LINEAMIENTOS TÉCNICOS PARA PROPORCIONAR ALTA DISPONIBILIDAD DE SERVICIO EN UNA NGSDP

Angélica María Burbano Cifuentes

Francisco Javier Calero Valenzuela

Universidad del Cauca

Facultad de Ingeniería Electrónica y Telecomunicaciones
Departamento de Telemática
Línea de Investigación en Servicios Avanzados de Telecomunicaciones

Popayán, Junio de 2011

LINEAMIENTOS TÉCNICOS PARA PROPORCIONAR ALTA DISPONIBILIDAD DE SERVICIO EN UNA NGSDP

Angélica María Burbano Cifuentes

Francisco Javier Calero Valenzuela

Trabajo de grado presentado como requisito para optar al título de Ingeniero en Electrónica y Telecomunicaciones

Director

Francisco Orlando Martínez Pabón
Magister en Ingeniería Telemática

Co-Director

Oscar Mauricio Caicedo Rendón
Magister en Ingeniería Telemática

Universidad del Cauca

Facultad de Ingeniería Electrónica y Telecomunicaciones
Departamento de Telemática
Línea de Investigación en Servicios Avanzados de Telecomunicaciones

Popayán, Junio de 2011

TABLA DE CONTENIDO

LISTA DE FIGURAS

LISTA DE TABLAS

Capítulo 1

1. Introducción

1.1. Planteamiento del problema

La saturación de redes, el estancamiento en los ingresos de los operadores y el número limitado de usuarios que acceden a los servicios de telecomunicaciones [1] [2], se ha evidenciado especialmente en áreas como la del servicio de telefonía fija, la cual ha alcanzado un número determinado de usuarios e ingresos que tienden a descender con el paso del tiempo; otros servicios como la telefonía móvil han presentado un crecimiento tan abrupto que ya casi ha conseguido el tope de clientes disponibles y por ende su nivel máximo de crecimiento [3]. En sentido contrario, la Internet, que ha presentado un gran crecimiento, muestra un número cada vez mayor de clientes e ingresos gracias a la diversidad de servicios y experiencias que ofrece a sus consumidores, quienes la prefieren por sus constantes innovaciones, algunas de las cuales se deben a la posibilidad que tienen los usuarios de desarrollar sus ideas en la creación de nuevos servicios y ofrecerlos a otros usuarios en la red [1] [4].

De acuerdo a este contexto, en la actualidad existen pocos clientes a quienes los operadores de telecomunicaciones les pueden vender líneas telefónicas, por lo cual se prevé que los nuevos servicios de datos que dan un valor adicional y atractivo para el usuario final, se convertirán en su principal fuente de sostenimiento y ganancias.

Con el ánimo de hacer frente a esta situación y con el interés de satisfacer las cada vez más cambiantes y diversas necesidades de los usuarios, las empresas de telecomunicaciones pretenden volver más atractivas sus plataformas, para lo cual apuntan hacia la convergencia de las redes y servicios, y también al acceso a los servicios ofrecidos en la Internet, ya que esto les permitiría, además de retener e incorporar nuevos clientes, diferenciarse de la competencia. En este sentido, la solución propuesta por los operadores de telecomunicaciones es la oferta de VAS (*Value Added Services)*. Con el fin de alcanzar dicho objetivo, los operadores de telecomunicaciones están implementando las SDP (*Service Delivery Platform*) que permiten la interconexión entre los diferentes tipos de redes y tecnologías, así como también el rápido desarrollo, despliegue, orquestación y ejecución de servicios. Sin embargo ante la falta de estandarización en la arquitectura de las SDP, muchos de los operadores de telecomunicaciones las adaptaron como sistemas propietarios, es decir que proporcionan interoperabilidad entre redes y servicios del mismo operador, pero no permiten la interconexión con otros, generando silos SDP [5].

Con el fin de permitir una arquitectura genérica y lograr el propósito para el cual fue diseñada la SDP, hoy en día se habla de NGSDP *(Next Generation SDP*), la cual es una SDP construida bajo los principios de SOA (*Service Oriented Architecture*) e integrada con IMS (IP Multimedia Subsystem), que además sirve como puente entre las redes de telecomunicaciones y la Web 2.0 [5]. Una NGSDP posee características como: bajo costo, bajo riesgo y rápido retorno de la inversión; además capacidad para soportar estándares, criterios y tecnologías, etc.

El futuro ambiente de convergencia planeado por los operadores de telecomunicaciones hará que los usuarios tradicionales tengan acceso a gran variedad de nuevos servicios y experiencias, lo que traerá grandes beneficios y retos que deben ser afrontados por dichos operadores; algunos de estos retos se centran en cómo lograr alta disponibilidad[1] de servicio en el contexto de la QoS (*Quality of Service*) a cumplir, en la prestación de servicios sobre una NGSDP. La alta disponibilidad implica que un dispositivo o red esté listo para ser usado cerca del 100% del tiempo como sea posible. La tolerancia a errores indica la capacidad de un dispositivo o la red para recuperarse del fallo de un elemento o dispositivo. Lograr una alta disponibilidad se basa en la eliminación de cualquier punto de fallo y en la distribución de la inteligencia en toda la arquitectura. En este sentido, se puede aumentar la disponibilidad mediante la adición de componentes redundantes, incluyendo dispositivos de red redundantes y conexiones a servicios de Internet redundante. Con el diseño adecuado, ni un solo punto de fallo afectará la disponibilidad de todo el sistema [6].

La definición de lineamientos[2] [7] técnicos en cuanto a alta disponibilidad de servicio es fundamental en el modelo de negocio de las telecomunicaciones para alcanzar el carrier grade[3] exigido por los operadores. Sin embargo, a pesar de existir numerosos métodos para conseguir servicios altamente disponibles, tales como Balanceo de cargas[4] y Clúster[5], en la actualidad hay deficiencia en cuanto a los lineamientos técnicos que deben ser tenidos en cuenta por parte de los operadores, investigadores y desarrolladores para poder lograrlos en la implementación de las SDP y NGSDP. A esto se suma la carencia de métodos con los que se puedan realizar pruebas y evaluaciones para comprobar que se esté cumpliendo con los estándares de calidad requeridos.

El gran número de clientes con los que cuentan los operadores de telecomunicaciones, es otro aspecto por el cual es de suma importancia contar con lineamientos técnicos que ayuden a proporcionar alta disponibilidad de servicio, además, cómo lo muestran numerosas encuestas [1], el deseo por nuevos servicios es muy alto, por lo cual se puede prever la gran demanda con la que van a contar estos, sin dejar de lado que el éxito o fracaso de un nuevo proyecto de este tipo depende en gran medida de estos aspectos, así como también son de gran utilidad para determinar las capacidades y limitaciones de las redes en un nuevo ambiente de convergencia.

A partir de lo expuesto, y siendo la alta disponibilidad de servicio un tema de tanta importancia en el marco de la QoS que deben proporcionar los operadores de telecomunicaciones, se propone la siguiente pregunta de investigación:

¿Cuáles son los lineamientos técnicos a tener en cuenta para proporcionar alta disponibilidad de servicio sobre una NGSDP?

[1] La disponibilidad en el trabajo de grado, se limita a la disponibilidad de la plataforma NGSDP y su hardware asociado, no a la disponibilidad asociada a la fuerza eléctrica (infraestructura eléctrica).
[2] Lineamiento: Directrices teóricas agrupadas por cada una de las dimensiones, áreas o perspectivas del proceso.
[3] Carrier grade: Sistema o componente hardware o software que es muy fiable, probado y demostrado en sus capacidades, para cumplir o superar el 99.999 % de disponibilidad sobre los recursos en la prestación de servicios, conocido como “cinco nueves”
[4] Balanceo de cargas: Permite dividir las tareas que tendría que soportar una máquina, con el fin de maximizar las capacidades de procesos de datos, así como de ejecutar tareas
[5] Clúster: Grupo de computadores que trabajan con un fin común; agrupan hardware, redes de comunicación y software para trabajar conjuntamente como si fuera un único sistema.

1.2. Objetivos del trabajo de grado

1.2.1. Objetivo General

Proporcionar lineamientos técnicos para cumplir con los criterios asociados a la alta disponibilidad de servicio en el contexto de una NGSDP.

1.2.2. Objetivos específicos

- Caracterizar el despliegue de servicios convergentes altamente disponibles en un ambiente NGSDP.
- Definir un conjunto de lineamientos técnicos para proporcionar servicios con alta disponibilidad en el contexto de una NGSDP.
- Evaluar los lineamientos propuestos a través de un prototipo consistente en el despliegue de un servicio que consuma capacidades de Internet y de una red de telecomunicaciones en una NGSDP simulada.

1.3. Aportes del trabajo de grado

- Caracterización del despliegue de servicios convergentes altamente disponibles en un ambiente NGSDP.
- Un conjunto de lineamientos técnicos para proporcionar un servicio altamente disponible en el entorno de una NGSDP.
- La evaluación de los lineamientos propuestos en el desarrollo de este trabajo de grado.

1.4. Estructura del trabajo de grado

- **Capítulo 2:** se abordan las definiciones formales de conceptos claves para el entendimiento del proyecto realizado; se construye una base inicial de conocimiento sobre los temas directamente relacionados con el presente trabajo de grado, en la cual se incluyen sus características más relevantes y las definiciones utilizadas para el mismo.
- **Capítulo 3:** se realiza una conceptualización alrededor de los servicios de telecomunicaciones y de los servicios Web y se explica el fenómeno de la convergencia en el sector de las telecomunicaciones; para finalizar el capítulo se abordan los aspectos que favorecen el despliegue de servicios convergentes sobre las plataformas NGSDP.
- **Capítulo 4:** en este capítulo se resumen los criterios relacionados con la alta disponibilidad de servicio, que en el presente trabajo de grado se consideran como los más relevantes para una plataforma de este tipo, además se plantean lineamientos para cumplir con cada uno de los criterios seleccionados.

- **Capítulo 5:** un prototipo de laboratorio en el que se evalúan los lineamientos planteados a través de un diagnóstico previo consignado en el Anexo B; en este contexto, el prototipo es sometido a dos diferentes escenarios de evaluación: *i)* sin la aplicación de lineamientos, *ii)* con la implementación de lineamientos. Al final del capítulo se presentan algunas conclusiones y el análisis respectivo.

- **Capítulo 6:** se resumen los aportes y conclusiones del trabajo y se plantean futuros proyectos relacionados.

- **Anexo A:** proceso de selección Filtro Telco.

- **Anexo B:** entrevista para el diagnóstico de los requerimientos de alta disponibilidad. Los cuestionamientos y resultados aquí presentados, corresponden a varias entrevistas realizadas a expertos de EMCALI, con el objetivo de obtener una valoración concreta sobre los requerimientos reales de alta disponibilidad en el contexto de un operador de telecomunicaciones, que finalmente guiaron el proceso de implementación del prototipo descrito en el capítulo 5.

- **Anexo C:** manuales de instalación y uso de herramientas utilizadas en el presente trabajo de grado.

- **Anexo D:** detalles técnicos de la implementación del prototipo.

- **Anexo E:** artículo publicado en el VI Congreso Ibero-americano de Telemática CITA 2011, Gramado RS (Brasil), 16-18 Mayo 2011.

Capítulo 2

Estado del Arte

2. Marco teórico

En el presente capítulo se realiza una contextualización de los términos más relevantes para el presente trabajo de grado, también se presentan algunos de los trabajos relacionados con la temática de este, junto con sus brechas y los aportes que realizan a la presente investigación.

2.1. Contextualización

Esta sección contiene los conceptos más relevantes que permitirán la comprensión del entorno en el que se desarrolla el problema central del presente trabajo de grado. Por lo cual a continuación se explican conceptos relacionados con las plataformas SDP, así como también aspectos relacionados con la disponibilidad y la alta disponibilidad en los sistemas de telecomunicaciones.

2.1.1. NGSDP

La SDP ha surgido a raíz de la evolución de las redes de telecomunicaciones, esto debido a que es un componente clave en las telecomunicaciones convergentes, en las cuales se sustituyen las arquitecturas de red verticales, por arquitecturas horizontales y comunes, en las cuales se facilita la entrega de servicios a través de diferentes tipos de redes, así como la creación de aplicaciones Web o del mundo IT que utilicen capacidades de redes del mundo de las telecomunicaciones, sin embargo y a pesar de todas los beneficios de las SDP, el termino representa un concepto sin una definición unificada [5] [8] [9] [10].

Según el grupo Moriana [11] las SDP han experimentado tres generaciones; inicialmente en él año 2000, el termino SDP fue introducido para describir una arquitectura común de servicios, específicamente diseñada para la entrega de contenido móvil y servicios de mensajería. Desde el año 2003 al 2006, la SDP evolucionó para soportar servicios de voz, localización, multimedia, presencia y de carga; esta fue conocida como la segunda generación de SDP, la cual además de soportar dichos servicios, adoptó tecnologías estándar del mundo IT y ofreció seguridad y gestión del acceso de las terceras partes a los servicios de red a través del surgimiento de servicios Web estándar para telecomunicaciones. La NGSDP o tercera generación de SDP (SDP 2.0), es construida bajo los principios SOA (Service Oriented Architecture), lo que permite la eficiente integración y orquestación de servicios así como también la gestión del ciclo de vida de la aplicación.

El término NGSDP es definido de forma más detallada por el grupo Moriana como un entorno completo que posibilita la eficiente creación, despliegue, orquestación y gestión de una o más clases de servicios. También posibilita el acceso seguro a las terceras partes, administración de las capacidades de servicio y además sirve de puente entre las telecomunicaciones y la web 2.0 [11]. Por esta razón la plataforma NGSDP se encuentra ubicada en la capa de servicio en la arquitectura de las telecomunicaciones y se posiciona como un elemento clave de esta (Figura 1) [11].

Según el grupo Moriana una SDP de tercera generación debe poseer cinco características: *i)* tecnologías estándar, *ii)* arquitectura horizontal, *iii)* integración basada en SOA, *iv)* entorno común, *v)* creación de servicios estándar. Estas características permiten diferenciar entre una verdadera NGSDP y un silo SDP.

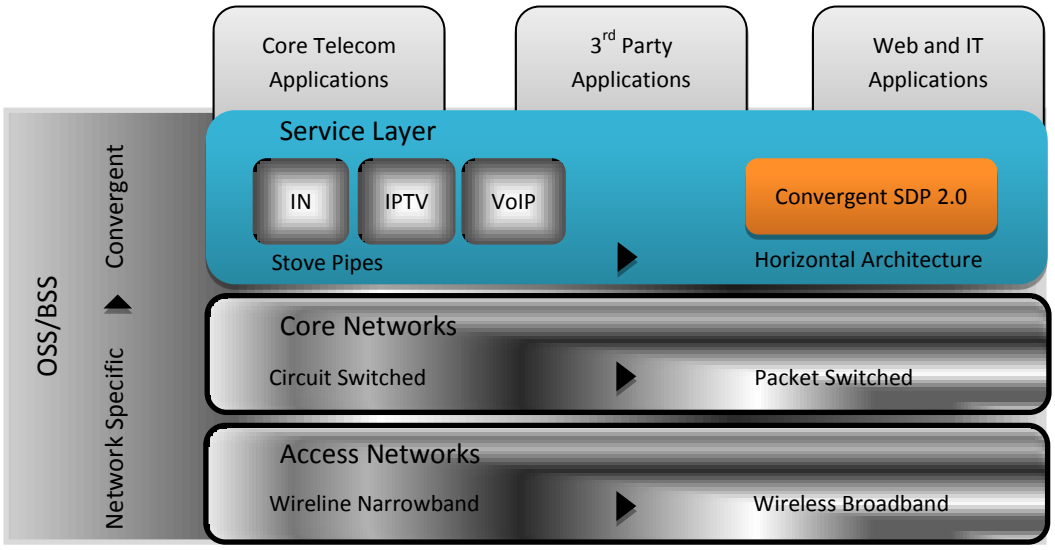

Figura 1. Ubicación de la NGSDP en arquitectura de las telecomunicaciones

2.1.1.1. Arquitectura de una NGSDP

A continuación se presenta y explica cada uno de los elementos de la arquitectura de una plataforma NGSDP (Figura 2). Según el grupo Moriana una NGSDP posee una arquitectura en capas compuesta por los siguientes elementos [12]:

- Capa de exposición de servicio.
- Capa de orquestación y Gestión de servicio.
- Capa de servicios de telecomunicaciones y Habilitadores de Servicio.
- Capa de Creación y Ejecución de Servicio.
- Capa de Abstracción de red.

Figura 2. Arquitectura de plataforma NGSDP. Fuente: Grupo Moriana

2.1.1.1.1. Capa de abstracción de red

Esta capa está conformada por habilitadores de servicio de bajo nivel que proveen acceso a las capacidades de red y a los servicios subyacentes como: mensajería móvil (SMS, MMS), localización, presencia, carga, control de llamadas, gestión de sesión, gestión multimedia, entre otros. También incluye la funcionalidad de gestión de políticas, controlando de esta forma el acceso de los diferentes servicios y componentes de la SDP a los elementos de la red.

Los habilitadores de servicio usualmente se comunican con elementos de la red a través de protocolos estándar como: SS7 (Signaling System No. 7), SIP (Session Initiation Protocol), Diameter, MM7[6], SMTP (Simple Mail Transfer Protocol), HTTP (Hypertext Transfer Protocol), entre otros. Las capacidades de servicio son expuestas hacia las capas superiores de la SDP a través de un conjunto habilitadores de servicio estándar como: SIP Servlets, Apis OSA/Parlay, JAIN SLEE, servicios Web , etc., [12].

2.1.1.1.2. Capa de Creación y Ejecución de Servicio

Esta capa se compone de uno o más entornos de ejecución de servicio incluyendo tecnologías estándar como Java EE Application Servers, así como también plataformas específicas de los operadores como: SIP Application Servers, JAIN SLEE (Java APIs for Integrated Networks Service Logic Execution Environments) y servidores Parlay[7] para telecomunicaciones. Los cuales típicamente son basados en Java y siguen estándares abiertos de las telecomunicaciones y del mundo IT.

Igualmente, incluye la gestión de los servicios de interacción y la función de intermediación de servicio, tal como SIP SCIM (SIP Service Capability Interaction Manager). El intermediador de

[6] MM7 interfaz para servicios de valor agregado.
[7] Parlay: conjunto de API estandarizadas para las telecomunicaciones.

servicios (Service broker) resuelve las interacciones de servicio y otros conflictos entre los servicios activados simultáneamente en la misma sesión, también puede tener funciones de traducción de protocolos y es un componente clave de la SDP convergente, donde los servicios usan tanto protocolos heredados como protocolos IMS/SIP en la gestión de sesión [12].

2.1.1.1.3. Capa de Servicios de Telecomunicaciones y Habilitadores de servicio

Esta capa contiene servicios críticos del núcleo de los operadores, tales como VPN (Virtual Private Network), Prepaid (pago por adelantado), IP Centrex o telefonía VoIP de clase 5. También permite la implementación de habilitadores de servicio de alto nivel como: conferencia, mensajería, presencia, distribución de llamadas, administración de identidad, de contenido y de medios, etc. Para el correcto funcionamiento de los servicios y de los habilitadores de servicio nombrados anteriormente, estos deben correr en una red cerrada, es decir directamente en las plataformas de servicio de los operadores de telecomunicaciones [12].

2.1.1.1.4. Capa de orquestación y Gestión de servicio

Esta capa incorpora los principios de SOA y los servicios Web a la plataforma SDP. Esto incluye los ESB (Enterprise Service Bus) para la integración con los sistemas OSS/BSS (Operational Support System/ Business Support Systems), así como también la integración al interior de la SDP. También incluye un motor BPEL (Business Process Execution) y otros mecanismos para el aprovisionamiento de servicios, la gestión del ciclo de vida y el desempeño de los mismos.

Adicionalmente, incluye características comunes de la SDP como repositorios de servicios, de perfiles y gestión de la identidad compartida entre todos los servicios desplegados [12].

2.1.1.1.5. Capa de Exposición de Servicio

Esta capa permite la gestión y el acceso seguro de terceros a los habilitadores de servicio y a las capacidades de bajo nivel de la red del operador de telecomunicaciones. Típicamente esta capa contiene un portal de gestión de las terceras partes, el cual les permite a los proveedores de servicio y a los desarrolladores de aplicaciones llegar a acuerdos de auto aprovisionamiento con el operador de telecomunicaciones. Otro elemento importante de esta capa es el Web Services Gateway, el cual implementa un conjunto de servicios Web estandarizados o propietarios para mapear los habilitadores de servicios.

Adicionalmente, contiene funciones de administración para el control del acceso de las terceras partes, entre las que se encuentran: autorización, autenticación y facturación, así como también políticas de ejecución y gestión de acuerdos de servicios de red [12].

2.1.2. SOA

SOA ha surgido como un exitoso sistema para la integración de los sistemas de soporte de operaciones y de facturación (OSS/BSS) de los operadores y actualmente se ha posicionado como un elemento clave para la capa de servicios de los operadores de telecomunicaciones.

SOA es una evolución de los sistemas de computación distribuidos y de la programación modular. Es una arquitectura proveniente del mundo IT, que no requiere de una tecnología específica para su implementación; además guía todos los aspectos de creación y el uso de procesos de negocio en el ciclo de vida de los servicios. Entre los beneficios de SOA se encuentra que el costo por aplicación es muy bajo, debido a que todo el software requerido para producir una nueva aplicación puede ser obtenido de otras aplicaciones ya existentes y solo es necesario el uso de orquestación[8] para crear un nuevo servicio. Desde el punto de vista de los operadores de telecomunicaciones, SOA se constituye en una arquitectura flexible que permite la innovación y adaptación de servicios y además el aprovechamiento de los estándares IT en el mundo de las telecomunicaciones.

En la infraestructura del mundo IT, SOA le permite a las diferentes aplicaciones el intercambio de datos para formar procesos de negocio, los cuales pueden requerir los servicios de múltiples plataformas IT. SOA separa las funciones en unidades de servicio, las cuales pueden ser distribuidas sobre la red y pueden ser combinadas y reutilizadas para crear nuevas aplicaciones de negocio. Las unidades de servicio se comunican entre sí mediante el intercambio de datos o mediante la coordinación de una actividad entre dos o más unidades de servicio.

La organización OASIS (Organization for the Advancement of Structured Information Standards) define a SOA como un paradigma para la organización y utilización de las capacidades distribuidas, que pueden estar bajo el control de diferentes propietarios, proporcionando un medio uniforme para ofrecer, descubrir, usar e interactuar con dichas capacidades [13].

Los servicios son unidades no asociadas de funcionalidad, quienes no se realizan llamados entre sí. Esto causa algunas confusiones a los operadores de telecomunicaciones, ya que desde la perspectiva de éstos los servicios son aquellos por los cuales los clientes finales pagan, por lo que el término “servicios SOA” se utiliza a menudo para diferenciar esta clase de servicios de los servicios que son entregados al cliente final. Típicamente los servicios SOA implementan funciones como el llenado de una solicitud en línea para una cuenta, consulta de extractos de cuenta en línea o reservas en línea de boletos de avión. En lugar de integrar los servicios a través de invocaciones mutuas en sus códigos fuente, se definen protocolos que describen cómo uno o más servicios se pueden comunicar entre sí. Por lo tanto esta arquitectura se basa en procesos de negocio que realizan enlaces y secuencias de servicios en el proceso de orquestación, para cumplir con los requisitos de negocio de los sistemas.

Generalmente los servicios Web son usados en la implementación de SOA. En estos se encuentra la descripción de las operaciones ofrecidas por el servicio, las cuales se encuentran escritas en el lenguaje para la descripción de servicios (WSDL, Web Services Description Language) y mediante el uso del catálogo de negocios de Internet UDDI (*Universal Description, Discovery and Integration*) son descubiertos. En la implementación de SOA, el lenguaje de marcado extensible (XML, e*Xtensible Markup Language*) ha sido utilizado para crear los datos que son puestos en la descripción del contenedor.

[8] En el ambiente de las NGSDP el término orquestación se refiere al proceso de creación de un servicio o aplicación a partir de otras aplicaciones existentes.

En realidad, SOA no está atada a una tecnología específica, y puede ser implementada usando una amplia variedad de tecnologías, como: SOAP (*Simple Object Access Protocol*), REST (Representational State Transfer), RPC (*Remote Procedure Call*), DCOM (Distributed Component Object Model), CORBA (*Common Object Request Broker Architecture*), servicios Web, WCF (Windows Communication Foundation). Las interfaces definidas por los servicios son la clave de SOA, ya que éstos pueden ser llamados de manera estándar para realizar sus tareas, sin que la aplicación tenga o necesite conocimiento de cómo el servicio realiza su tarea [14].

SOA trae grandes beneficios al mundo de las telecomunicaciones, como la introducción de nuevos servicios rentables de manera rápida y eficiente, y la entrega y la facturación de éstos a los usuarios, sin importar el tipo de conexión de la cual dispongan (banda ancha, wireless, etc.). En este contexto, los operadores son forzados a redefinir sus modelos de negocio. En síntesis, SDP se basa en los principios SOA a fin de hacer uso de las capacidades del mundo IT, además mediante el uso de los adaptadores SOA en la arquitectura de la NGSDP los operadores de telecomunicaciones pueden exponer de forma segura los elementos subyacentes como: servicios, contenido y las funciones de los sistemas de soporte de operaciones y de negocios (OSS/BSS) [12].

2.1.3. IMS

IMS es una arquitectura multimedia de nueva generación abierta, estandarizada y de fácil manejo que combina Internet y el mundo móvil con servicios de línea fija, utilizando tecnologías celulares para facilitar el acceso ubicuo y tecnologías de Internet para ofrecer nuevos servicios atractivos a los usuarios [15].

IMS fue definido por el 3GPP (3rd Generation Partnership Project), en estrecha colaboración con el IETF (Internet Engineering Task Force), y originalmente fue diseñado para evolucionar las redes UMTS (Release 5). Posteriormente surgieron complementos (Releases 6, 7 y 8), que permiten a los operadores ofrecer servicios interactivos e interoperables de manera rentable e independiente de la red de acceso [16]. IMS ha sido adoptado por otros organismos de estandarización como 3GPP2 Y ETSI (European Telecommunications Standards Institute).

IMS no define las aplicaciones o servicios que se le pueden ofrecer al usuario final, solo define la infraestructura y las capacidades del servicio que los operadores o proveedores de servicio pueden emplear para construir sus propias aplicaciones.

IMS presenta una arquitectura de red horizontal que define cuatro capas, como se muestra a continuación (Figura 3):

- **Capa de acceso**: representa las tecnologías de acceso. Se encuentran ubicadas las diferentes redes de acceso de los operadores de telecomunicaciones como: HFC (Hybrid Fiber Coaxial), xDSL (x Digital Subscriber Line), FTTP (Fiber to the Premises), Wi-Fi, WiMax (World Wide Interoperability for Microwave Access), entre otros.
- **Capa de Transporte:** es la capa física del núcleo IP y está compuesta de enrutadores interconectados.

- **Capa de Control de Sesión**: se encarga de la lógica de señalización, permitiendo establecer, modificar y terminar sesiones por medio de los servidores de control de estado de llamada (CSCF, Call Session Control Function), y de ejecutar servicios de valor agregado para el usuario por medio del servidor de funciones de recursos multimedia (MRF, Multimedia Resource Function).
- **Capa de Aplicación o Servicio:** esta capa le permite al operador ofrecer servicios propios o de terceros, mediante los servidores de aplicaciones (AS, Application Server), y su control a través de SIP, permitiéndole diferenciarse de su competencia. También está formada por el HSS (Home Subscriber Server) que es una base de datos donde se almacena la información de los usuarios.

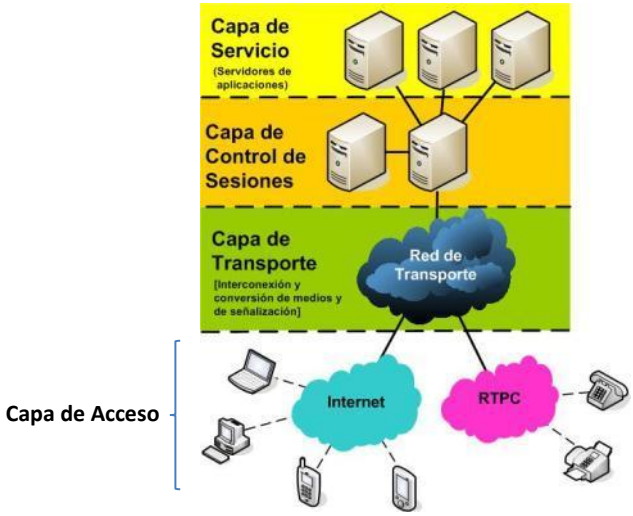

Figura 3. Modelo de Referencia de capas de IMS [17]

La arquitectura IMS (Figura 4) está compuesta de varios elementos e interconexiones, algunos de ellos se explican a continuación [18] [19] :

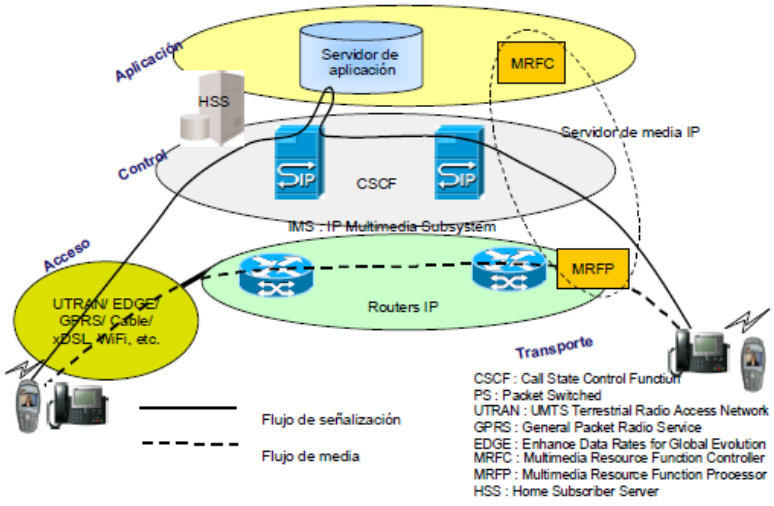

Figura 4. Arquitectura de Redes y Servicios IMS [18]

Terminal IMS: es una aplicación que emite y recibe solicitudes SIP. Es un software instalado en un PC, un teléfono IP o una estación móvil UMTS y se denomina UE (User Equipment).

HSS (Home Subscriber Server): es la base principal de almacenamiento de datos de usuarios y de los servicios a los cuales están suscritos. Se almacenan las identidades del usuario, informaciones de registro, parámetros de acceso e información que permite la invocación de los servicios suscritos. La entidad HSS interactúa con las entidades de la red a través del protocolo DIAMETER.

CSCF (Call Session Control Function): el control de llamada iniciado por un terminal IMS tiene que ser asumido en la red a la cual el usuario suscribe sus servicios IMS ya que el usuario puede suscribirse a una gran cantidad de servicios y algunos de ellos pueden estar no disponibles o pueden funcionar de manera diferente. Eso induce la definición de tres entidades: P-CSCF (Proxy-CSCF), I-CSCF (Interrogating-CSCF) y S-CSCF (Serving-CSCF).

- ✓ P-CSCF (Proxy-CSCF): es el primer punto de contacto entre un terminal IMS y la red. Puede estar colocado tanto en la red local como en una red de otra compañía. Sirve para enrutar la conexión hacia los I-CSCF. Además el P-CSCF Coordina con la red de acceso, autorizando el control de recursos y la calidad de las llamadas y/o sesiones (QoS).

- ✓ S-CSCF (Serving-CSCF): es el nodo central en el plano de señalización de IMS. Es un servidor SIP y coordina con otros elementos de la red el control de las llamadas y/o sesiones. Es el nodo en la arquitectura IMS que se conecta con el HSS, para descargar o actualizar perfiles de usuarios utilizando el protocolo DIAMETER el cual es usado para funcionalidades de AAA (Authentication, Authorization and Accounting). El S-CSCF tiene también una función de control de servicio que le permite interactuar con los AS para soporte de servicios y aplicaciones. Provee seguridad para la sesión.

- ✓ I-CSCF (Interrogating-CSCF): es el punto de contacto en la red de un operador para todas las conexiones destinadas a un suscriptor de la red de este operador, o para un suscriptor visitando su red. Pueden existir múltiples I-CSCF en una red. La dirección IP de este servidor se publica en el DNS (Domain Name System) del dominio al que pertenece. De esta manera, otros servidores remotos pueden encontrarlo y usarlo como punto de reenvío de paquetes SIP hacia ese dominio.

BGCF (Breakout Gateway Control Function): identifica la red donde se accede a la red pública conmutada (PSTN). Si se determina que el acceso ocurre en la misma red donde el BGCF está localizado, entonces este selecciona un MGCF; este será responsable por el interfuncionamiento con la red PSTN. Si el punto de acceso está en otra red, el BGCF enviará la señalización de esta sesión a un BGCF.

MGCF (Media Gateway Control Function): provee la función de interfuncionamiento de señalización entre los elementos de la red IMS y las redes heredadas (PSTN). El MGCF controla un conjunto de MGWs a través de la señalización H.248, la cual permite el establecimiento de recorridos para las sesiones.

MRFC (Multimedia Resource Function Controller): controla los recursos de media del elemento Multimedia Resource Function Processor (MRFP), para proveer tonos, anuncios y conferencias.

Signaling Gateway: provee la conversión de señalización en ambas direcciones en la capa de transporte entre SS7 y la señalización basada en IP (por ejemplo ISUP/SS7 e ISUP/SCTP/IP).

Todos los elementos de la red IMS se pueden apreciar a continuación (Figura 5).

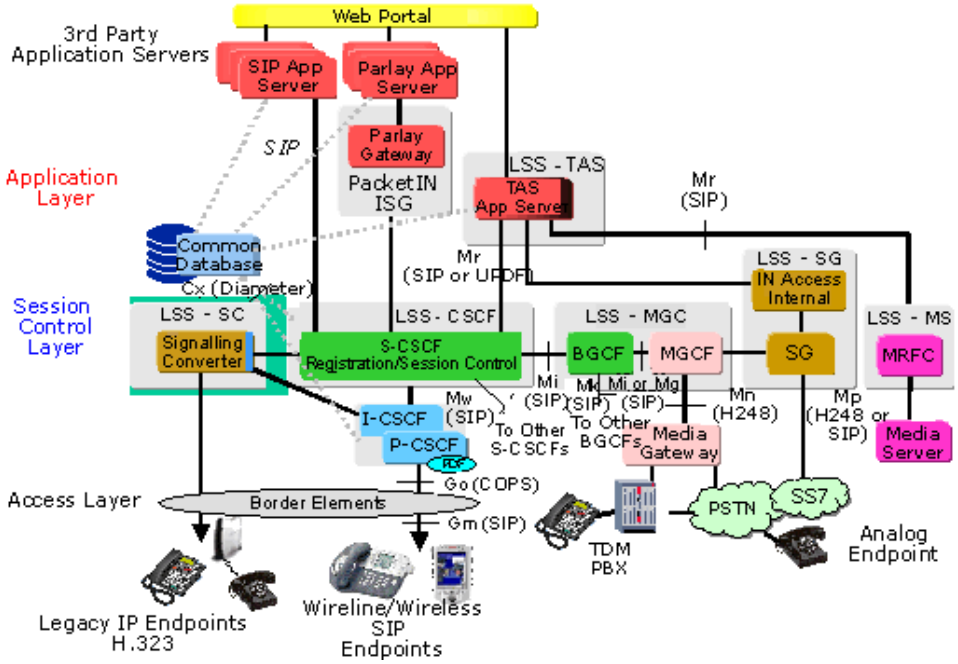

Figura 5. Componentes de la red IMS [19]

IMS en el contexto de una NGSDP: el control de servicio para sesiones multimedia IP de la arquitectura IMS, se fija entre la red y la SDP. Un operador no requiere la implementación de toda la arquitectura IMS, pero si necesita un control de sesión IP basado en SIP/soft-switch o IMS. Para el correcto funcionamiento de la NGSDP en la infraestructura de un operador de telecomunicaciones, la NGSDP debe soportar protocolos como SIP y Diameter y capacidades de servicio propias de IMS, así como también interfaces estandarizadas y soporte a la creación de servicios SIP.

2.1.4. Disponibilidad

La disponibilidad es un concepto que permite expresar la capacidad del sistema de estar en condiciones de funcionamiento adecuado en un momento determinado dentro de un intervalo de tiempo determinado [20].

En términos generales, se puede decir que la disponibilidad es una medida de la frecuencia con la que se puede utilizar la aplicación. Es decir la disponibilidad es un cálculo porcentual del tiempo en que la aplicación está realmente disponible para resolver las solicitudes de servicio en comparación con el tiempo de ejecución total disponible previsto. El cálculo formal de la disponibilidad está expresado de manera general mediante la ecuación básica (Ec.1),la cual

relaciona el tiempo promedio entre fallas (MTBF, Mean Time Between Failure) y el tiempo medio para la reparación una vez ocurrida la falla (MTTR, Mean Time To Repair), donde A, representa la disponibilidad (Availability) general [21]:

A= MTBF/(MTBF + MTTR) **Ec.1. Fórmula básica para el cálculo de disponibilidad [22] [23].**

La fórmula anterior de propósito general, puede ser ampliada como se muestra en la ecuación (Ec.2); ésta fórmula expresa de manera adecuada la disponibilidad desde el punto de vista de prestación del servicio, donde AST (Agreed Service Time) corresponde al tiempo acordado de servicio y DT (Down Time) es el tiempo de interrupción del servicio durante las franjas horarias de disponibilidad acordadas, haciendo el cálculo más práctico debido a que es más fácil conocer estos tiempos.

La disponibilidad se expresa con mayor frecuencia a través del índice de disponibilidad, el cual se determina dividiendo el tiempo durante el cual el servicio está disponible, por el tiempo total [24].

Según [25] la fórmula para calcular la disponibilidad, está dada de la siguiente forma:

%Disponibilidad= ((AST-DT)/AST)*100 **Ec.2. Fórmula del cálculo de la disponibilidad de los servicios**

A medida que aumenta la complejidad y los niveles de acceso a las aplicaciones, se incrementa la probabilidad de que surjan errores: en el diseño original, en los servicios de soporte técnico, o en el mantenimiento. Esto genera posibles paradas del sistema o de los servicios, lo cual implica la disminución de la disponibilidad. Los tiempos de inactividad originan problemas y pérdidas a largo plazo, a la empresa y a los usuarios.

Se pueden producir errores en las aplicaciones por diferentes motivos, como [23]:

- Comprobación inadecuada.
- Problemas relacionados con cambios en la administración.
- Falta de control y análisis continuos.
- Errores en las operaciones.
- Código poco consistente.
- Ausencia de procesos de diseño de software de calidad.
- Interacción con aplicaciones o servicios externos.
- Condiciones de funcionamiento distintas (cambios en el nivel de uso, sobrecargas máximas).
- Sucesos inusuales (errores de seguridad, desbordamientos en la difusión).
- Errores de hardware (discos, controladores, dispositivos de red, servidores, fuentes de alimentación, memoria, CPU).
- Problemas de entorno (red eléctrica, refrigeración, incendios, inundaciones, polvo, catástrofes naturales).

2.1.5. Alta disponibilidad

La alta disponibilidad garantiza un grado absoluto de continua funcionalidad dentro de una ventana de tiempo determinada y consiste en el funcionamiento de la infraestructura tecnológica las 24 horas del día, los siete días de la semana y los 365 días del año, proporcionando el 99.999% o cinco nueves de disponibilidad, sin importar que se presenten fallos en la red [26] [27] [28] [29].

La disponibilidad se define como un número garantizado de nueves como se muestra en la Tabla 1, en la cual se muestra el número de minutos o segundos de parada (downtime) estimados en un sistema en relación con el número de minutos al año, es decir 525.600 minutos. Para propósitos de mercadeo, este número de minutos simboliza el 100% de disponibilidad. Para lograrlo, se requiere integrar soluciones que se componen de energía, procesadores, discos duros, software, personal capacitado, consultoría y procesos detallados; además se requiere redundancia en los componentes críticos.

Availability %	Downtime in Minutes	Downtime per Year	Vendor Jargon
90	52,560.00	36.5 days	one nine
99	5,256.00	4 days	two nines
99.9	525.60	8.8 hours	three nines
99.99	52.56	53 minutes	four nines
99.999	5.26	5.3 minutes	five nines
99.9999	0.53	32 seconds	six nines

Tabla 1. Porcentajes de disponibilidad en sistemas de Telecomunicaciones [27]

Desde el punto de vista de los usuarios, los servicios de telecomunicaciones tradicionales siempre están disponibles, suministrando una calidad cercana al 100% del tiempo en las llamadas o sesiones y garantizando que el discado a un determinado número sea casi instantáneo.

Para ofrecer esta experiencia al usuario, las compañías de telecomunicaciones aplican a sus redes redundancia en el enrutamiento y hacen uso de componentes en la infraestructura de red que brinden alto rendimiento, baja latencia en el procesamiento y tolerancia a fallos. Además despliegan capacidades sofisticadas en la gestión de red y funcionan con estrictos acuerdos de nivel de servicio (SLA, Service Level Agreement). Las actualizaciones ocurren incrementalmente, para proveer un servicio continuo mientras se adapta y mejora el sistema.

Los tiempos de caídas y de paradas planeados y no planeados casi nunca se presentan en las redes de telecomunicaciones tradicionales, mientras que los periodos planeados y no planeados son muy comunes en el mundo IT e Internet y estos pueden durar varias horas. En un escenario de convergencia se planea proporcionar la misma calidad de los servicios tradicionales de telecomunicaciones a los nuevos servicios convergentes, con el fin de mantener y atraer nuevos usuarios, y de esta forma aumentar los ingresos de los operadores de telecomunicaciones.

Para responder a las expectativas de los usuarios de telecomunicaciones, OpenCloud sugiere que las redes de nueva generación (NGN, New Generation Network) deben presentar características similares al carrier grade brindado por los operadores de telecomunicaciones en los servicios tradicionales [30].

En este sentido, OpenCloud, define las características de carrier grade que deben ser cumplidas en las NGN, con el fin de garantizar que los nuevos servicios convergentes funcionen adecuadamente en el momento que sean solicitados por los usuarios, de la siguiente manera [30]:

- **Ningún punto único de fallo:** redundancia hardware o software a fin de aumentar la fiabilidad del sistema.
- **99.999% de disponibilidad:** el objetivo esperado para las NGN es extender el carrier grade de los servicios tradicionales de telecomunicaciones a los nuevos servicios convergentes ofrecidos sobre esta.
- **Disponibilidad continúa de servicios:** disponibilidad continua, sin importar los posibles fallos en la plataforma.
- **Disponibilidad continúa de sesiones de servicio:** múltiples cambios en dispositivos, manteniendo las sesiones de usuario continuas
- **Auto monitoreo y auto reparación:** plataforma que se auto monitorea, a fin de minimizar posibles fallos; metodología proactiva que minimiza los tiempos de indisponibilidad y las perdidas asociadas a estos tiempos.
- **Protección de sobrecarga y gestión de la calidad de servicio**: a fin de garantizar alta disponibilidad en los servicios prestados, se debe monitorear constantemente el número de usuarios y peticiones a las que el sistema puede responder de manera adecuada.
- **Baja latencia en la ejecución:** bajo retardo en la transmisión de la información, aspecto clave en sistemas de alta disponibilidad.
- **Alto rendimiento:** procesamiento de gran cantidad de peticiones a fin de satisfacer las crecientes necesidades de los usuarios.

2.2. Trabajos Relacionados

"Lineamientos para composición de servicios de telecomunicaciones en un entorno JAIN SLEE basado en software de libre distribución"

En [31] los autores muestran como el movimiento dinámico del mercado a posicionado a las tecnologías de composición de servicios como un factor estratégico para el sector de las telecomunicaciones, debido a que permiten hacer uso de los servicios previamente desarrollados como bloques constructores, los cuales trabajando de forma conjunta crean nuevos y más complejos servicios de una forma rápida y flexible. Bajo este enfoque los autores se centran en la especificación JAIN SLEE, la cual define un modelo orientado a componentes para la estructuración de la lógica de aplicaciones y servicios de comunicaciones como un conjunto de componentes reutilizables y orientados a objetos; la cual facilita la composición de servicios de valor agregado,

así como la reducción del Time to Market, trayendo de esta forma importantes beneficios para los operadores de telecomunicaciones.

El trabajo proporciona diversos aspectos relacionados e importantes para el presente trabajo de grado, como lo son: SDP, JAIN SLEE, SOA, los múltiples proveedores de las plataformas SDP; realizando una descripción detallada de cada uno de estos. Adicionalmente aportan un modelo para la construcción de lineamientos. Sin embargo en la evaluación del prototipo no se realizan pruebas a fin de determinar la disponibilidad del entorno JAIN SLEE seleccionado.

"Criterios Técnicos para el aprovisionamiento de VAS en una NGN dentro del Contexto Colombiano"

En [32] los autores muestran las razones que motivan la creación de servicios de valor agregado, y la necesidad existente en el mercado de guías que muestren claramente los aspectos a tener en cuenta al momento de crear esta clase de servicios.

El trabajo proporciona una base de conocimiento alrededor de las SDP y sus funciones dentro de una NGN, además de esto se presentan criterios técnicos para el aprovisionamiento de VAS en Colombia y un caso de estudio que los evalúa, lo cual es de gran aporte para el trabajo de grado. Sin embargo, en este caso de estudio los criterios técnicos propuestos y el ambiente de ejecución utilizado, son aplicados al contexto específico de la NGN en el contexto Colombiano y no se realizan mediciones a fin de comparar entre las diferentes soluciones JAIN SLEE cual ofrece mejor disponibilidad. Este documento también proporciona un modelo para la construcción del prototipo de laboratorio del presente trabajo de grado.

"Designing High-Availability Services"

En [6] se muestra como la implementación de la alta disponibilidad en la redes de hoy en día, se ha convertido en un requerimiento fundamental para las empresas de telecomunicaciones, debido a que mejora aspectos claves para estos, como: i) aplicaciones críticas siempre disponibles, ii) mejorar la satisfacción y lealtad de empleados y clientes, iii) reducir los costos asociados a las reparaciones reactivas, iv) reducir pérdidas de dinero, v) minimizar las pérdidas de productividad. Además muestra como está caracterizada la alta disponibilidad para Cisco y propone una serie de preguntas que deben ser formuladas a los operadores de telecomunicaciones que deseen implementar servicios con alta disponibilidad a fin de identificar las necesidades relacionadas con los servicios, las falencias en las redes y los procedimientos de las organizaciones. También desarrolla algunos lineamientos para lograr alta disponibilidad en redes empresariales y de campus; finalmente propone una serie de pasos que deben ser tenidos en cuenta por las organizaciones a fin de llevar buenas prácticas dentro de las mismas.

Los autores proporcionan un modelo para la alta disponibilidad y un modelo para la construcción de lineamientos, y además aporta aspectos que pueden ser tenidos en cuenta para mejorar las buenas prácticas dentro de las organizaciones; y una serie de preguntas con las que se determinan las necesidades con respecto a la alta disponibilidad dentro de las organizaciones. No obstante, la solución proporcionada se basa en el uso de herramientas propietarias, específicamente el hardware y software de la compañía CISCO.

"Propuesta de Solución de Alta Disponibilidad de los Servicios Críticos del Centro de Datos de la Universidad del Cauca"

En [22] se propone una solución de alta disponibilidad para los servicios críticos en un centro de datos universitario (CDU), mediante una guía metodológica en la cual se encuentran una serie de lineamientos reunidos en siete fases; estas fases cuentan con pasos específicos y se encuentran estructurados de manera secuencial. Se facilitan definiciones y conceptos de disponibilidad y alta disponibilidad, y también de aspectos relacionados con ellos como los clústeres y la virtualización, importantes para el trabajo a desarrollar. Se muestra la solución para un caso de estudio específico, el CDU de la Universidad del Cauca, donde se hace uso de la guía metodológica propuesta, de una forma satisfactoria. Sin embargo en la implementación no se realizó un estudio comparativo a fin de determinar las herramientas de gestión y de monitoreo más adecuadas a fin de medir la disponibilidad del sistema.

"Providing Open Architecture High Availability Solutions"

En [33] el HA Forum explica el significado que tiene la alta disponibilidad para un sistema y cuáles son las capacidades que se necesitan para brindar este aspecto. También se explican cuales son los bloques de componentes hardware y software apropiados y necesarios para la creación de Sistemas de Arquitectura Abierta (Open Architecture Systems). También se definen las cinco etapas que deben existir en la gestión de fallos en un sistema altamente disponible y que permiten obtener notificaciones de procesos que se estén llevando a cabo y de sus posibles fallos, proporcionando así la mejor solución posible para que no se vea afectado la disponibilidad del servicio.

El documento aporta valiosos conceptos relacionados con la alta disponibilidad y gestión de fallos en sistemas altamente disponibles, aspectos de gran importancia y altamente relacionados con el tema a desarrollar en la presente investigación. Sin embargo no se muestra ningún caso de estudio en el cual se verifique el correcto funcionamiento de las etapas proporcionadas.

Capítulo 3

Caracterización de servicios convergentes en el ambiente NGSDP.

3. Introducción

En el presente capítulo se realiza la conceptualización acerca de los servicios de telecomunicaciones y de los servicios Web, después se explica el fenómeno de la convergencia en el sector de las telecomunicaciones; esta explicación se realiza tanto a nivel internacional, como a nivel nacional y finalmente se explican claramente los aspectos que facilitan el despliegue de servicios convergentes sobre las plataformas NGSDP.

3.1. Servicios de Telecomunicaciones

El mercado de servicios de telecomunicaciones es uno de los más dinámicos a nivel mundial y desempeña un papel central en el comercio de servicios, contribuyendo al desarrollo de este sector.

Así, los servicios de telecomunicaciones son aquellos que se ofrecen a terceros o al público en general, para que por medio de un circuito o una red de telecomunicaciones un usuario pueda establecer comunicación desde un punto de la red a cualquier otro punto de la misma o a otras redes de telecomunicaciones [34].

Según la Recomendación F.500 de la Unión Internacional de Telecomunicaciones (ITU, International Telecommunication Union) [35], los servicios de telecomunicación son la utilidad o provecho que resulta de la prestación, uso y aplicación del conjunto de capacidades y facilidades de la telecomunicación, destinados a satisfacer intereses y necesidades de los usuarios y al mejoramiento de la calidad de vida de la población.

En Colombia, el artículo 33 de la Ley 80 de 1993 [36] entiende por servicios de telecomunicaciones: “aquellos que son prestados por personas jurídicas, públicas o privadas, debidamente constituidas en Colombia, con o sin ánimo de lucro, con el fin de satisfacer necesidades específicas de telecomunicaciones a terceros, dentro del territorio nacional o en conexión con el exterior.” Según el Decreto Ley 1900 de 1990 [37], en Colombia se clasifican los servicios de Telecomunicaciones de la siguiente manera: Servicios Básicos (comprenden los servicios portadores y los teleservicios), de Difusión, Telemáticos y de Valor Agregado, Auxiliares de Ayuda y Especiales (

Tabla **2).**

Servicios portadores: aquellos que proporcionan la capacidad necesaria para la transmisión de señales entre dos o más puntos definidos de la red de telecomunicaciones. Comprenden los servicios que se hacen a través de redes conmutadas de circuitos o de paquetes y los que se hacen

a través de redes no conmutadas. Forman parte de éstos, entre otros, los servicios de arrendamiento de pares aislados y de circuitos dedicados.

Categoría	Servicio
Servicios Básicos	Telefonía Fija
	Telefonía Móvil
	Telefonía Móvil- Celular
	Telegrafía y Télex
Servicios de Difusión	Televisión
	Radio
Servicios Telemáticos	Telefax
	Videotex
	Datafax
Servicios de Valor Agregado	Correo Electrónico
	Transferencia Electrónica de Fondos
Servicios Auxiliares	Meteorología
	Navegación Aérea o Marítima
Servicios Especiales	Radioaficionados
	Investigación

Tabla 2. Clasificación de los servicios de Telecomunicaciones – Tomado del Decreto Ley 1900 de 1990

Teleservicios: aquellos que proporcionan en sí mismos la capacidad completa para la comunicación entre usuarios, incluidas las funciones del equipo terminal. Forman parte de estos, entre otros, los servicios de telefonía tanto fija como móvil y móvil-celular, la telegrafía y el télex.

Servicios de difusión: aquellos en los que la comunicación se realiza en un solo sentido a varios puntos de recepción en forma simultánea. Forman parte de éstos, entre otros, las radiodifusiones sonora y de televisión.

Servicios Telemáticos: aquellos que utilizando como soporte servicios básicos, permiten el intercambio de información entre terminales con protocolos establecidos para sistemas de interconexión abiertos. Forman parte de estos, entre otros, los de telefax, publifax, teletex, videotex y datafax.

Servicios de Valor Agregado: aquellos servicios que utilizan como soporte servicios básicos, telemáticos y de difusión, o cualquier combinación de estos, que proporcionen la capacidad completa para el envío o intercambio de información, agregando otras facilidades diferenciables del servicio soporte o satisfaciendo nuevas necesidades específicas de telecomunicaciones, independientemente de la tecnología que utilice. Es el caso de las señales de video, audio, voz, texto y otras, que usan como soporte las redes de telecomunicaciones del estado u otras, las redes de servicios básicos de telefonía móvil, telefonía pública básica conmutada y servicios portadores. Para que el servicio de valor agregado se diferencie del servicio básico, es necesario que el usuario de aquél perciba de manera directa alguna facilidad agregada a la simple telecomunicación, que le proporcione beneficios de telecomunicaciones adicionales, independientemente de la tecnología

o el terminal utilizado; o que el operador de servicios de valor agregado efectúe procesos lógicos sobre la información que posibiliten una mejora, adición o cambio al contenido de la información de manera tal que genere un cambio neto de la misma independientemente del terminal utilizado. Este cambio a su vez, debe generar un beneficio inmediato y directo, que debe ser percibido por el usuario del servicio [38] [39].

3.2. Servicios Web

Según el W3C (World Wide Web Consortium) un Servicio Web es un sistema de software diseñado para apoyar la interacción máquina a máquina sobre una red. Tiene una interfaz descrita en un formato procesable por máquina llamado, WSDL (*Web Services Description Language*). Otros sistemas interactúan con el servicio Web de una manera prescrita usando mensajes SOAP (*Simple Object Access Protocol*), típicamente transmitido a través de HTTP (Hypertext Transfer Protocol) con una serialización XML (*eXtensible Markup Language*) en conjunción con otras normas relacionadas con la Web [40] . En general un servicio Web es cualquier aplicación que corre en un equipo local, solo que la información necesaria para llevar a cabo una tarea específica es enviada a un servidor y el resultado de esa tarea, es devuelto al usuario, en forma de contenido Web.

Los servicios Web han tenido gran éxito y acogida debido a sus grandes beneficios, como: i) interoperabilidad entre aplicaciones software y plataformas de distintos fabricantes, debido a la utilización de protocolos estándar y abiertos, ii) al apoyarse en el protocolo HTTP, los servicios Web pueden aprovechar los sistemas de seguridad firewall sin necesidad de cambiar las reglas de filtrado, iii) reduce las limitaciones geográficas debido a que permite que servicios y software de diferentes compañías, ubicadas en diferentes lugares geográficos puedan operar conjuntamente para ofrecer servicios integrados.

Desde el surgimiento de la era Web se pueden identificar tres generaciones, las cuales han afectado los servicios Web prestados. Estas tres generaciones han sido denominadas por el W3C como: i) Web 1.0, ii) Web 2.0 y iii) Web 3.0.

Web 1.0

Es la forma más básica que existe para transmitir información vía Web (1991-2003). Es una Web estática de solo lectura, no permite la participación activa del usuario final, lo que significa que la información se encuentra centralizada y solo es actualizada por el administrador del sitio Web; su principal propósito es difundir información; el diseño y producción están a cargo de las personas encargadas de transmitir la información. El formato utilizado para trasmitir la información es HTML.

Web 2.0

El termino Web 2.0, es un concepto difuso. Facilita la interoperabilidad entre las diferentes aplicaciones, el diseño es centrado en el usuario y le permite una participación activa a través de opciones que le dan voz propia en la web, permitiéndole administrar sus propios contenidos, opinar sobre otros, enviar y recibir información con otros usuarios de su mismo estatus o instituciones que así lo permitan [41] [42].

Un buen resumen de lo que significa el término es obra de Wade Roush[9], quien argumenta que el término web 2.0 se refiere a tres cosas: i) nuevos mecanismos de relación y comunicación entre las personas, utilizando las tecnologías de redes sociales (con servicios como Facebook, YouTube, Digg o Wikipedia), ii) utilización de estándares web para la creación de servicios distribuidos en Internet ("mashups") iii) y la mejora en las interfaces de las páginas web hasta llegar a imitar casi a la perfección la experiencia de usuario de las aplicaciones clásicas que se ejecutan en un computador [41].

En la Tabla 3 se pueden observar claramente las diferencias entre el concepto de la Web 1.0 y Web 2.0

Web 1.0		Web 2.0
DoubleClick	-->	Google AdSense
Ofoto	-->	Flickr
Akamai	-->	BitTorrent
mp3.com	-->	Napster
Britannica Online	-->	Wikipedia
personal websites	-->	blogging
evite	-->	upcoming.org and EVDB
domain name speculation	-->	search engine optimization
page views	-->	cost per click
screen scraping	-->	web services
publishing	-->	participation
content management systems	-->	wikis
directories (taxonomy)	-->	tagging ("folksonomy")
stickiness	-->	syndication

Tabla 3. Diferencias entre Web 1.0 y Web 2.0 [41]

Web 3.0

El paso a seguir en la evolución de la Web, todavía no cuenta con un norte definido, algunos especialistas apuntan hacia la Web Semántica como el siguiente gran hito en la evolución de la Web, sin embargo otros apuntan hacia otras mejoras en la interacción de la Web y a la combinación con la inteligencia artificial y también aparecen términos como el “Web 3D”, “Data Web” y los “microformatos”. A continuación se muestran las características de algunas de estas tendencias [43] [40].

- ***La Web Semántica***

El concepto de web semántica se está desarrollando bajo el mando de Tim Berners-Lee[10]. Se trata de dotar de significado a las páginas Web, de ahí el nombre de Web semántica, ya que al día de hoy los contenidos en las páginas Web no son entendibles para los computadores, solo tienen sentido para las personas; el concepto busca añadir información adicional a la estructura de la página Web, de tal forma que esta pueda ser entendida por los computadores, los cuales

[9] Wade Roush: jefe y editor de Xconomy, empresa relacionada con la innovación y la vida digital.
[10] Tim Berners-Lee: fundador de la 3WC y considerado el padre de la Web.

mediante técnicas de inteligencia artificial serían capaces de emular y mejorar la obtención de conocimiento, algo que hasta el día de hoy solo puede ser hecho por los humanos.

- ***Los microformatos***

Los microformatos surgen del trabajo de la comunidad de desarrolladores de Technorati[11]. Su objetivo es estandarizar un conjunto de formatos en los cuales se almacenaría conocimiento básico, como la información de contacto de una persona (microformato hCard), una cita (microformato hCalendar), una opinión (microformato hReview), una relación en una red social (microformato XFN) y así hasta un total de 9 especificaciones concluidas y 11 en proceso de definición. La principal limitación es que cada tipo de significado requiere de la definición de un microformato específico. A cambio ya es posible utilizarlos, como así lo hace un conjunto reducido de sitios web.

3.3. Servicio Convergente

El termino convergencia es frecuentemente usado para describir la visión futura del desarrollo de diferentes áreas orientadas hacia un objetivo común. Sin embargo en el entorno de los proveedores de telecomunicaciones, este término tiene cinco diferentes orientaciones que están en evolución, debido a la creciente demanda por parte de los usuarios. Estas áreas son [44]:

Convergencia en la industria: convergencia entre los operadores de telecomunicaciones con el mundo de IT y con los proveedores de contenido multimedia (terceras partes).

Convergencia a nivel de red: convergencia entre los tipos de red existentes: fijas, móviles y de banda ancha (se tendrá acceso a las redes y al núcleo (core) de estas, y en el futuro se convertirán hacia All-IP).

Convergencia de servicios: se refiere a la prestación homogénea de servicios multimedia para los consumidores y usuarios empresariales; se tendrá la facilidad para transferir voz, video, e-mail y servicios de chat a través de múltiples dispositivos como: terminales móviles, PCs y televisores.

Convergencia en pagos: manejo común para todos los usuarios (post-pago o pre-pago), los cuales son considerados de igual forma por el sistema de facturación, sin embargo cuenta con diferentes reglas de negocio aplicadas al control de crédito.

Convergencia en datos del subscriptor: la información de los usuarios se encuentra almacenada en un solo lugar, simplificando la autentificación, autorización y la facturación.

Como se puede apreciar de la anterior información, en la actualidad el deseo por nuevas aplicaciones y servicios han introducido necesidades que originalmente no fueron tenidas en cuenta en el diseño de la primera generación de redes de paquetes, además aspectos como la competencia entre los operadores de telecomunicaciones, el auge del tráfico digital (la utilización a gran escala de la Internet), la fuerte demanda de nuevos servicios multimedia que agregan valor

[11] Technorati: es un motor de búsqueda de Internet para buscar blogs, que compite con Google, Yahoo!, PubSub e IceRocket.

añadido a los servicios, entre otros, ha motivado la introducción de un nuevo concepto que tenga en cuenta las nuevas necesidades de la industria de las telecomunicaciones. Este concepto es denominado NGN, el cual según la recomendación Y.2001 de la ITU-T [45], está basado en paquetes para suministrar servicios de telecomunicaciones en múltiples tecnologías de acceso de banda ancha apropiadas para garantizar calidad en los servicios ofrecidos.

Entre las múltiples ventajas de las NGN, se encuentra el acceso de los usuarios a redes y proveedores de servicios y/o a los servicios de su elección, también soporta movilidad generalizada coherente y ubicua de servicios a los usuarios. Por lo tanto la evolución desde las redes tradicionales de telecomunicaciones hacia las NGN se fundamentan en la convergencia de aplicaciones y servicios soportados y transportados sobre diferentes redes de acceso y núcleos hacia una red unificada con la capacidad de soportar cualquier tipo de servicio.

Teniendo claro, por que surgió el concepto de las NGN y que este agrupa todas las perspectivas de convergencia de la telecomunicaciones, ahora se debe aclarar el concepto de servicio convergente. Cuáles son sus alcances, limitaciones, tipos y las características con las que un servicio de esta clase debe cumplir.

3.3.1. Servicio Convergente en el ámbito Internacional

En las NGN se pueden identificar tres clases de servicios convergentes, estos son: servicios residenciales, servicios empresariales y servicios de movilidad [46]. La tendencia general de la industria está orientada a que cualquier clase de servicio pueda ser entregado en cualquier tipo de pantalla, además de disfrutar de la personalización y las aplicaciones multimedia integradas para negocios y entretenimiento. Una característica esencial de las NGN es la capacidad para suministrar gran variedad de servicios (voz, video, audio y datos) basados en sesiones de usuario SIP, sin importar el tipo de transporte (unidifusión, multidifusión y difusión), por lo cual en las NGN es posible la utilización indistinta de las tecnologías alambradas e inalámbricas para la entrega de servicios y además puede emplearse de manera coherente en cualquier instante o lugar a través de diferentes entornos que emplean equipos de terminales convergentes (terminales capaces de aceptar todos los servicios). De esta forma será posible la prestación del servicio de telefonía portable, es decir que este servicio sería llevado al dispositivo móvil, fijo o softphone, según sean las necesidades del cliente; también será posible que los proveedores ofrezcan servicios convergentes de video que serán proporcionados a dispositivos HDTV (*High Definition TV*), PC, teléfono celular, PDA (Personal Digital Assistant) o dispositivos inalámbricos, basados únicamente en las preferencias de los clientes. Esta tendencia lleva a la sustitución de los medios tradicionales, es decir se pasa de la telefonía tradicional a VoIP y a la entrega de servicios de video por los medios de IPTV y VoD-IP (Video on Demand-IP).

En la Figura 6 se presenta un esquema de la evolución de los servicios en varias líneas de las tecnologías de la comunicación, como IPTV, telefonía y servicios multimedia relacionados con las redes fijas, móviles, Internet y servicios de negocio.

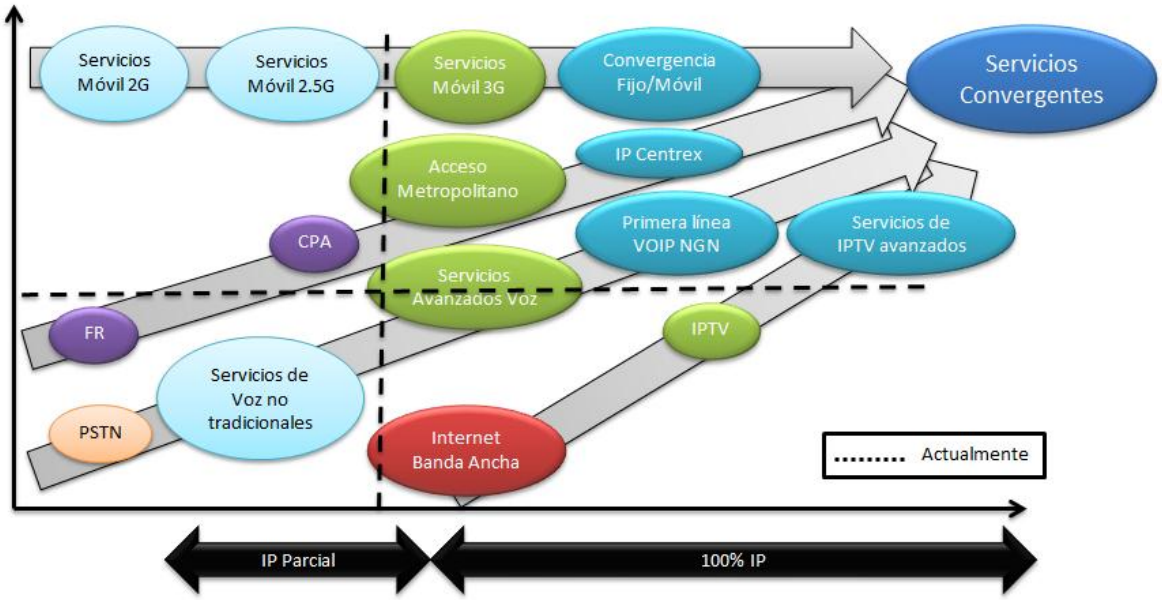

Figura 6. Convergencia de Servicios [46]

Los servicios que debe soportar la NGN han sido descritos por ITU-T, FGNGN, WG1, SR y se presentan a continuación [46].

- ***Servicios interactivos:***

• Servicios de conversación en tiempo real.

• Servicios interactivos de multimedia punto a punto, incluyendo voz en tiempo real interactiva, video y otros medios.

• Servicios de comunicación colaborativos (servicios de conferencia multimedia con intercambio de archivos y aplicaciones, e-learning, juegos, etc.).

• Push to talk sobre NGN (PoN).

• Mensajería Instantánea (IM) y Servicios de mensajería (SMS, MMS, etc.).

• Mensajería en grupo.

• Servicios existentes sobre PSTN/ISDN (emulación y simulación de PSTN/ISDN).

• Servicios de comunicación de datos (transferencias de archivos, fax, mail electrónico, etc.).

• Aplicaciones en línea (Ventas en línea, comercio electrónico, etc.).

•Servicios de activación por voz.

- ***Servicios no interactivos***

• Servicios de entrada de contenido (radio y video streaming, video y música bajo demanda, distribución de canales de TV digital, distribución de información financiera, distribución de imágenes médicas y profesionales, publicidad electrónica).

• Servicios en redes de sensores.

• Servicios "Push"[12].

• Servicios de acciones de control remotas, tales como aplicaciones de control de hogar, telemetría, alarmas, etc.

• Servicios de Broadcast/Multicast.

• Administración de dispositivos sobre la red.

- **Servicios Mixtos**

• Servicios de VPN (*Virtual Private Network*).

• Servicios administrados para empresas (IP Centrex, etc.).

• Servicios de información (información de tiquetes para el cine, estado del tráfico, servicios avanzados de "push", etc.).

• Servicios generales de presencia y notificación (visualización de contactos de un usuario, su estado actual y cualquier servicio relacionado con notificaciones).

• Servicios soportados en OSA (*Open Services Architecture*) para 3GPP Release 6 y 3GPP2 (*2 – 3G/CDMA2000*).

- ***Servicios de Red***

• Servicios Básicos de Transporte (BTS13): proveen conectividad básica punto a punto, punto-multipunto, multipunto-multipunto. En cuanto a los aspectos básicos del transporte incluyen servicios de mejor esfuerzo, seguridad limitada, etc.

• Servicios de transporte mejorado (ETS14): proveen los servicios de conectividad básicos, pero adicionalmente garantizan servicios diferenciados como QoS, nivel de seguridad avanzada y acceso a VPN.

- ***Servicios regulados***

• Servicios de telecomunicaciones de emergencia (ciudadano a autoridades, entre autoridades, y autoridades a ciudadanos).

• Servicios de intercepción legal.

• Servicios de emisión de alerta de emergencia.

3.3.2. Servicio Convergente en el ámbito Colombiano

Hasta el momento en Colombia no existe una definición clara del régimen regulatorio hacia la convergencia, pero existen algunos artículos y decretos que lo nombran y que permiten tener un poco de conocimiento sobre este. Se busca en un futuro definir un régimen regulatorio

[12] El término "push services" describe el contenido que es enviado desde un servidor, directamente a una terminal de subscriptor.

convergente, acorde a las nuevas necesidades y hábitos de los usuarios Colombianos y al desarrollo de nuevos medios [47].

Según el decreto 2870 de 2007 [39], en el artículo 1, "se establece un marco reglamentario que permita la convergencia en los servicios públicos de telecomunicaciones y en las redes de telecomunicaciones del estado, asegurar el acceso y uso de las redes y servicios a todos los habitantes del territorio, así como promover la competencia entre los diferentes operadores".

Anteriormente, los operadores Colombianos ofrecían a sus usuarios simplemente el servicio de telefonía local y de larga distancia; un caso más específico se encuentra en la desaparecida empresa "TELECOM" la cual hasta 1998 era un monopolio de este tipo de servicios. Con la llegada de la telefonía móvil, la telefonía de larga distancia tuvo una gran caída y se sustituye por la telefonía fijo-móvil, haciendo que los operadores de telecomunicaciones empiecen a diseñar estrategias que le permitan ofrecer paquetes de servicios de telecomunicaciones para evitar su extinción del mercado. Es por esto que se busca la convergencia comercial, los operadores de telecomunicaciones ofrecen a sus usuarios paquetes, combos, etc. que contengan una gran variedad de servicios de telefonía, Internet, televisión, servicios móviles, entre otros [48]. Actualmente los operadores más exitosos son los que han podido reemplazar los ingresos decrecientes de los servicios tradicionales con paquetes de varios servicios.

Según [49], la convergencia es la posibilidad tecnológica de provisión sobre múltiples redes tanto de los servicios tradicionales de comunicaciones así como de sus innovaciones en los campos de voz, datos, sonidos e imágenes.

Las dimensiones en las que se puede manifestar la convergencia tecnológica son [50]:

- Convergencia de Servicios
- Convergencia de Equipos Terminales
- Convergencia de Redes o Medios de Transmisión
- Convergencia de Mercados

Ante la falta de regulaciones claras en cuanto a la definición y alcance de lo que se significan los servicios convergentes en el ámbito Colombiano (no se hace claridad si se adopta la definición internacional), en el presente trabajo de grado se adoptara la definición de servicio convergente internacional, presentada en la sección 3.3.1.

3.4. Caracterización del despliegue de servicios en el ambiente NGSDP

En la anterior sección se explicó claramente el término de convergencia y las múltiples ramificaciones que este posee. Debido a que el rápido despliegue de los nuevos servicios convergentes es un factor crítico para los operadores de telecomunicaciones a fin de diferenciarse de la competencia; a continuación se explica como la SDP de tercera generación facilita dicho aspecto.

3.4.1. Despliegue de servicios en una NGSDP

Con el fin de facilitar el despliegue de nuevos servicios convergentes creados a partir de componentes del mundo de las telecomunicaciones, mundo IT y de la Web, la SDP 2.0 se muestra como el entorno más apropiado. Entre las características que ésta posee y por la cual es posicionada de esta forma, se encuentran: i) flexibilidad, debido a que no tiene restricción a la innovación, es decir, permite nuevos proveedores de servicio y nuevos servicios, ii) gestión, posibilita el acceso a las capacidades de red del operador de telecomunicaciones pero de forma controlada, iii) integración de los sistemas OSS y BSS, entre otros. En la Figura 7 se puede apreciar la implementación de servicios sin la utilización de la plataforma SDP, mientras que en la Figura 8 se aprecia claramente el gran cambio que se tienen en la implementación de servicios mediante la utilización de las plataformas SDP [51].

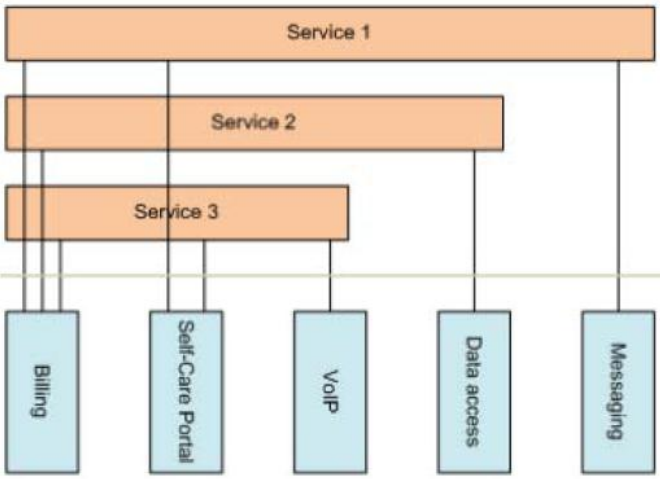

Figura 7. Implementación de servicios sin SDP [51]

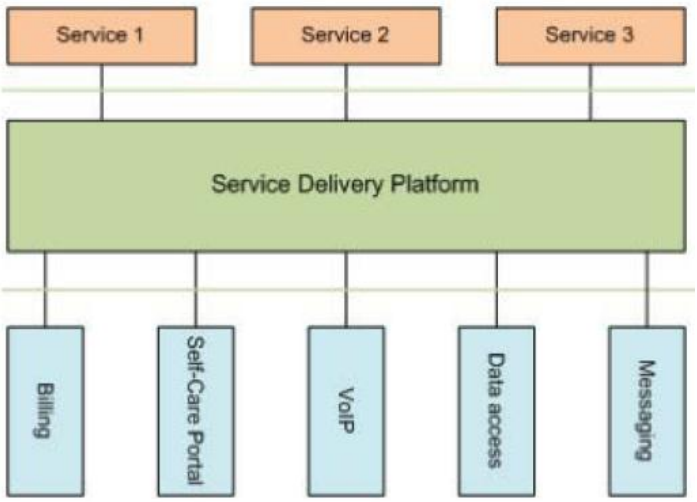

Figura 8. Implementación de servicios con SDP [51]

Adicionalmente aspectos cruciales al momento de ofrecer nuevos servicios, como: "*time to market*", factor de innovación y velocidad en términos del desarrollo y soporte, son mejorados por la plataforma mediante la implantación de las plataformas NGSDP ; adicionalmente a las ventajas técnicas mostradas la NGSDP también trae considerables beneficios económicos a los operadores de telecomunicaciones, dichos beneficios pueden observarse entre otros en el desarrollo de servicios basados en SDP (Figura 9).

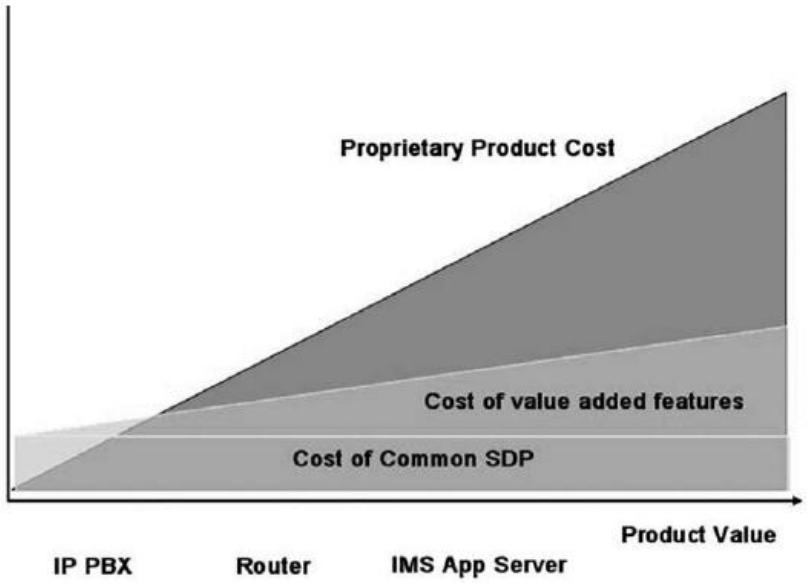

Figura 9. Análisis de costos asociados al desarrollo de servicios [46]

A continuación se muestran detalladamente las características que facilitan el despliegue de servicios convergentes en las plataformas SDP 2.0.

3.4.1.1. Arquitectura de integración

Históricamente la creación de servicios de telecomunicaciones implicaba la construcción de todo un sistema para su soporte, teniendo en la actualidad dificultades como [46] [44] [52] [53]:

Información de usuarios no unificada: la información de los usuarios se encuentra dispersa entre los diferentes sistemas de telecomunicaciones.

Datos no unificados: datos de facturación, consumo, entre otros; se encuentran dispersos y se requieren sistemas de gestión para todos estos en cada sistema de telecomunicaciones (por ejemplo: wireless, PSTN, cable. etc.).

Tiempo de venta (Time to Market): el desarrollo de un nuevo servicio de telecomunicaciones en término promedio toma de doce a dieciocho meses.

Inflexibilidad hacia nuevos modelos de negocio: se requieren múltiples procesos de negocio que deben ser probados y refinados.

Silos de sistemas: cada servicio de telecomunicaciones posee su propia infraestructura hardware, sistemas de gestión y facturación.

Entornos propietarios: implica grandes esfuerzos y altos gastos en el desarrollo y despliegue de nuevos servicios.

La tendencia actual es la unificación de las arquitecturas, pasando de un esquema vertical a uno horizontal (NGN), integrar los sistemas OSS/BSS y ofrecer los servicios de la Web sobre esta arquitectura (Figura 10).

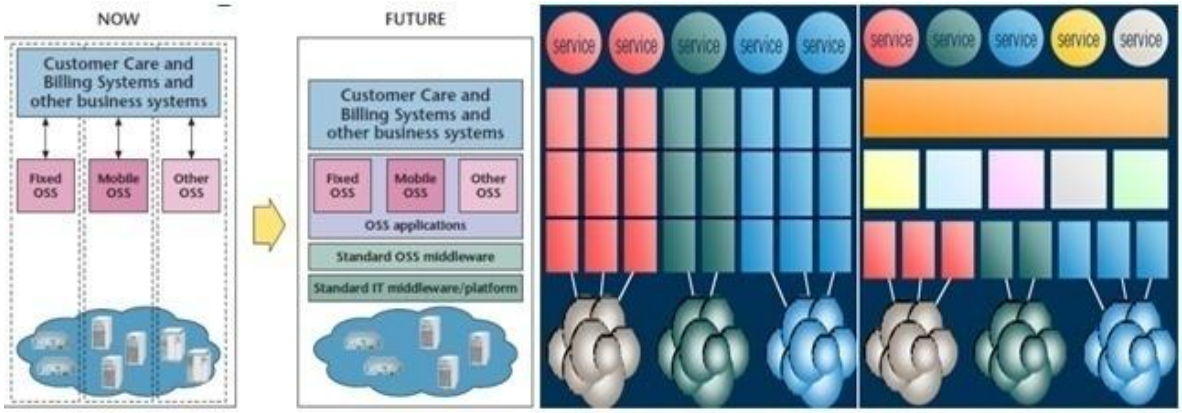

Figura 10. Cambio en la infraestructura de los operadores [54]

En este punto toma gran importancia el uso de las NGSDP en las arquitecturas de los operadores de telecomunicaciones, esto debido a las características con las que fue diseñada la SDP de tercera generación [11] [12] [54] [55], como:

Capacidades IMS: la SDP de tercera generación soporta los protocolos SIP y Diameter, capacidades de servicio (p.e: presencia), así como también estándares e interfaces como: ISC, Ro, Rh, etc., y creación de servicios SIP.

Principios SOA: la SDP 2.0 es construida alrededor de los principios SOA, lo cual facilita los procesos de negocio. Estos principios son [31]:

i) **Flexibilidad de negocios**: se mejora la flexibilidad en cuanto a los cambios necesarios según las necesidades del negocio.

ii) **Mejores procesos de negocio:** con SOA los bloques de construcción son servicios de negocio.

iii) **Fácil integración:** mediante la definición clara de interfaces, se generan piezas modulares que facilitan su integración.

iv) **Reúso de activos:** es posible la reutilización de componentes, facilitando de esta forma la construcción de nuevos servicios.

v) **Reducción de riesgos:** se aumenta la calidad, se incrementa el desarrollo y se mejoran sus tiempos.

Convergencia de redes: facilita la convergencia a nivel de servicio y migración de servicios hacia redes All-IP. Soporta la integración entre las diferentes redes existentes y las redes de nueva generación (2,5G, 3G móvil, PSTN, IMS). El acceso a las capacidades de red se hace posible mediante un conjunto de habilitadores de servicio, lo cual permite el despliegue de un servicio sobre las diferentes redes [54] [55].

Gestión: provee un conjunto común de capacidades de gestión de servicio, incluyendo la gestión del ciclo de vida, desempeño, operación, mantenimiento, gestión de procesos de negocio para el aprovisionamiento de servicios, etc. (utilizando mecanismos SOA) [54] [55].

De acuerdo a lo anterior, se puede concluir que la SDP 2.0, actúa como un puente entre el mundo de las telecomunicaciones y la Web, posibilitando de esta forma su convergencia (Figura 11) [11] [54] [55].

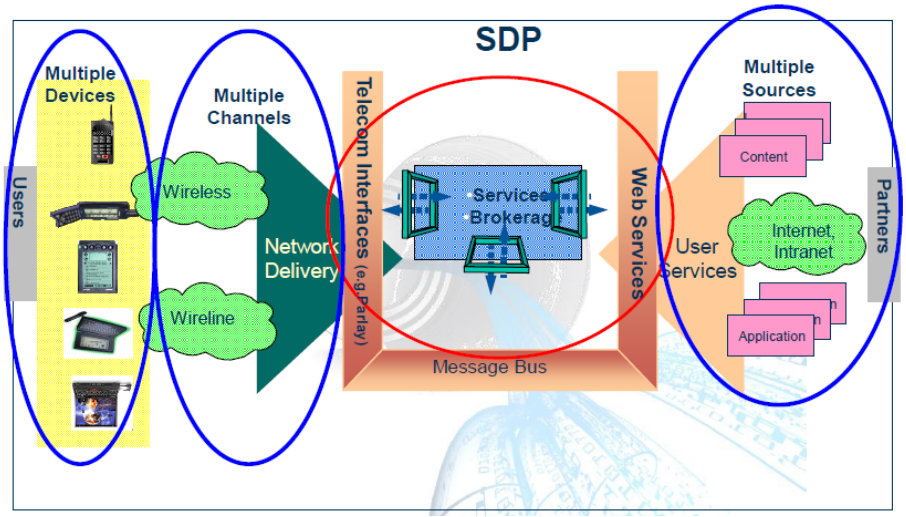

Figura 11. SDP 2.0 puente entre el mundo de las telecomunicaciones y la Web [54]

3.4.1.2. Habilitadores de servicio

En la arquitectura en capas de las NGSDP, los habilitadores de servicio son de gran ayuda en el rápido despliegue de servicios. En la capa de abstracción de red los servicios de telecomunicaciones son abstraídos. Esta operación es posible debido a que la SDP de tercera generación reconoce estándares y protocolos propietarios como SS7, SIP, Diameter, MM7, SMTP, HTTP, etc., permitiendo de esta manera la provisión de servicios del core de las telecomunicaciones, a las capas superiores, para la reutilización de sus funcionalidades en aplicaciones más complejas, de forma rápida y eficaz (Time to Market), permitiendo de esta forma la reducción en los costos de inversión y de operación. En la capa de habilitadores de servicio y servicios de telecomunicaciones se permite la implementación de habilitadores de servicio de alto nivel como: conferencia, mensajería, presencia, distribución de llamadas, administración de identidad, administración de contenido y de medios, etc., [12] [44] [56] .

Todos los anteriores habilitadores de servicio son expuestos a las terceras partes por medio de servicios web y herramientas estándar (Parlay X), facilitando de esta manera la construcción y posterior despliegue de nuevos servicios. Esta exposición hacia las terceras partes se realiza en la capa de exposición de servicio, poniendo de esta forma a disposición de los desarrolladores de servicio todas las funcionalidades de los servicios de telecomunicaciones mediante dichas interfaces. En esta capa también se tiene funcionalidades para controlar el acceso de las terceras partes como: autorización, autenticación y facturación así como también políticas para la gestión y la ejecución.

En consecuencia, se facilita la creación y posterior despliegue de servicios convergentes en las plataformas SDP 2.0, además se garantiza la portabilidad de los servicios y se facilita la migración de estos hacia redes All-IP.

Otro aspecto importante para el despliegue de servicios convergentes en la arquitectura de la SDP 2.0, se encuentra en la capa de orquestación y gestión de servicio, donde se encuentran repositorios de servicios, de perfiles de usuarios y la gestión de la identidad compartida por todos los servicios desplegados en la SDP.

Algunos de los habilitadores de servicio que se encuentran en la plataforma NGSDP son: presencia, localización, mensajería, control de llamada, control de medios, gestión de la movilidad, gestión de sesión, gestión de políticas, facturación, evaluación de la calidad (rating), contabilidad.

En la Figura 12 se puede apreciar gráficamente como se realiza la abstracción de los servicios de telecomunicaciones y como son expuestos hacia las terceras partes.

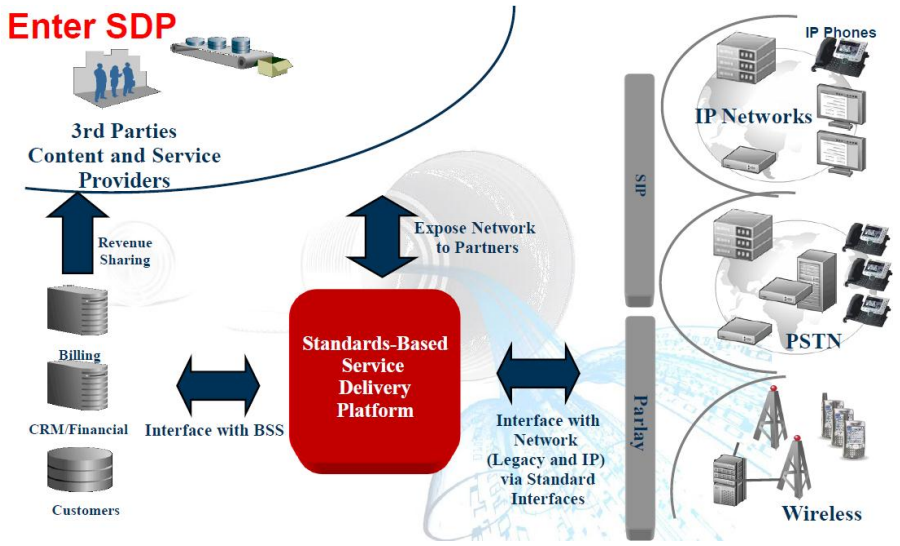

Figura 12. Abstracción y exposición de habilitadores de servicio hacia las terceras partes [54]

3.4.1.3. SEE (Service Execution Environment)

La capa de ejecución y creación de servicio está conformada por uno o más SEE (Service Execution Environment), los cuales son de gran importancia para el despliegue de servicios en esta plataforma [12]. Entre los entornos de ejecución aplicables a esta plataforma se encuentra el JAIN SLEE (Java *Service Logic Execution Environment*), Parlay, SIP Servlets, entre otros. Sin embargo y como se muestra en [31] JAIN SLEE se posiciona como el entorno de ejecución más adecuado para las telecomunicaciones convergentes, esto gracias a que su arquitectura aporta beneficios a la migración de redes y al despliegue de servicios [31] [52] [30]:

Entre las características que poseen los entornos de ejecución JAIN SLEE se encuentran:

Portabilidad de servicios: al implementar la filosofía WORA (Write Once, Run Anywhere) del mundo Java. Los componentes puede ser desarrollados y después ser desplegados en plataformas JAIN SLEE de diferentes proveedores sin la necesidad de recompilar o modificar el código fuente [31] [52].

Desarrollo de servicios: los JSLEE facilitan el rápido desarrollo de servicios y a costos eficientes, debido a que los componentes principales existentes pueden ser reutilizados para la creación de los nuevos servicios por parte de los operadores de telecomunicaciones o por parte de los proveedores de servicios. Por otro lado la barrera de entrada al mercado de desarrollo de servicio de esta clase de reducirá debido a que los desarrolladores tendrán la posibilidad de lograr considerables ingresos y con bajos costos en los entornos de desarrollo [31] [52].

Independencia de red: la abstracción de las redes posibilita que las redes tradicionales sean usadas para creación y desarrollo de nuevos servicios en cualquier tipo de red y su posterior migración hacia las redes ALL-IP, facilitando de esta forma portabilidad en los servicios [31] [52].

Extensibilidad: los JSLEE son entornos que puede ser ampliados con nuevos servicios y funciones y pueden ser integrados con las redes actuales y futuras [31] [52].

Integración: facilita la integración de los sistemas de facturación, gestión de servicios y de los sistemas OSS [31] [52].

Soporte para Servicios de Voz y Servicios Web: JAIN SLEE permite la interoperabilidad con J2EE, permitiendo el desarrollo de soluciones basadas en servicios convergentes de voz, datos y servicios Web [31] [52].

Capítulo 4

Lineamientos para alta disponibilidad en el contexto de una NGSDP

En el presente capítulo se conforma una base inicial de criterios asociados a la alta disponibilidad; estos criterios son definidos por algunas de las empresas con mayor reconocimiento a nivel mundial[13]. Dicha base de criterios es filtrada, a fin de seleccionar los criterios técnicos con los que una plataforma NGSDP altamente disponible debe cumplir. Finalmente son definidos los lineamientos técnicos que permiten cumplir con cada uno de los criterios seleccionados; además se plantean una serie de preguntas generales para cada criterio, esto con el fin de identificar o recomendar cuál o cuáles son los lineamientos más convenientes de implementar para cada operador.

4.1. Criterios técnicos asociados a la alta disponibilidad

A continuación se presentan los criterios técnicos relacionados con alta disponibilidad, los cuales han sido definidos por diferentes empresas prestadoras de servicio según sus necesidades. A partir de estos se realiza un filtrado y selección de los criterios más importantes para un operador de telecomunicaciones que desea implantar una NGSDP en su modelo de negocio.

Cabe la pena resaltar que también fueron consultadas empresas directamente relacionadas con el modelo mundo de las telecomunicaciones, como: Ericsson, Siemens, entre otros. Sin embargo no se tuvieron en cuenta para conformar la base inicial de criterios, debido a que las soluciones planteadas por estas son orientadas hacia la adquisición de sus productos, lo cual está directamente relacionado con la adquisición de productos hardware. También se debe mencionar que criterios directamente relacionados con la disponibilidad software como: protección de memoria, detección y protección de overload, heartbeatieng, monitoreo de procesos, detección de fugas de memoria, time outs, run time diagnostics, etc., no fueron tenidos en cuenta debido a que no se contaba con las herramientas necesarias para su correcta implementación y evaluación, y a que se considera que los criterios expuestos a continuación se consideran los más relevantes a fin de alcanzar alta disponibilidad en el contexto de una NGSDP.

4.1.1. Cisco Systems

Cisco Systems es una empresa multinacional, principalmente dedicada a la fabricación, venta, mantenimiento y consultoría de equipos de telecomunicaciones, actualmente es el líder mundial en soluciones de red e infraestructuras para Internet [57].

Cisco define la alta disponibilidad de servicio mediante la siguiente serie de criterios [6]:

[13] Se consultaron empresas

- **Confiabilidad, tolerancia a fallos de dispositivos de red:** fiabilidad de Hardware y Software para automáticamente identificar y superar fallos.
- **Redundancia de dispositivos y enlaces:** todos los dispositivos, los módulos dentro de los dispositivos y las conexiones pueden ser redundantes. Asegurando de esta forma mayor disponibilidad.
- **Balanceo de cargas:** permite a un dispositivo tener múltiples caminos para un determinado destino, asegurando que no se sobrecargue un solo camino de información, y de esta forma evitar posibles retardos y pérdidas de información.
- **Resiliencia de tecnologías de red:** tecnología que asegura una rápida recuperación ante las fallas en cualquier dispositivo o enlace, permitiendo que los momentos de indisponibilidad en un sistema sean pequeños o nulos.
- **Diseño de red**: topologías y configuraciones de red bien definidas para asegurar que no exista ni un punto de fallo.
- **Buenas prácticas:** procedimientos documentados para desarrollar y mantener una infraestructura de red robusta.

4.1.2. OpenCloude

OpenCloude es una empresa formada en el año 2000, que provee a la industria de las telecomunicaciones servidores de aplicaciones en tiempo real, para el desarrollo y despliegue ágil de aplicaciones, y una gestión eficiente de los servicios de comunicación entre las personas a través de redes, tecnologías actuales y de nueva generación. OpenCloude ofrece servicios de consultoría especializados, implementación, capacitación y servicios de soporte para clientes en todo el mundo [58].

OpenCloude define la alta disponibilidad mediante el siguiente conjunto de criterios técnicos [30]:

- **Rendimiento, funcionamiento, desempeño:** se refiere al desempeño en términos de tiempo de respuesta y el rendimiento de una red, equipo, servicio u otro.
- **Fiabilidad y disponibilidad:** fiabilidad se refiere a la probabilidad de que un dispositivo o sistema lleve a cabo sus actividades prescritas sin fallas durante un tiempo determinado. Disponibilidad se define como el porcentaje de tiempo total que el sistema está funcionando correctamente.
- **Escalabilidad:** capacidad del sistema para manejar el crecimiento continuo de trabajo o para estar preparado para hacerse más grande sin perder calidad en los servicios ofrecidos. En general, se define como la capacidad del sistema informático de cambiar su tamaño o configuración para adaptarse a las circunstancias.
- **Contención ante fallas:** se refiere a la capacidad del sistema para reaccionar ante posibles fallas, de forma que el rendimiento del sistema no se vea afectado en caso de un siniestro.
- **Construcción en redundancia:** se refiere a la duplicación de información o de los componentes de un sistema a fin de incrementar su fiabilidad.

- **Gestión y sostenimiento[14]:** Se refieren a las actividades necesarias con el fin de lograr los objetivos planteados de manera eficiente y a la probabilidad de realizar las acciones de reparación en un tiempo determinado.

4.1.3. IBM

IBM (International Business Machines) es una empresa multinacional que fabrica y comercializa herramientas, programas y servicios relacionados con la informática. Tiene una presencia principal en prácticamente todos los segmentos relacionados con las tecnologías de la información; más de la mitad de sus ingresos vienen de sus ramas de consultoría y servicios, y no de la fabricación de equipos. Además es una firme patrocinadora del software libre [59].

IBM define la alta disponibilidad mediante la siguiente serie de criterios técnicos [29]:

- **Presupuesto:** cada solución de alta disponibilidad tiene un costo asociado y un presupuesto. El costo de la solución debe compararse con las ventajas obtenidas para cada negocio. Desde el punto de vista técnico, se puede obtener una disponibilidad constante con un tiempo de inactividad igual a cero, pero el costo de la protección ofrecida por la solución podría ser demasiado elevado.

- **Requerimientos de tiempo de funcionamiento:** hace referencia a la cantidad total de tiempo en que el sistema está disponible para las aplicaciones de usuarios finales. El valor se indica como el porcentaje del total de horas de trabajo planificadas.

  Los porcentajes de tiempo de funcionamiento y los valores de tiempo de inactividad correspondientes son:

  - ✓ Menos del 90%, tiempo de inactividad de 876 horas o más al año.
  - ✓ 90 - 95%, tiempo de inactividad de 438 a 876 horas al año.
  - ✓ 95 - 99%, tiempo de inactividad de 88 a 438 horas al año.
  - ✓ 99,1 - 99,9%, tiempo de inactividad de 8,8 a 88 horas al año.
  - ✓ 99,99%, tiempo de inactividad de aproximadamente 50 minutos al año.
  - ✓ 99,999%, tiempo de inactividad de aproximadamente 5 minutos al año.

  Generalmente, el costo por cada hora de parada se utiliza como factor determinante para los requisitos de tiempo de funcionamiento.

- **Cobertura de paradas:** se refiere a los tipos de paradas imprevistas que pueden existir en un sistema y cómo estas se pueden resolver para no afectar de forma considerable la disponibilidad del mismo. Los tipos de paradas son: reducción de la ventana de copia de seguridad, mantenimiento planificado, paradas imprevistas, siniestros de sitio.

  El impacto en las aplicaciones puede definirse de la siguiente manera:

[14] Sostenimiento o "Maintainability" se refiere a un producto que puede ser mantenido a fin de: corregir defectos, reunir nuevos requerimientos, hacer frente a cambios en el entorno.

- ✓ No es un problema. La disponibilidad de la aplicación es lo más importante.
- ✓ El rendimiento se puede ver afectado mientras la solución de disponibilidad se está entregando.
- ✓ Se admite cierta degradación del rendimiento.
- ✓ Una ligera degradación del rendimiento.
- ✓ No se percibe ningún impacto en el rendimiento.

- **Tiempo de recuperación objetivo (RTO):** es el tiempo que lleva la restauración de una parada (planificada, imprevista o siniestra) y la reanudación de las operaciones normales para una aplicación o conjunto de aplicaciones. Las distintas tecnologías de resiliencia de datos tendrán tiempo de RTO diferentes.

  Los valores posibles para el RTO son:

  - ✓ Más de 4 días.
  - ✓ De 1 a 4 días.
  - ✓ Menos de 24 horas.
  - ✓ Menos de 4 horas.
  - ✓ Menos de 1 hora.
  - ✓ Casi inmediato.

- **Punto de recuperación objetivo (RPO):** es el punto en el tiempo con respecto al fallo en el cual se necesita la conservación de datos. El proceso de restauración conserva los cambios de datos producidos en este periodo de tiempo previo al fallo o siniestro.

  Los valores de RPO son:

  - ✓ Última operación de salvar (semanalmente, diariamente, etc.).
  - ✓ Inicio del último desplazamiento (8 horas).
  - ✓ Última ruptura importante (4 horas).
  - ✓ Último lote de trabajo (de 1 hora a decenas de minutos).
  - ✓ Última transacción (de segundos a minutos).
  - ✓ Se pueden perder los cambios realizados durante la incidencia (coherencia de interrupción de la alimentación).
  - ✓ Pérdida de datos cero.

- **Requerimientos de resiliencia:** el negocio debe identificar qué es necesario proteger cuando el sistema que hospeda la aplicación sufre una parada. Los requisitos de resiliencia son el conjunto de aplicaciones, datos y entornos de sistemas que es necesario conservar en caso de una parada del sistema de producción. Estas entidades siguen estando disponibles tras una anomalía aunque el sistema que los hospeda en el momento sufra una parada.

  Las posibles opciones son:

  - ✓ No hay nada que necesite ser resiliente.
  - ✓ Datos de aplicaciones.
  - ✓ Datos de aplicaciones y del sistema.

- ✓ Programas de aplicaciones.
- ✓ Estado de aplicaciones.
- ✓ Entorno de aplicaciones.
- ✓ Conservar todas las comunicaciones y conexiones de clientes.

- **Cambio automático y manual ante fallas:** el negocio debe definir el control proporcionado a la automatización durante las paradas imprevistas. En caso de anomalía, la aplicación puede conmutar automáticamente o de forma manual al sistema de copia de seguridad, incluyendo el inicio del entorno de todas las aplicaciones.

- **Requerimientos de distancia**: la distancia entre los sistemas o la dispersión geográfica, tiene ventajas pero cuenta con varios límites físicos y prácticos. Generalmente, cuanto mayor es la distancia entre los sistemas, mayor es la protección que obtendrá ante los siniestros de toda el área. Sin embargo, esta distancia vendrá con los impactos del entorno de aplicaciones. Cuando más separados estén los sistemas, más latencia se añadirá a la transmisión de datos.

- **Número de sistemas de copia de seguridad:** las distintas tecnologías de resiliencia de datos ofrecen números diferentes de posibles sistemas de copia de seguridad y de copias de datos de aplicaciones. El número de sistemas de copia de seguridad, así como los conjuntos de datos necesarios proporcionan guías para determinar la tecnología de resiliencia de datos necesaria.

- **Acceso a una copia secundaria de datos:** se refiere a las diferentes restricciones en el acceso al conjunto de datos de copias de seguridad, las cuales tienen que ver con operaciones de salvar y las consultas o informes.

- **Rendimiento del sistema:** la implementación de la alta disponibilidad puede tener implicaciones de rendimiento. Los requisitos del negocio pueden determinar qué tecnología de resiliencia de datos se necesita.

4.1.4. Oracle

Oracle es el estándar para las aplicaciones y la tecnología de base de datos para las empresas de todo el mundo. La compañía es líder mundial proveedor de software para la administración de la información, y es la segunda empresa de software independiente más grande del mundo. La tecnología de esta puede encontrarse en casi todos los sectores. Oracle es la primera empresa de software en desarrollar e implementar software empresarial 100 por ciento activado por Internet en toda su línea de productos: base de datos, aplicaciones comerciales y herramientas para el soporte de decisiones y el desarrollo de aplicaciones [60] .

Oracle define la alta disponibilidad mediante la siguiente serie de criterios técnicos [61]:

- **Fiabilidad:** la fiabilidad hardware es uno de los componentes para las soluciones de alta disponibilidad. La fiabilidad en software se constituye en un componente fundamental en la implementación de una solución de alta disponibilidad; en esta se incluyen las bases de datos, los servidores Web y las aplicaciones.

- **Recuperabilidad**: pueden existir diversas formas de recuperarse ante un error en el sistema. Por esta razón, es de gran importancia determinar qué tipo de errores pueden ocurrir en el entorno de alta disponibilidad y cómo recuperarse de estos de manera oportuna y que se ajuste a los requisitos empresariales.
- **Oportuna detección de errores:** se refiere a la rápida detección de las fallas inesperadas en el sistema, las cuales se pueden originar por la falla de uno o más componentes de la arquitectura. Aunque es posible la rápida recuperación ante una parada inesperada, son necesarios considerables minutos a fin de determinar el problema que causo el mismo.
- **Funcionamiento continuo:** proveer acceso continuo a los datos es esencial cuando los periodos de paradas no son aceptables o son poco aceptables para la realización de actividades de mantenimiento. Actividades tales como la restructuración de componentes software y la adición o cambio de componentes hardware deben ser completamente transparentes para el usuario final en una arquitectura de alta disponibilidad.
- **Buenas prácticas:** se refiere a las operaciones y procedimientos al interior de las organizaciones a fin de mejorar las reacciones ante posibles siniestros.
- **Clúster**: conjunto de computadoras funcionando en paralelo que se comportan como si fuesen una única computadora. La aplicación de clúster dentro de una organización incrementa su capacidad de procesamiento, tanto en hardware y software mediante la utilización de tecnologías estándar, aportando de esta manera beneficios como: alto rendimiento, alta disponibilidad, alta eficiencia y alta escalabilidad.

  La aplicación de clúster en las bases de datos por ejemplo, le permite a dos o más computadores de un clúster acceder concurrentemente a una sola base de datos compartida, creando un sistema de base de datos único que abarca múltiples sistemas de hardware y aparece frente a la aplicación como una base de datos unificada. Lo anterior genera beneficios de disponibilidad y escalabilidad para las aplicaciones, como los que se describen a continuación:

  - ✓ Tolerancia a fallas dentro del clúster, en especial a fallas del computador.
  - ✓ Flexibilidad y eficiencia de costos en la planificación de la capacidad, de manera que un sistema pueda escalar a cualquier capacidad deseada a pedido y a medida que las necesidades de negocio cambian.
- **Backup y Recuperación:** cada organización debe implementar un procedimiento para los backups de datos; el administrador debe poder recuperar los datos críticos del negocio desde el backup.

4.2. Elección de criterios técnicos asociados a la alta disponibilidad

A continuación se presentan los criterios técnicos que en el presente trabajo de grado se consideran cómo los más relevantes para el modelo de negocio de los operadores de telecomunicaciones que desean implantar o que ya posean una plataforma NGSDP. Para la selección de los mismos fueron diseñados una serie de filtros, los cuales fueron construidos teniendo en cuenta trabajos relacionados con la alta disponibilidad [6] [22] [32] [33], la visión

práctica de los operadores de telecomunicaciones[15] y toda la documentación adquirida a lo largo de la investigación [31][32][6][22][33]; de esta manera se obtiene un conjunto final de criterios, que se consideran deben ser cumplidos por un sistema altamente disponible.

Para obtener los criterios finales y como un aporte al presente trabajo de grado fueron diseñados cuatro filtros para la depuración de la base inicial de criterios, los cuales son explicados a continuación.

Filtro identificación: se refiere a la clasificación de los criterios en: hardware, software o conceptuales. En este sentido, el principal resultado de este filtro es la exclusión de los criterios hardware.

La exclusión de este tipo de criterios, responde a que en los sistemas de hoy en día, los periodos de inactividad "downtime", se presentan en la mayoría de los casos por fallas en los componentes software de la arquitectura [30]. Por esta razón, se decide concentrar el enfoque del presente trabajo de grado en este tipo de aspectos.

Filtro análisis conceptual: tomando el resultado arrojado por el filtro identificación, es decir los criterios software y los criterios conceptuales, el filtro selecciona los criterios de mayor importancia para el modelo de negocio de los operadores de telecomunicaciones. Para esta selección fueron tenidos en cuenta trabajos y documentación relacionada con la alta disponibilidad [31][32][6][22][33].

Filtro unificación: se comparan los criterios arrojados por el filtro "análisis conceptual" a fin de evitar la duplicación de los mismos, esto debido a que puede existir el caso en que uno o varios criterios sean definidos por más de una fuente de criterios.

Como resultado de este filtro se obtiene el conjunto de criterios iniciales.

Filtro Telco: mediante la colaboración de un operador de telecomunicaciones[16] son seleccionados del conjunto de criterios iníciales aquellos criterios que según la visión práctica y experiencia de los operadores, son los más relevantes a fin de alcanzar sistemas de telecomunicaciones altamente disponibles.

En la Figura 13 se muestra de forma gráfica y general el funcionamiento de los filtros y el resultado de los mismos.

[15] Se contó con la colaboración del operador de telecomunicaciones Colombiano EMCALI
[16] El filtro Telco no contó con la participación de más empresas de telecomunicaciones, debido a la complejidad de la consecución de citas de este tipo.

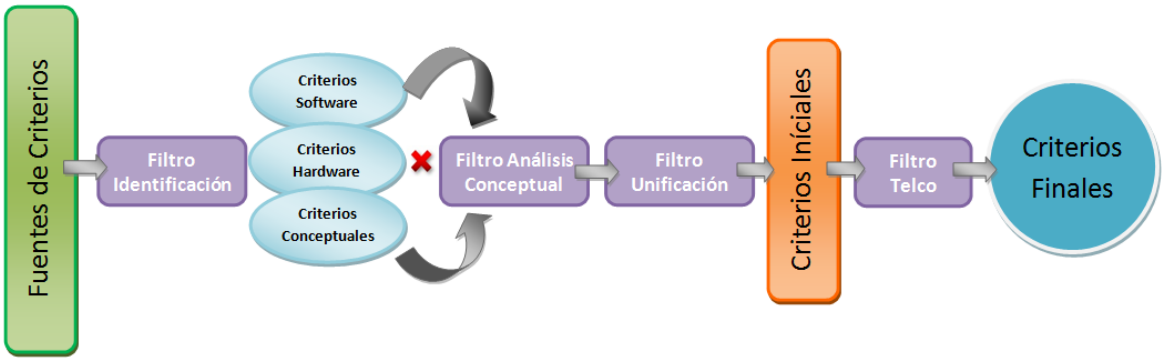

Figura 13. Fases para la elección de los criterios
(Imagen propia)

4.2.1. Filtro identificación

Para la visualización de la fase de identificación se construyó la Tabla 4, la cual se conforma por cuatro columnas que se enumeran de izquierda a derecha; en la primera se ubican las empresa prestadoras de servicios y sus respectivos criterios asociados a la alta disponibilidad; en la segunda, se señalan aquellos criterios que para su implementación son necesarios medios software; en la tercera aquellos criterios que para su implementación necesitan medios hardware y en la cuarta columna aquellos criterios conceptuales relacionados con el modelo de negocio de las telecomunicaciones.

Para señalar aquellos criterios que para su implementación necesitan medios software, hardware o son conceptuales se utilizó el símbolo de aprobación " ✓ ".

Como resultado de este filtro se obtienen los criterios que para su implementación requieren medios software y los criterios conceptuales; son excluidos los criterios hardware.

Empresa/Criterio	Software	Hardware	Conceptual
Cisco			
Confiabilidad, tolerancia a fallos de dispositivos de red	✓	✓	-
Redundancia de dispositivos y enlaces	-	✓	-
Balanceo de cargas	✓	✓	-
Resiliencia de tecnologías de red	-	✓	-
Diseño de red	-	-	✓
Buenas prácticas	-	-	✓
OpenCloud			
Rendimiento	-	-	✓
Fiabilidad y disponibilidad	✓	✓	-

Empresa/Criterio	Software	Hardware	Conceptual
OpenCloud			
Escalabilidad	✓	✓	-
Contención ante fallas	✓	✓	-
Construcción en redundancia	✓	✓	-
Gestión y sostenimiento	-	-	✓
IBM			
Presupuesto	-	-	✓
Requerimientos de tiempo de funcionamiento	-	-	✓
Cobertura de paradas	-	-	✓
Tiempo de recuperación objetivo (RTO)	-	-	✓
Punto de recuperación objetivo (RPO)	-	-	✓
Requerimientos de resiliencia	-	-	✓
Cambio automático y manual ante fallas	✓	✓	-
Requerimientos de distancia	-	-	✓
Número de sistemas de copia de seguridad	✓	✓	-
Acceso a una copia secundaria de datos	✓	✓	-
Rendimiento del sistema	-	-	✓
Oracle			
Fiabilidad	✓	✓	-
Recuperabilidad	✓	✓	-
Oportuna detección de errores	✓	-	-
Funcionamiento continúo	✓	-	-
Buenas prácticas	-	-	✓
Clúster	✓	✓	-
Backup y Recuperación	✓	✓	-

Tabla 4. Clasificación de Criterios

La anterior selección puede ser apreciada de forma gráfica en la Figura 14.

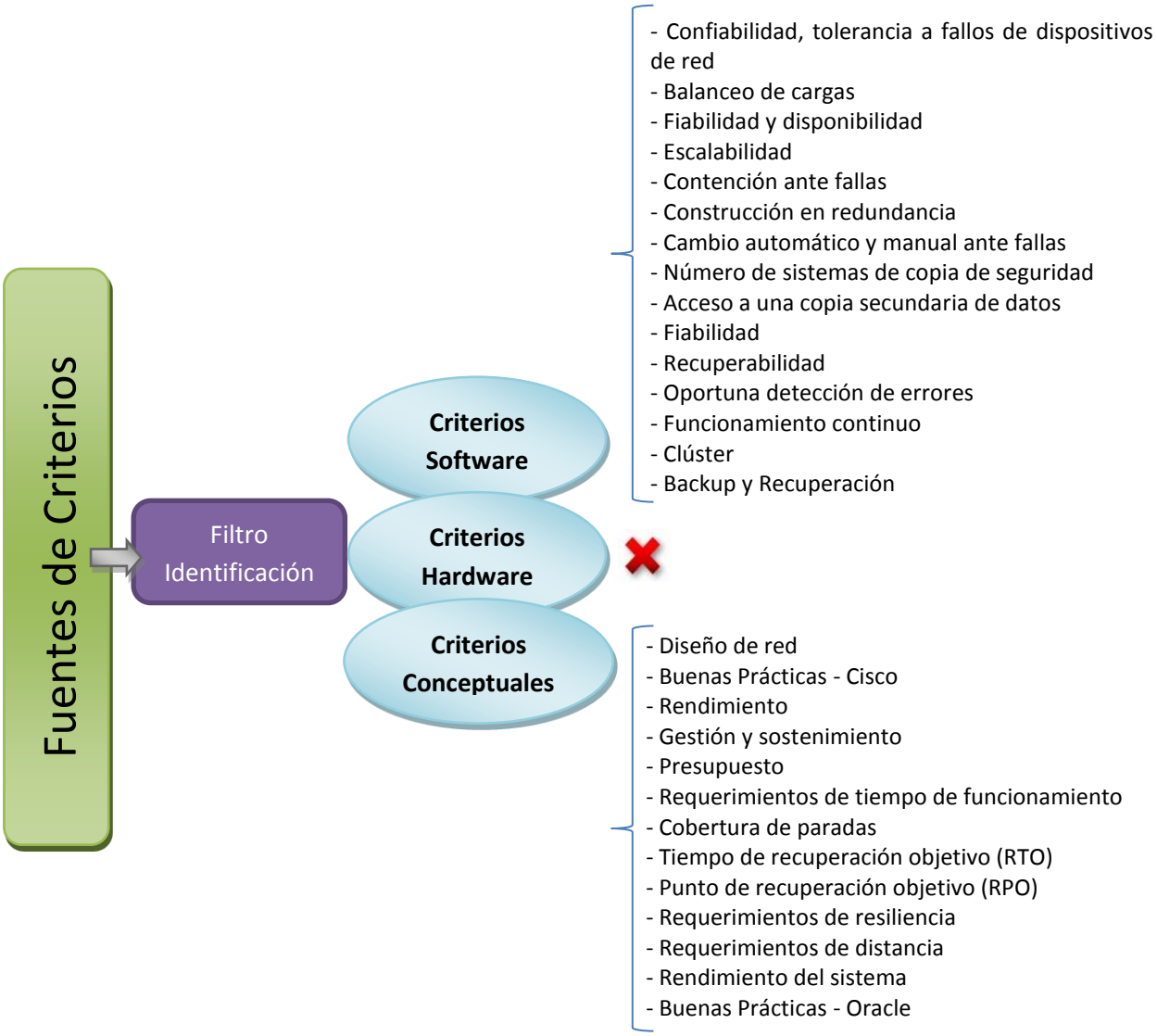

Figura 14. Filtro Identificación
(Imagen propia)

4.2.2. Filtro análisis conceptual

Toma los criterios arrojados por el filtro de identificación y selecciona aquellos que se consideran como los más relevantes para el modelo de negocio de los operadores de telecomunicaciones. Entendiendo como los criterios más relevantes aquellos que le permiten al operador incrementar la calidad en los servicios prestados y sus ganancias.

Para la elección de los siguientes criterios se tuvo en cuenta trabajos relacionados con la alta disponibilidad y la documentación adquirida a lo largo de la investigación [6][22][32][33].

En la Figura 15 se muestran los criterios obtenidos después de esta fase de filtrado.

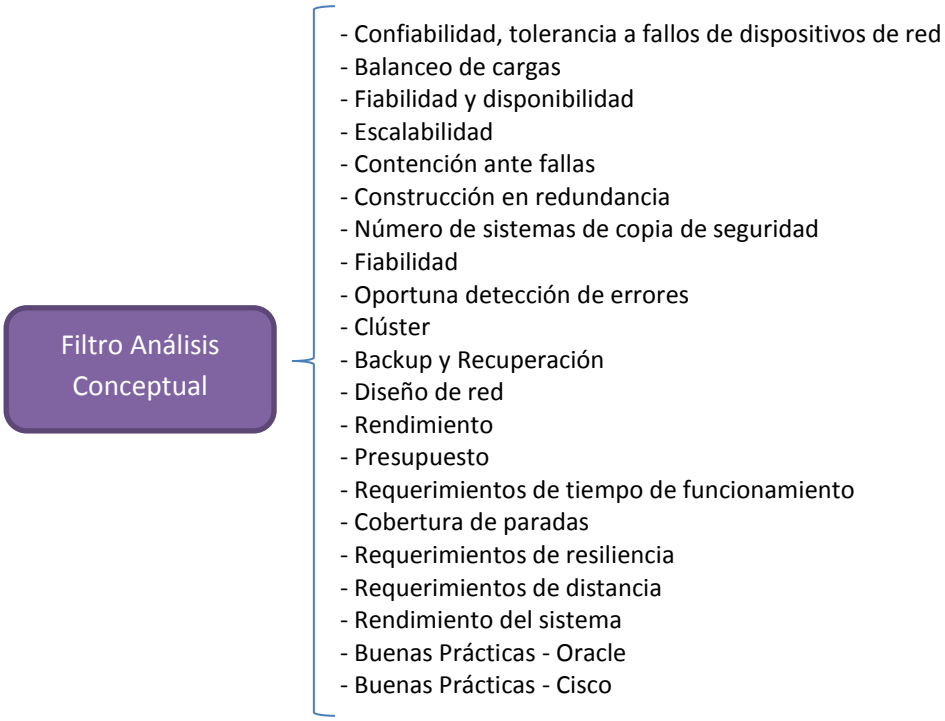

Figura 15. Filtro Análisis Conceptual
(Imagen propia)

A continuación se muestran las razones por las que fueron escogidos cada uno de los criterios.

- **Confiabilidad, tolerancia a fallos de dispositivos de red:** fue seleccionado debido a que la confiabilidad es un factor de gran importancia para los sistemas altamente disponibles; esto debido a que de la confiabilidad de los equipos, aplicaciones y servicios depende la disponibilidad total del sistema.

- **Balanceo de cargas:** es de gran importancia a fin de evitar sobrecargas en el sistema o denegación del servicio a los usuarios**.**

- **Fiabilidad y disponibilidad:** fue elegido debido a que de la fiabilidad de los elementos del sistema influye en el porcentaje de disponibilidad total del sistema.

- **Escalabilidad:** indispensable a fin de manejar el crecimiento continuo de usuarios y peticiones al sistema sin perder la calidad en los servicios prestados.

- **Contención ante fallas:** de gran importancia a fin de mejorar el tiempo de reactivación del sistema ante un falla y que de esta manera no se vea afectada en gran medida la disponibilidad del sistema.

- **Construcción en redundancia:** fue elegido debido a que un sistema altamente disponible debe contar con dispositivos redundantes, que le permitan la rápida recuperación ante las posibles fallas en el sistema.

- **Número de sistemas de copia de seguridad:** fundamental a fin de duplicar y en caso de siniestro recuperar la información necesaria para reactivar el sistema rápidamente.

- **Oportuna detección de errores:** a fin de evitar periodos extensos de downtime en el sistema.

- **Clúster:** de gran importancia a fin de dividir las tareas computacionales entre diferentes equipos o nodos y de esta manera aumentar el rendimiento del sistema.

- **Diseño de red:** fue eligió debido a que es indispensable tener un buen diseño de red, a fin de alcanzar alta disponibilidad y calidad de servicio.

- **Buenas prácticas:** diversas investigaciones han demostrado que las buenas prácticas dentro de una organización, son uno de los aspectos más importante a tener en cuenta, ya que con el cumplimiento de estas se incrementa considerablemente el rendimiento de la organización [6] [61].

- **Rendimiento:** es un criterio de gran importancia tanto para las organizaciones IT, como para las empresas de telecomunicaciones, ya que con este se define el desempeño deseado del sistema. A partir de este se trazan los requerimientos para cada una de las aéreas y equipos.

- **Presupuesto:** a partir del presupuesto de la organización se puede dar un estimado de la disponibilidad que teóricamente se podría alcanzar.

- **Requerimientos de distancia:** para cada organización es de gran importancia definir los requerimientos de distancia, ya que con estos se define seguridad y fiabilidad del sistema en caso de catástrofe, pero también trae implicaciones para el rendimiento del sistema, los cuales se traducen en la disminución de la QoS del sistema.

- **Requerimientos de tiempo de funcionamiento:** es muy importante fijar los requerimientos de funcionamiento del sistema, de estos puede salir un estimado del presupuesto necesario para cumplir con el requerimiento.

- **Cobertura de paradas:** se escogió debido a que un sistema altamente disponible debe contar con mecanismos que le permitan resolver las fallas de forma rápida y que no se afecte de forma considerable la disponibilidad del sistema.

4.2.3. Filtro unificación

Tomando los criterios arrojados por el filtro análisis conceptual, el filtro unificación compara los criterios entre sí, e identifica aquellos que son comunes en al menos dos de las empresas de referencia, en el caso de que exista duplicación de los mismos, estos son unificados en un solo criterio y enviados al conjunto inicial de criterios; en caso que algunos de los criterios no encuentren duplicados, estos pasan a formar parte del conjunto inicial de criterios.

Lo anterior se puede apreciar gráficamente de la siguiente forma (Figura 16).

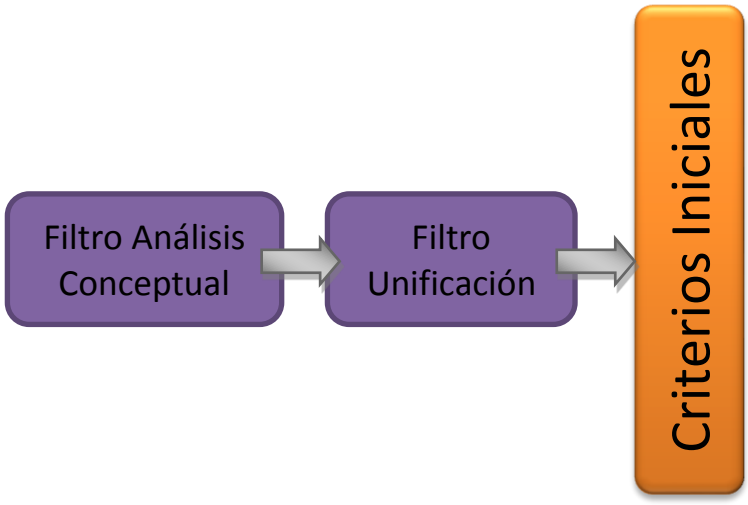

Figura 16. Filtro unificación
(Imagen propia)

Los criterios unificados en esta etapa de selección fueron:

Criterio	Cisco	IBM	Oracle	OpenCloude	Definición Criterio
Confiabilidad	✓	X	X	✓	Probabilidad de que un dispositivo o sistema lleve a cabo sus actividades prescritas sin fallas durante un tiempo determinado
Copias de seguridad	X	✓	✓	X	Las distintas tecnologías de resiliencia de datos ofrecen posibilidades de escoger diferentes sistemas de copia de seguridad y de copias de datos de aplicaciones.
Buenas Prácticas	✓	X	✓	X	Procedimientos documentados para desarrollar y mantener una infraestructura de red robusta.

De esta forma, la base inicial de criterios queda conformada por 16 criterios, los cuales se muestran a continuación (Figura 17):

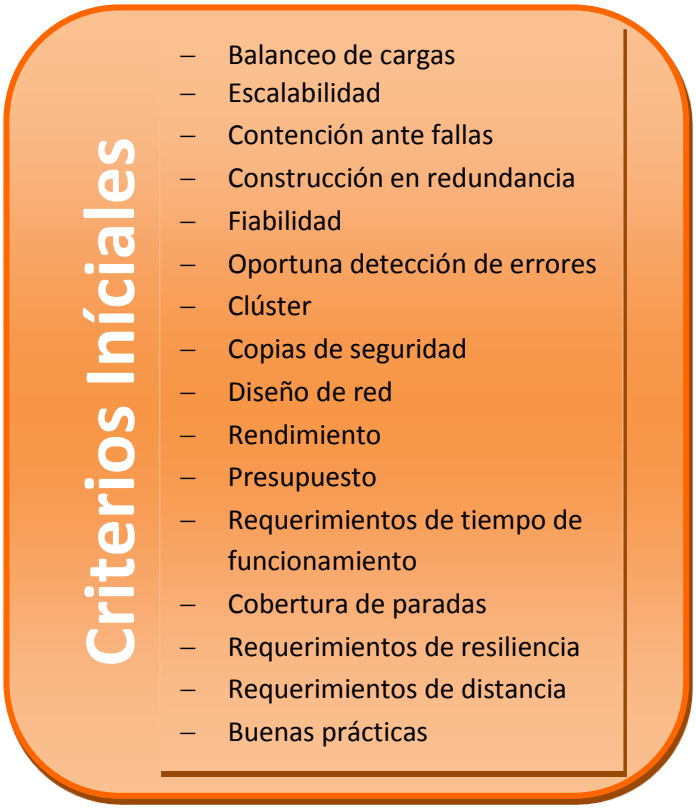

Figura 17. Criterios iniciales
(Imagen propia)

4.2.4. Filtro Telco

Mediante la colaboración del operador de telecomunicaciones EMCALI y algunos de sus expertos, en esta etapa de filtrado fueron seleccionados del conjunto de "criterios iníciales" aquellos criterios que según la experiencia adquirida a través de la práctica, los expertos del operador de telecomunicaciones, consideran como los más relevantes a fin de alcanzar sistemas altamente disponibles; de esta forma queda conformado el conjunto final de criterios relacionados con la alta disponibilidad, los cuales se consideran deben ser cumplidos a cabalidad por una NGSDP altamente disponible (Figura 18). En el anexo A se puede observar completamente el proceso realizado.

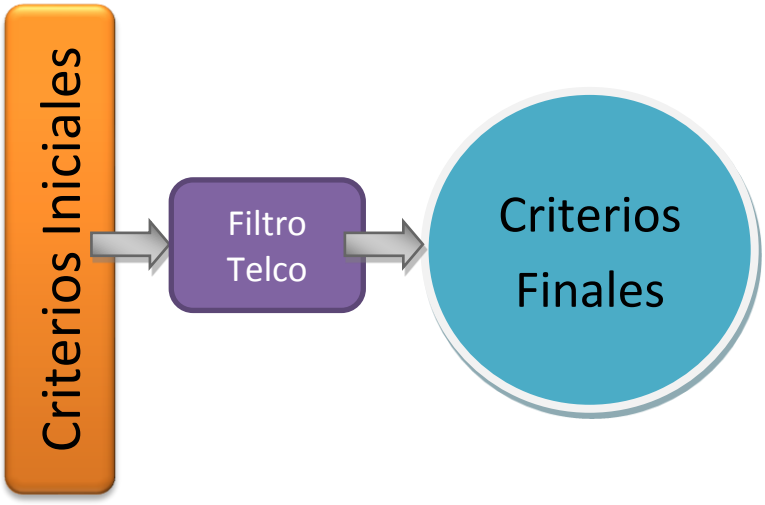

Figura 18. Filtro Telco
(Imagen propia)

El conjunto final de criterios quedó conformado por los siguientes diez criterios (Figura 19):

Figura 19. Criterios Finales
(Imagen propia)

4.3. Definición de Lineamientos Técnicos

A continuación son presentados los lineamientos técnicos para dar cumplimiento a los criterios relacionados a la alta disponibilidad definidos en la sección anterior. Para este propósito se tendrá el siguiente esquema (Figura 20), el cual se explica en la Tabla 5:

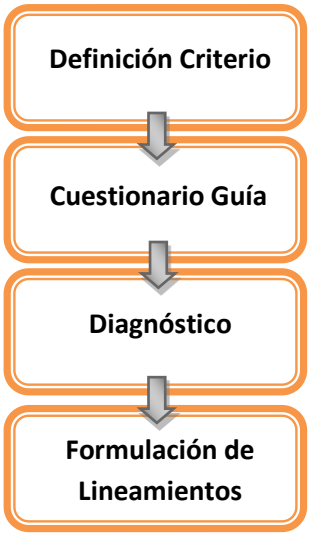

Figura 20. Fases para la formulación de lineamientos
(Imagen propia)

Definición criterio	Se realiza una completa descripción del criterio a tratar y los beneficios que trae para la organización su implementación.
Cuestionario guía	Con el fin de guiar al operador de telecomunicaciones hacia la elección de los posibles lineamientos a aplicar, se realiza una serie de preguntas a fin de identificar si es necesaria la implementación del criterio y que lineamiento es el más adecuado.
Diagnóstico	Se formula una recomendación, en la cual se indica cuál es el lineamiento más adecuado a implementar. A fin de formular un diagnóstico correcto, fue diseñada una encuesta de alta disponibilidad, la cual fue realizada a algunos de los expertos de la empresa de telecomunicaciones EMCALI (Anexo B); esto debido a que para realizar un diagnóstico adecuado se deben conocer profundamente el funcionamiento interno y los equipos de las organizaciones. Este diagnóstico será utilizado en la implementación del prototipo en el presente trabajo de grado.
Formulación de lineamientos	Son presentadas las diferentes alternativas a fin de dar cumplimiento a un determinado criterio.

Tabla 5. Formulación de lineamientos

Partiendo del conjunto final de criterios relacionados con la alta disponibilidad (Figura 21) a continuación se hace la presentación de cada uno de estos de acuerdo al esquema anteriormente presentado (Figura 20).

Figura 21. Conjunto de criterios relacionados con la alta disponibilidad

4.3.1. Balanceo de cargas

El balanceo de cargas tiene el propósito de equilibrar la carga de trabajo entre los diferentes sistemas de telecomunicaciones o del mundo IT. Su intención es aumentar el rendimiento de los sistemas evitando los denominados cuellos de botella, que se generan por la mala distribución de la carga de trabajo entre los equipos de cómputo y redes; el balanceo de cargas permite analizar continuamente las peticiones entrantes y asignar los medios más adecuados para su resolución, por lo tanto es de gran importancia contar con un balanceador de cargas en el sistema.

La implementación y el correcto funcionamiento de esta técnica, son de gran importancia en un sistema de alta disponibilidad; por lo cual, los operadores de telecomunicaciones que deseen implantar esta técnica a fin de mejorar la disponibilidad de la plataforma NGSDP, se deben plantear los siguientes interrogantes. Esto con el fin de identificar las falencias existentes en sus sistemas y la técnica que mejor se acomode para cada uno de ellos:

1. ¿Se está realizando balanceo de cargas en el sistema?
2. ¿Teniendo en cuenta las gráficas del monitor de recursos, el sistema está presentado síntomas de sobrecarga en la capacidad de funcionamiento?
3. ¿Ha percibido que el sistema no responde de manera adecuada a las peticiones de servicio entrantes?
4. ¿Qué clase de balanceo de cargas se está realizando en el sistema y cuál es el rendimiento que se está obteniendo?
5. Los cuellos de botella se presentan por la mala distribución de la carga de trabajo en: i) los niveles de acceso y transporte de la información ii) o por el contrario esta mala distribución se presenta en los niveles de aplicación.
   i. _____ (ir a la pregunta 6)
   ii. _____ (ir a la pregunta 7)

6. La mala distribución de la carga de trabajo a nivel de los componentes de red y transporte se presenta debido a:

   A. Cuellos de botella presentados por la mala distribución de trabajo entre las capas de enlace de datos y de red (OSI)___

   B. No se tiene redundancia en puertos___

   C. Mala distribución de las peticiones hacia los servidores___

7. La mala distribución de la carga de trabajo a nivel de aplicación se presenta debido al mal funcionamiento de:

   A. Mala distribución de las peticiones___

   B. Mal funcionamiento del servidor DNS___

   C. Mala distribución de la carga de trabajo en la base de datos___

   D. Mala distribución de la carga de trabajo SIP___

   E. Mala distribución de la carga de trabajo entre los componentes computacionales del sistema___

Teniendo en cuenta las anteriores respuestas, el operador de telecomunicaciones tiene diferentes opciones a fin de implementar, modificar o complementar el balanceo de cargas en su modelo de negocio; a fin de mejorar la disponibilidad del sistema.

Balanceo de cargas a nivel de infraestructura

- **Balanceo de cargas de capa 2:** también llamado agregación de puertos, consiste en la agregación de dos o más enlaces en uno solo, proporcionando redundancia y tolerancia a fallos; cada uno de los enlaces agregados sigue un camino físico diferente. La implementación de este tipo de balanceo de cargas se realiza con software de red, es decir a nivel de enlace de datos del modelo de referencia OSI (*Open System Interconnection*). Por lo cual para el presente trabajo de grado no se tendrá en cuenta este tipo de implementación, debido a que se toman aspectos físicos de la red, que se salen del alcance del trabajo de grado como se especificó en el anteproyecto de grado [62].

- **Balanceo de cargas de capa 4:** distribuye las peticiones hacia los servidores en la capa de transporte, la asignación de los servidores se realiza sin saber el tipo de información de cada petición por lo cual no se garantiza la calidad de servicio; es orientado a la conexión [62].

- **MPLS load balancing**: el balanceo de cargas en las redes MPLS distribuye el tráfico de la red entre los diferentes LSP (Label Switched Paths), esto en función del tráfico que tenga cada LSP a cada momento, evitando de esta forma la congestión de la red y mejorando el rendimiento del sistema. Otra alternativa para realizar balanceo de cargas en las redes MPLS es mediante el

establecimiento dinámico de nuevas rutas LSP; esto mediante el muestreo estadístico del tráfico y las funciones de notificación realizadas por los LSR[17] (Label Switching Router) [63].

Balanceo de cargas a nivel de aplicación

- **Balanceo de cargas de capa 7:** la distribución del trabajo a realizar se hace a nivel de la capa de aplicación del modelo de referencia OSI, es decir que se analizan y distribuye las peticiones hacia los diferentes servidores basándose en el contenido de cada una de ellas. Con este tipo de balanceo de cargas el sistema provee calidad de servicio y se mejora el rendimiento de los servidores, sin embargo la cabecera en la capa de aplicación es grande, y su análisis también; limitando de esta forma la escalabilidad del sistema [64].

- **Balanceo de cargas en DNS:** consiste en la instalación de un servidor de nombres (DNS, Domain Name System) que devuelva una dirección IP distinta cada vez que un cliente solicite el identificador numérico correspondiente a un nombre en particular. Entre las técnicas que permiten la implementación de este balanceo de cargas se encuentran Round Robin y NLB (Network Load Balancing), las cuales se explican a continuación [65] [66].

  - **Round Robin:** selecciona una opción entre varias direcciones IP de forma consecutiva. Una vez que se hace uso de la última dirección, se vuelven a ordenar de manera distinta realizando una permutación cíclica de las mismas mediante el algoritmo de Round Robin, así el cliente hace la solicitud a un servidor distinto cada vez que inicie una conexión [66]. De esta forma, se balancean por igual las peticiones entre los servidores disponibles. Esta técnica se caracteriza por su sencillez ya que no necesita el uso de ningún hardware ni software adicional [67].

  - **NLB (Network Load Balancing):** NLB es una tecnología de Microsoft disponible desde Windows NT 4, disponiendo actualmente de varios años de funcionamiento en entornos de producción en todo tipo de empresas; distribuye de forma transparente las solicitudes de cliente entre los servidores en un clúster NLB virtual mediante el uso de direcciones IP y un nombre común. Desde la perspectiva del cliente, el clúster NLB parece ser un único servidor. No utiliza un despachador central. Cuando las peticiones de resolución de dominio son resueltas, estas se redirigen hacia uno de los servidores asociados al servidor DNS, estas distribuciones se realizan mediante esquemas como: posición geográfica, acuerdos de horarios, entre otros [68].

- **Balanceo de cargas en bases de datos:** las peticiones de acceso que se realizan a la base de datos se distribuyen entre clúster de servidores de bases de datos, alcanzando de esta forma alta disponibilidad y escalabilidad en la base de datos [69].

- **Balanceo de cargas SIP:** se utiliza para lograr rendimiento, escalabilidad y alta disponibilidad de los servicios SIP en cualquier servidor, ya que puede estar limitado el número simultáneo de conexiones y sesiones que se pueden manejar en un determinado momento. El objetivo es aumentar la disponibilidad de los servicios, por lo cual las nuevas solicitudes se reparten entre

[17] LSR: elemento que conmuta etiquetas.

los servidores disponibles utilizando un algoritmo de selección. A gran escala los servicios corporativos poseen múltiples servidores con el fin de atender varias operaciones solicitadas por los clientes al mismo tiempo. Varios servidores instalados o incluso clúster de servidores trabajan para procesar el tráfico de entrada y/o salida evitando esperas y procesos abrumadores que pueden provocar cuellos de botella o bloqueos y logrando así que el servicio se pueda sostener sin vigilancia y sin degradar la calidad de servicio (QoS) [70] [71].

- **Balanceo de cargas computacional:** divide una tarea computacional entre los diferentes nodos de un clúster, logrando de esta forma que el sistema en su totalidad proporcione un mayor rendimiento en comparación con que la misma tarea fuera realizara por un solo nodo [62].

4.3.2. Escalabilidad

La escalabilidad es la capacidad del sistema para manejar el crecimiento continuo de trabajo de manera fluida o para estar preparado para hacerse más grande sin perder calidad en los servicios ofrecidos en un sistema o red a medida que las necesidades del negocio lo ameriten [27]. La escalabilidad está compuesta por dos dominios, el dominio Hardware o escalabilidad horizontal y el dominio Software o escalabilidad vertical; el logro de una verdadera escalabilidad es un acto de equilibrio que implica la elección de hardware escalable, junto con el software adecuado que está específicamente diseñado para aprovechar la misma [72], ya que no es suficiente poseer potentes medios hardware si no se cuenta con un software lo suficientemente robusto que los respalde y viceversa.

En un futuro ambiente de convergencia, donde se espera un aumento considerable en cuanto a las peticiones realizadas a los sistemas de telecomunicaciones, es fundamental contar con sistemas escalables que permitan responder de manera adecuada a este crecimiento. A fin de determinar las capacidades de escalabilidad con las que cuenta el operador, así como para determinar las falencias y las posibles soluciones, el operador de telecomunicaciones debe plantearse las siguientes preguntas.

1. ¿Su sistema cuenta con escalabilidad?
2. Teniendo en cuenta el número de usuarios con los que cuenta la organización y que idealmente el funcionamiento debe ser el mismo cuando se atienden pocos o muchos usuarios. ¿Cuál es el comportamiento del sistema (funcionamiento) con un número elevado de usuarios?
3. ¿Cuál es la granularidad con la que el sistema puede escalar? Tenido en cuenta que una granularidad baja significa que el sistema puede escalar en pasos pequeños.
4. ¿Cuál es el funcionamiento del sistema a grandes escalas en comparación con el funcionamiento del mismo a bajas escalas? Teniendo en cuenta que idealmente se debe tener la misma operación en los estándares como en el funcionamiento a bajas escalas.
5. Teniendo en cuenta que la escalabilidad del sistema cuenta con varias facetas, como: *i)* tamaño de la base de datos, *ii)* número de usuarios, *iii)* número de consultas y *iv)* transacciones realizadas en el sistema. Para ofrecer escalabilidad en su sistema, ¿se necesita el mejoramiento de una o varias facetas dentro de la organización?

Teniendo en cuenta las respuestas a las anteriores preguntas el operador de telecomunicaciones tiene diferentes opciones a fin de implementar o cambiar la escalabilidad en su modelo de negocio, con el fin de mejorar la capacidad de su sistema; las cuales se muestran a continuación:

Escalabilidad Horizontal:

La escalabilidad horizontal consiste en adicionar nuevos nodos físicos, con funciones idénticas a las ya existentes, y la redistribución de la carga entre todos ellos dentro de un sistema. Los sistemas presentan escalabilidad de hardware, añadiendo más servidores a una red con equilibrio de carga, con el fin que las solicitudes de entrada puedan ser distribuidas entre todos ellos y funcionar esencialmente como un único equipo. Al dedicar varios equipos a una tarea común mejora la tolerancia de errores de la aplicación. Escalar en horizontal presenta un desafío mayor de administración debido al mayor número de equipos [73] .

En la Figura 22, se puede apreciar gráficamente la adición de un nuevo nodo para lograr este tipo de escalabilidad; la capacidad de servicio lograda versus la capacidad de procesamiento logrado con la misma se pueden apreciar en la Figura 23.

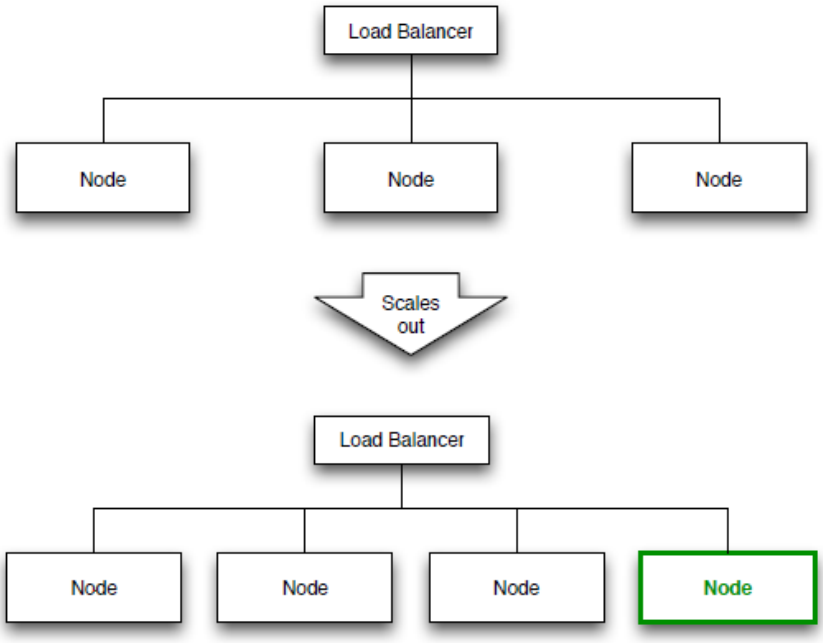

Figura 22. Escalabilidad Horizontal

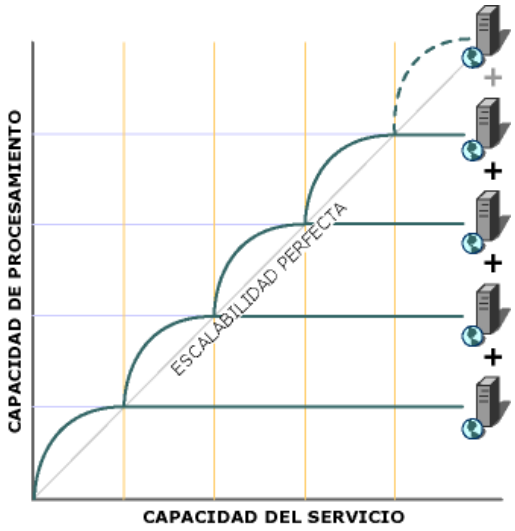

Figura 23. Capacidad de procesamiento - Escalabilidad Horizontal

Escalabilidad Vertical:

La escalabilidad vertical o de software, se puede implementar en equipos con los que ya cuente la organización (agregación de nodos virtuales), o en el mismo hardware mediante el incremento de [27]: *i)* el número de procesadores o procesadores más rápidos, *ii)* la cantidad de memoria principal, *iii)* migrando las aplicaciones a un equipo más potente, para así alojar más servidores virtuales, o nodos. Este método permite un aumento en la capacidad sin requerir cambios en el código fuente. Desde el punto de vista administrativo, las cosas permanecen igual ya que sigue existiendo un único equipo, que se debe administrar [73]. El escalamiento vertical, puede traer posibles problemas, ya que el uso de un único equipo en el cual se encuentran las aplicaciones, puede crear un único punto de error, lo que disminuye la tolerancia de errores del sistema [73].

En la Figura 24, se puede apreciar gráficamente la adición de un nuevo nodo virtual, para lograr escalabilidad. La capacidad de servicio lograda versus la capacidad de procesamiento logrado con la misma se pueden apreciar en la Figura 25.

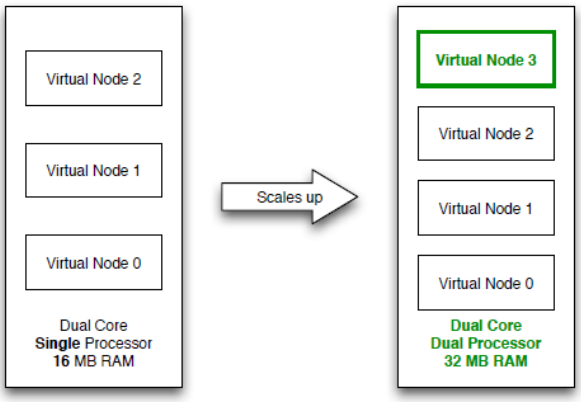

Figura 24. Escalabilidad Vertical [73]

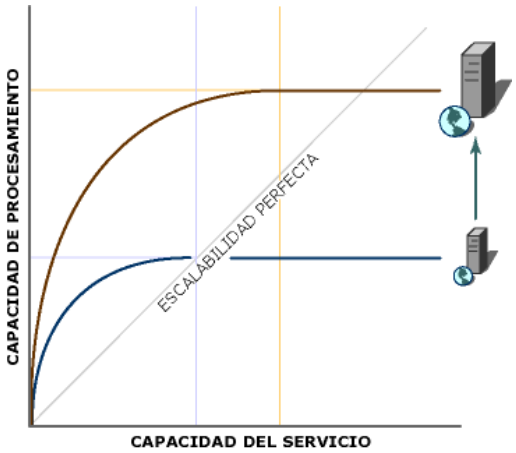

Figura 25. Capacidad de procesamiento - Escalabilidad vertical [73]

4.3.3. Fiabilidad

La fiabilidad es la característica de un elemento expresada por la probabilidad que los componentes, partes y sistemas lleven a cabo una función requerida bajo condiciones dadas en un intervalo de tiempo indicado. Desde el punto de vista cualitativo, la fiabilidad se puede definir como la capacidad del elemento para seguir funcionando. Cuantitativamente, la fiabilidad especifica la probabilidad de que las interrupciones de funcionamiento no ocurran durante un intervalo de tiempo indicado [74] [75] [76]. Según [77], la fiabilidad es un conjunto de atributos que inciden en la capacidad del sistema para mantener su nivel de rendimiento bajo condiciones establecidas por un período determinado de tiempo (ISO 9126: 1991, 4.2).

La fiabilidad incluye i) la corrección (asegurar que los servicios que proporciona el sistema son los que se han especificado), ii) precisión (asegurar que la información se proporciona al usuario con el nivel de detalle adecuado), y iii) oportunidad (asegurar que la información que proporciona el sistema esté cuando se requiera). Además, un sistema puede convertirse en no fiable si sus datos han sido corrompidos [78].

La fiabilidad y la capacidad de mantenimiento son funciones, que ayudan a determinar la disponibilidad en un sistema. Si la fiabilidad se mantiene constante, esto no implica directamente una alta disponibilidad. Pero a medida que el tiempo de reparación incrementa, es decir disminuye la capacidad de mantenimiento, la disponibilidad disminuye (ver Tabla 6). Un sistema con una fiabilidad baja y tiempo de reparación corto, podría tener una alta disponibilidad [79].

Fiabilidad	Capacidad de Mantenimiento	Disponibilidad
Constante	Disminuye	Disminuye
Constante	Aumenta	Aumenta
Aumenta	Constante	Aumenta
Disminuye	Constante	Disminuye

Tabla 6. Relación entre Fiabilidad, Capacidad de Mantenimiento y Disponibilidad [79]

Como se puede apreciar de la anterior información, la fiabilidad es un aspecto que influye directamente en el porcentaje de disponibilidad del sistema, de aquí la importancia de una correcta y oportuna determinación de la misma.

Para esto, se debe hacer una declaración numérica de la fiabilidad, la cual debe ir acompañada de la definición de la función deseada, las condiciones de funcionamiento, y la duración de la misión. La definición de la función deseada es el punto de partida para cualquier análisis de confiabilidad, ya que definen las posibles fallas. Las condiciones de funcionamiento tienen una influencia importante sobre la fiabilidad, y por lo tanto se deben especificar con cuidado. La función requerida y/o las condiciones de funcionamiento pueden ser dependientes del tiempo. En estos casos, el perfil de la misión tiene que ser definido y todas las cifras de fiabilidad serán relacionadas con él. A menudo, la duración de la misión se considera como un parámetro t, así la función de fiabilidad se define como R (t), que es la probabilidad de que ningún fallo se produzca en el intervalo (0, t]. La condición del elemento en t = 0 (nuevo o inexistente) influye en los resultados finales. Para considerar esto, las cifras de fiabilidad a nivel de sistema deben tener índices S*i* (por ejemplo, R_{Si} (t)), donde S es sinónimo de sistema e *i* es el estado en t = 0 [76].

Adicionalmente, se debe considerar que la fiabilidad varía con el paso del tiempo y que la fiabilidad hardware y software son significativamente diferentes. En la Figura 26, se muestran las fallas características de hardware; se pueden identificar tres fases de izquierda a derecha: i) la fase "Burn in" de lanzamiento o de desarrollo, ii) la fase "Useful Life" o vida útil y iii) la fase "Wear out" o final del ciclo de vida, adicionalmente se puede observar el símbolo "λ(t) Failure intensity" o intensidad de fallas, el cual es expresado en función del tiempo y puede ser calculado como se muestra a continuación (Ec.3) [80].

Ec. 3. Ecuación para el cálculo de Intensidad de fallas

$$\boldsymbol{Mean\ Time\ To\ Failure} = \frac{\mathbf{1}}{\boldsymbol{\lambda(t)}}$$

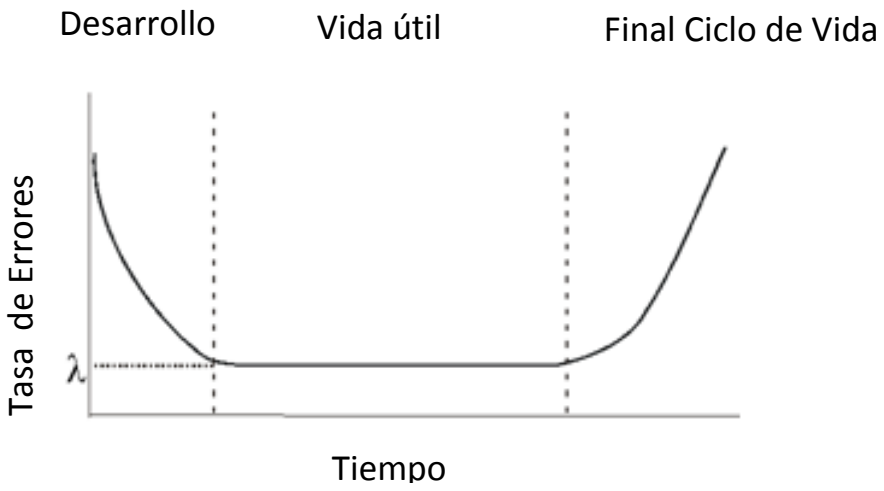

Figura 26. Curva de bañera para la fiabilidad hardware [80]

La fiabilidad del software (

Figura 27), posee dos grandes diferencias en comparación a la fiabilidad hardware, la primera diferencia se nota en la primera fase “Test/Debug” o fase de pruebas de vida útil, en la cual software va a experimentar un aumento drástico en la tasa de errores cada vez que se realice una actualización; la segunda diferencia se encuentra en la tercera fase “Obsolescence” u obsolescencia, el software no tiene una tasa de error tan alta como el hardware y además ya se está acercando a la obsolescencia y no hay motivo para realizar actualizaciones o cambios [80].

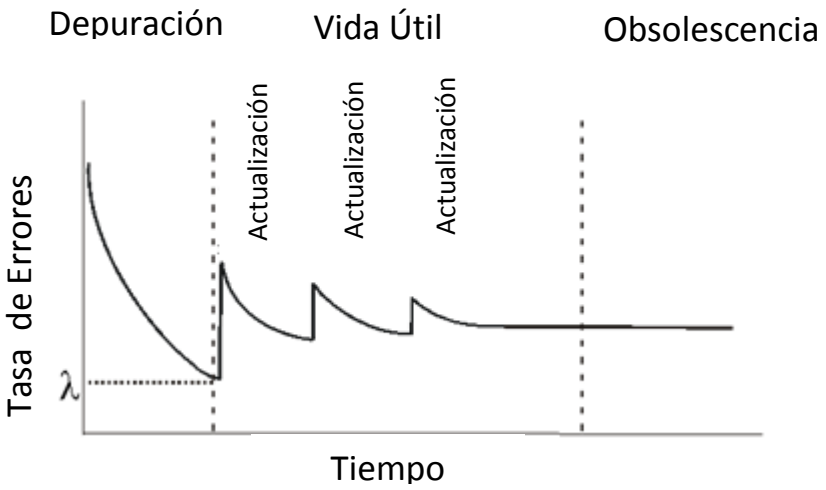

Figura 27. Curva para la fiabilidad de software [80]

La fiabilidad de software, es la aplicación de técnicas estadísticas a los datos recopilados durante el desarrollo y operación del sistema para especificar, predecir, estimar y evaluar la fiabilidad de los sistemas basados en software. "Software Reliability Engineering” (SRE) es un estándar el cual provee las mejores prácticas, para realizar las pruebas más fiables, rápidas y económicas y se puede aplicar a cualquier sistema que esté utilizando software [81] [82].

Por lo anterior y a fin alcanzar alta disponibilidad, es vital la determinación tanto de la fiabilidad software como de la hardware, sin embargo, y siendo el propósito principal del presente trabajo de grado guiar a los operadores de telecomunicaciones hacia el mejoramiento de la disponibilidad mediante medios software, a continuación se proponen una serie de preguntas a fin de proporcionar una guía para estos hacia la selección del método de estimación de la fiabilidad para sus respectivos casos.

1. ¿Conoce la fiabilidad de su sistema?
2. ¿Utiliza algún método para determinar la fiabilidad de su sistema?
3. Su sistema cuenta con:
    - **A.** Software ya desarrollado, es decir compra el software a otras empresas.

B. Software a desarrollar, cuenta con un grupo de desarrolladores que implementaran la solución.

C. Todas las anteriores.

Las anteriores preguntas son una guía para que los operadores de telecomunicaciones puedan definir cuál de los modelos, es el más indicado para su caso específico, sin embargo y debido a la complejidad de la estimación de la fiabilidad software, cualquier modelo debe tener suposiciones adicionales y solo ciertos factores pueden ser puestos a consideración, como un ejemplo de la anterior premisa se puede considerar el caso en el cual se mide la fiabilidad del sistema en un entorno, no se puede suponer que la fiabilidad del mismo será la misma en otro entorno en el que el sistema se usa de forma diferente; por lo cual se puede concluir que no existe un modelo que se pueda utilizar en todas las situaciones, de aquí la importancia de que cada operador escoja de manera cuidadosa un modelo de predicción que se adapte a sus condiciones de funcionamiento específicas.

Teniendo en cuenta la anterior información y las respuestas del cuestionario, el operador de telecomunicaciones tiene diferentes opciones a fin de implementar o cambiar el modelo de predicción y/o estimación de fiabilidad software, estos modelos se pueden dividir en dos categorías, los cuales se basan en i) observación y recolección de datos, para los cuales son utilizados los modelos de predicción y ii) el análisis de fallas, para el cual son utilizados los modelos de estimación [80].

Modelos de Predicción: se basa en la predicción de la fiabilidad al principio de la fase de desarrollo (software desarrollado al interior de la organización) y las mejoras requeridas para aumentar la fiabilidad y por ende la disponibilidad del sistema; pueden ser implementadas posteriormente. Entre los métodos disponibles se encuentran [9] [83]:

- Musa's Execution Time Model.
- Putnam's Model.
- MIL-HDBK-217.
- Bellcore/Telcordia.
- Rome Laboratory Model TR-92-52.
- Rome Laboratory Model TR-92-15.
- RDF 2000.
- NTT Procedure.
- Siemens SN29500.
- China 299B.
- PRISM.

Modelado de Estimación: el uso de modelos de estimación hace que la fiabilidad del software pueda ser estimada para los sistemas software ya desarrollado (software comprado). Incluye, entre otros, los modelos nombrados a continuación:

- Classical Fault Count/Fault Rate.
    - Exponential Distribution Model.
    - Weibull distribution Model.

- Modelos Bayesianos Fault Rate.
    - Thompson y Chelson Model.

4.3.4. Copia de seguridad

Es la actividad de copia total o parcial de información o bases de datos a fin de que se conserven de forma segura, y que puedan utilizarse para restaurar el sistema original en caso de una pérdida eventual de la información, fallo del equipo o de otra catástrofe. Hacer copias de seguridad es una rutina en la operación de las empresas y es parte de un plan de protección contra los desastres; si este proceso no se diseña y se prueba minuciosamente, es posible que no proporcione las capacidades de recuperación de desastres y de protección de datos deseado [84]. Este proceso puede realizarse de dos formas:

> **Manual:** se copian directamente los archivos a respaldar por medio de comandos o por medio del explorador de archivos de los respectivos sistemas operativos.
>
> **Automático:** por medio de una aplicación especializada, se programan los archivos a guardar, conformado un sistema de archivos de respaldo, los cuales se van actualizando en tiempo real a medida que se vayan registrando cambios en estos.

Independientemente de los posibles daños que se puedan producir en el sistema, es fundamental que el mecanismo o mecanismos de copia de seguridad que implementen los operadores de telecomunicaciones, estén diseñados de tal forma que permita la continuidad y recuperación de todos los datos importantes, sin interrumpir el funcionamiento y la alta disponibilidad del sistema.

Por lo anterior, y teniendo en cuenta el futuro ambiente de convergencia de las telecomunicaciones, facilitado en gran medida por la NGSDP, en el cual se espera el aumento de los clientes, peticiones, datos suministrados, servicios multimedia, entre otros, es aconsejable que los operadores de telecomunicaciones diseñen estratégicamente mecanismos de copia de seguridad periódicamente, indicando con claridad cuáles son los datos de los cuales se necesita realizar una copia, con qué frecuencia, con qué método, y que plan de recuperación de desastres se implementará para restablecer el funcionamiento normal en el caso de que surja algún problema. A fin de proporcionar una guía para los operadores de telecomunicaciones, en la cual se les recomiende el método de copia de seguridad más adecuado para su caso específico, los operadores deben responder el siguiente cuestionario.

1. ¿Realiza copias de seguridad?

2. ¿Tiene algún plan de contingencia contra desastres o fallas que puedan afectar su sistema?

3. ¿Se pierde información, datos, y archivos de su sistema y no es posible recuperarlos?

4. ¿Tiene su sistema algún método de recuperación de datos?

5. ¿Qué tipo de copias de seguridad realiza?

6. ¿Cada cuánto realiza sus copias de seguridad?

7. ¿Está conforme con los resultados obtenidos de sus copias de seguridad?

8. ¿Con su método de copia de seguridad puede recuperar la información totalmente?

9. ¿El rendimiento de la copia de seguridad es el deseado?

Teniendo en cuenta las anteriores respuestas y a fin de mejorar la fiabilidad y por ende la disponibilidad del sistema, los operadores de telecomunicaciones tienen diferentes opciones a fin de implementar o cambiar el tipo de copia de seguridad en su sistema, como se muestra a continuación [85] [86]:

Copia de seguridad completa: Este tipo de copia de seguridad permite la copia total de la información. Su gran ventaja es que permite almacenar toda la información en un solo lugar, facilitando de esta forma el proceso de restauración de la información, sin embargo el tiempo y el espacio de almacenamiento son considerablemente elevados (dependiendo del volumen de la información), por esto las copias de seguridad completas no son realizadas con mucha frecuencia, solo son realizadas cuando el volumen de la información lo amerita.

Con el fin de mejorar el rendimiento de las copias de seguridad completas, a menudo son combinadas con las copias de seguridad incrementales o diferenciales.

Copia de seguridad incremental: solo copia los datos que han variado desde la última copia de seguridad realizada; se suele utilizar la hora y fecha de modificación en los archivos para compararla con la hora y fecha de la última copia de seguridad. Una copia de seguridad incremental se puede ejecutar el número de veces que se desee, pues sólo guarda los cambios más recientes. Por lo tanto, guarda una menor cantidad de datos que una copia de seguridad completa, haciendo que sus operaciones sean más rápidas y que ocupen menor espacio.

En la Figura 28 se muestra gráficamente como funciona una copia de seguridad completa e incremental.

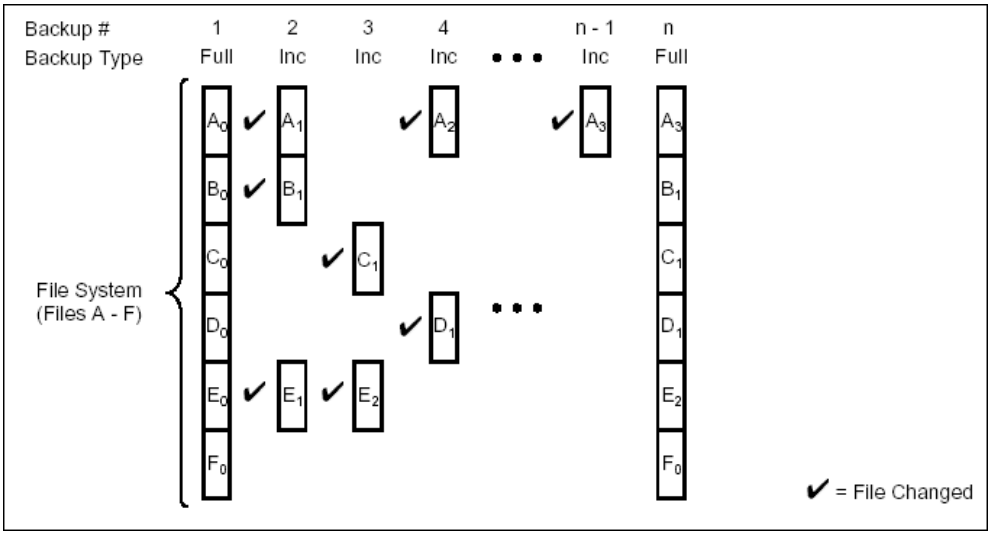

Figura 28. Copia de seguridad completa e incremental [85]

En la Figura 28 se puede apreciar de izquierda a derecha, "Backup #, Backup type" lo cual simboliza el número y tipo de copia de seguridad, así como también "File System (Files A-F)" que

simbolizan los archivos de las copias; la copia de seguridad uno, es completa o "Full" y por lo tanto guarda todos los datos, con cambios y sin cambios. De la copia 2 a n-1 son incrementales, es decir que solo se copian aquellos archivos que han sido modificados desde la última copia de seguridad, independientemente del tipo. Por ejemplo, los archivos A, B y E cambian después de la copia de seguridad completa y son apoyados en la copia de seguridad incremental 2. La copia de seguridad 4 copia los archivos A y D porque ambos archivos cambian en algún momento después de la copia de seguridad 3. El archivo F no ha cambiado, por lo que no existe una copia de seguridad de este archivo en cualquiera de las copias de seguridad incrementales anteriores, pero fue incluido en ambas copias de seguridad completas.

Copia de seguridad diferencial: es similar a la copia de seguridad incremental debido a que almacena todos los datos nuevos o que han variado desde la última copia de seguridad. Sin embargo, cuando se vuelve a ejecutar continúa copiando los datos que han cambiado. Es decir que almacena más datos que una copia de seguridad incremental, pero menos que una completa.

En la Figura 29 se muestra cómo funciona la copia de seguridad diferencial.

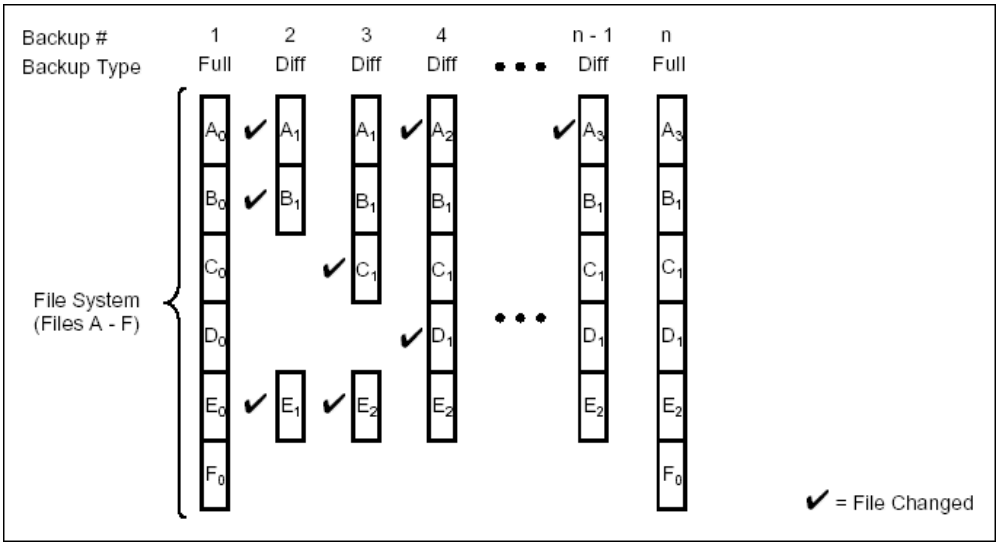

Figura 29. Copia de seguridad diferencial [85]

La copia de seguridad 1, como la anterior definición, es una copia de seguridad completa. Las copias de seguridad, de la 2 a n-1 son copias de seguridad diferenciales. Por ejemplo, los archivos A, B y E cambian después de la copia de seguridad completa y son apoyados tanto en la copia de seguridad 2, así como en todas las copias de seguridad diferenciales posteriores. El archivo C cambia en algún momento después de la copia de seguridad 2 y es respaldado en la copia de seguridad 3 y en todas las copias siguientes. El archivo F no cambia, por lo que no está en las copias de seguridad diferenciales, pero está incluido en las 2 copias de seguridad completas.

Copia de seguridad completa sintética: este tipo de copia de seguridad no lee ni copia los datos directamente del disco, al contrario de esto, sintetiza los datos de las anteriores copias de seguridad completas, incrementales y diferenciales guardadas y se restaura en un medio alterno para su almacenamiento, reduciendo así el tiempo de recuperación.

Copia en espejo: consiste en la copia de los datos almacenados en el disco de copia de seguridad hacia otro juego de discos.

Copia de seguridad incremental inversa: es utilizada para añadir soporte a la copia de seguridad de tipo incremental. Al aplicar este tipo de copia a un servidor espejo, el resultado será una versión previa del mismo.

Protección de datos continua (CDP): permite un mayor número de puntos de restauración que las opciones de copia de seguridad tradicionales.

4.3.5. Clúster

Un clúster, es una solución que permite aumentar la fiabilidad y la disponibilidad de los sistemas; esto debido a que permite utilizar un conjunto de recursos computacionales autónomos interconectados que se ven como un solo elemento hacia el exterior, lo cual permite eliminar los denominados puntos de fallos, es decir que, ante el caso de un fallo de uno de los componentes del clúster, se hace el relevo del mismo, asegurando de esta forma el continuo funcionamiento del sistema [87] [88] [89].

Además de garantizar escalabilidad, fiabilidad y disponibilidad de los sistemas, los clúster juegan un papel de gran importancia en la solución de problemas y en la ejecución de diversas aplicaciones sobre los sistemas de telecomunicaciones. Adicionalmente es de fácil implementación debido a la flexibilidad en la utilización de componentes hardware, software y sistemas operativos [90].

Para el correcto funcionamiento de un clúster son necesarias las conexiones adecuadas entre los elementos y un sistema para el manejo del mismo, que se encargue de la interacción con los usuarios y con los procesos que corren sobre este.

Un clúster debe poseer las siguientes características [91]:

- Un clúster consta de 2 o más nodos.
- Los nodos del clúster están conectados entre sí por al menos un canal de comunicación.
- Los clúster necesitan un software de control especializado. Existen dos modos: el control centralizado, donde hay un nodo maestro con el cual se configura todo el sistema, y el control descentralizado, en el cual cada nodo se administra y gestiona individualmente.

La arquitectura de un Clúster comprende componentes hardware y software, como se muestra en la Figura 30 [92]. Esta arquitectura permite la ejecución de aplicaciones secuenciales y paralelas.

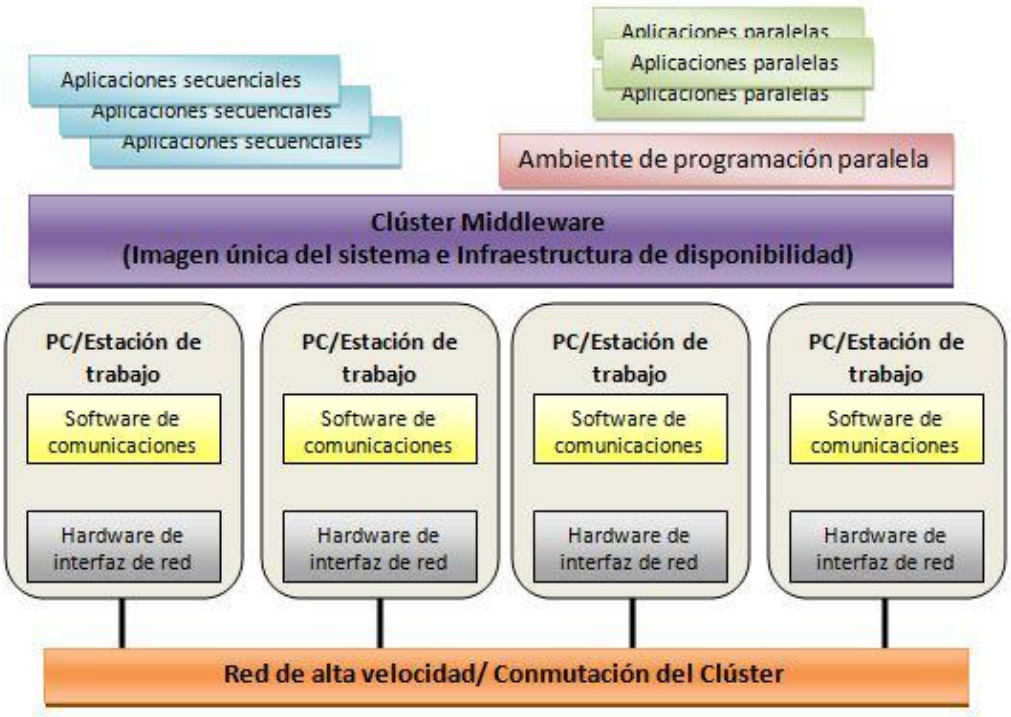

Figura 30. Arquitectura básica de un Clúster [91]

El Clúster Middleware forma parte del software, se encuentra entre el sistema operativo de los nodos del Clúster y las aplicaciones, por lo cual tiene una visión global de los recursos del sistema y la disponibilidad de los mismos. Los componentes hardware están conformados por el conjunto de computadores o nodos, los cuales presentan interfaces y software de comunicación; además para la interconexión con los otros nodos, hacen uso de una red de área local en la cual se utilizan tecnologías de alta velocidad, como Fast Ethernet, Gigabit Ethernet, o hardware especializado que provee un ancho de banda adecuado y tiempos de latencia bajos [93].

Debido a que existen diversas forma de implementar sistemas de clúster los operadores de telecomunicaciones deben escoger el que mejor se adapte a sus necesidades específicas; para esto se debe tener en cuenta los recursos hardware y software con los que cuentan, además de considerar la forma como se desea el funcionamiento del sistema.

Para lo anterior el operador debe responder el siguiente cuestionario guía, el cual proporcionará directrices en cuanto al método más adecuado de implementar.

1. ¿Conoce que es un clúster?
2. ¿Cuenta con algún tipo de clúster en su sistema? ¿De qué tipo?
3. ¿Este tipo de clúster es eficiente y se adecua a sus necesidades?
4. ¿Está conforme con el rendimiento del clúster?

5. Su sistema está orientado a: ¿brindar alta disponibilidad en los servicios? ¿brindar alto rendimiento en sus recursos?
6. En su sistema ¿existen puntos de falla?

Teniendo en cuenta las anteriores respuestas, el operador de telecomunicaciones tiene diferentes opciones que le permiten implementar o cambiar el tipo de clúster en su sistema. Un Clúster se puede dividir en diferentes categorías; a continuación se muestra la clasificación según su funcionalidad [94], ya que estos son considerados como los más relevantes para el presente trabajo de grado, debido a que se pueden implementar en el contexto de una NGSDP.

Clúster de Alto Rendimiento

Este tipo de clúster está formado por varios nodos, y sirve para resolver problemas complejos que requieren gran cantidad de recursos computacionales. Esto se logra interconectando los nodos entre sí, a través de redes de alta velocidad, obteniendo de esta manera un sistema de gran rendimiento que actúa como uno solo. Cualquier operación que necesite altos tiempos de CPU y millones de operaciones, puede ser utilizada en un clúster de alto rendimiento [87] [90].

Clúster de Alta Disponibilidad

Un clúster de alta disponibilidad tiene como objetivo brindar un alto grado de disponibilidad y confiabilidad a uno o varios servicios compartidos en los nodos. La característica principal de este clúster es que ante la existencia de algún problema o fallo de uno de los nodos, el resto asume dicho fallo y las tareas del nodo con problemas. Estos clúster disponen de herramientas con capacidad para monitorizar los servidores o los servicios caídos y automáticamente migrarlos a un nodo secundario para garantizar la disponibilidad del servicio. Estos mecanismos de alta disponibilidad se brindan de forma transparente y rápida para el usuario.

Este tipo de clúster es de mucha importancia y es una alternativa real en empresas que brindan servicios altamente disponibles, esto debido a que su principal función es estar controlando y actuando para que un servicio, o varios servicios, se encuentren activos a sus usuarios durante el máximo período de tiempo. Brindar alta disponibilidad no hace referencia a conseguir una gran capacidad de cálculo, si no lograr que los nodos funcionen en conjunto y que todos realicen la misma función.

La flexibilidad y robustez que poseen este tipo de clúster los hace necesarios en sistemas cuya funcionalidad principal es el intercambio masivo de información y el almacenamiento de datos sensibles, dónde se requiere que el servicio esté presente sin interrupciones.

Existen dos tipos de configuraciones en este tipo de clúster [90]:

- ***Configuración activo - pasivo:*** los nodos activos se encargan de ejecutar las aplicaciones, mientras que los nodos pasivos, actúan como respaldo.
- ***Configuración activo - activo:*** todos los nodos están activos para una o más aplicaciones, pero también sirven como respaldo para las aplicaciones que se ejecutan en otros nodos.

Si un nodo falla, las aplicaciones que se ejecutaban en él migran a uno de los nodos de respaldo.

Clúster de Balanceo de cargas

Un clúster de balanceo de carga permite que un conjunto de nodos compartan la carga de trabajo y de tráfico de sus clientes, es decir se divide el trabajo para todos los nodos que forman parte del clúster de tal manera que se obtiene un mejor tiempo de acceso a las aplicaciones y una mejor confiabilidad del sistema, además se aprovecha el paralelismo de los nodos para evitar que se presenten fallos [90].

4.3.6. Buenas Prácticas

Es la documentación de las experiencias exitosas dentro de una organización, que pueden ser replicadas en su conjunto o en alguna de sus partes. Una buena práctica es una técnica o metodología que a través de la experiencia y la investigación, ha demostrado conducir a un resultado fiable deseado, para garantizar el éxito [95].

El utilizar buenas prácticas en cualquier campo es un compromiso, el cual implica adoptar todos los conocimientos y la tecnología a disposición para poder realizar las prácticas sistematizadas del trabajo.

Las buenas prácticas se deben considerar como un elemento muy importante en el desarrollo de cualquier tipo de organización, ya que en primer lugar lo que se busca con estos procedimientos es aprender de los errores propios y de terceros. Las mejores prácticas provienen de las buenas y malas experiencias de otras organizaciones, las cuales reflejan el conocimiento adquirido en ambientes de producción y proporcionan la oportunidad de mejorar operaciones a un costo mínimo [96].

De acuerdo con la AP&QC (American Productivity & Quality Center)[97], existen tres principales obstáculos en la adopción de una buena práctica, los cuales son: i) la falta de conocimiento sobre las mejores prácticas actuales, ii) la falta de motivación para hacer los cambios necesarios para su adopción y iii) la falta de conocimientos y habilidades necesarias para hacerlo.

Existen varias recomendaciones que se deben seguir, para que cualquier organización pueda contar con unas buenas prácticas documentadas, que le permitan seguir avanzando tecnológicamente.

Por lo anterior y a fin de mejorar la disponibilidad y la capacidad de reacción ante desastres, los operadores de telecomunicaciones, deben plantear las siguientes preguntas al interior de la organización a fin de identificar cuáles son los procedimientos que necesitan mejores prácticas:

1. ¿En su organización, son documentados los procesos que se llevan al interior de la misma?
2. ¿Todos los procesos son documentados?
3. ¿Por qué no llevan a cabo la documentación de los procesos dentro de la organización?

4. ¿La documentación de los procesos fue creada en su empresa? ¿Pertenece a otro grupo que le facilito dicha información?
5. ¿Su empresa desea implementar buenas prácticas en todos los procesos que maneja?
6. ¿Estaría dispuesto a capacitar a todos sus empleados para que los procesos, sean documentados de manera satisfactoria?
7. Al interior de la organización, ¿se tienen claramente identificados las aplicaciones críticas y los procedimientos que deben ser llevados a cabo para restaurar, activar, y desplegar dichas aplicaciones?

Teniendo en cuenta las anteriores respuestas, y considerando que para un operador de telecomunicaciones que ya tenga o vaya a implantar una NGSDP, es muy importante contar con buenas prácticas, se le recomienda seguir los siguientes consejos generales [96]:

Comprensión del entorno: el personal implicado en la creación de las buenas prácticas dentro de una organización deben conocer su entorno, los diversos elementos que intervienen en él, y la interacción entre dichos componentes. Además deben ser capaces de describir con precisión y rapidez el entorno a un extraño para ayudar a acelerar los tiempos de respuesta de las solicitudes y a minimizar los errores por una comprensión incompleta o incorrecta del entorno. Esto proporciona el contexto necesario para que una persona pueda hacer recomendaciones sobre las alternativas en el desarrollo de las buenas prácticas.

Definir procedimientos: a fin de alcanzar buenas prácticas, deben estar documentados los eventos que pueden ocurrir dentro del entorno donde se encuentre la organización, por ejemplo el tiempo de paradas planeado y no planeado, una situación crítica, una rutina de tipo "Heads-up"[18] en las comunicaciones. Además se debe establecer en que formato va a quedar la documentación, por escrito, digital, o de las dos maneras y si va a existir algún tipo de repositorio donde encontrar la documentación.

Crear un plan IT-Risk[19]: se deben identificar los riesgos potenciales y sus impactos. Crear un plan de gestión de riesgos para aquellos entornos, aplicaciones, etc., que representen un impacto significativo para la empresa, o que son más probables de ocurrir. Identificar los disparadores para la ejecución de este plan de gestión de riesgos, ya que en momentos de crisis, la información es muy valiosa y se necesita llevar a cabo el plan predefinido. La ejecución de este tipo de planes llevan a obtener un mejor tiempo de respuesta, menor pérdida de datos y mejores decisiones.

Mantener acuerdos de apoyo para los sistemas clave: cualquier sistema que sea fundamental para la empresa debe contar con acuerdos de apoyo o con un sistema de reemplazo que pueden ser intercambiados con previo aviso.

Enfoque en la formación profesional del personal: las personas encargadas de apoyar los sistemas deben tener la capacidad y recursos disponibles para hacerlo. Deben estar muy

[18] Heads-up: llamada de advertencia frente a un peligro que se avecina.
[19] IT-Risk**:** riesgo relacionado con la tecnología de la información, es un término relativamente nuevo, es una faceta de una multitud de riesgos que son relevantes para el mundo TI y el mundo real.

familiarizados con el entorno y con las herramientas que necesiten para cumplir a cabalidad sus funciones. Además deben entender cuál es la interacción entre los diferentes componentes del sistema y saber qué hacer en caso de que se presenten inconvenientes o fallos. La formación del personal de la empresa trae grandes beneficios para la organización, ya que se cuenta con empleados capacitados en varias áreas, y no con individuos que solo conozcan un área específica**.**

Además para cada área específica existen pequeñas recomendaciones que pueden llegar a ser útiles cuando se desee implantar las buenas prácticas en una empresa [96]:

Administración del sistema: documentación acerca de instalación de parches y sistemas operativos, actualizaciones de software, correcciones en caliente, optimización, cambios de configuración, entre otros. Documentación sobre la identificación y solución de problemas de forma proactiva. Información sobre el mantenimiento del sistema, servicios, aplicaciones, y la frecuencia con la cual se deben realizar dichos mantenimientos. Etiquetar correctamente todos los componentes del sistema. Documentar los planes en caso de fallas en el sistema y contar con contratos de apoyo para asegurar que el sistema siempre este funcionando. Inspeccionar a diario el sistema en busca de luces de avería y/o escucha de sonidos inusuales. Mantener siempre al alcance de los empleados la información de apoyo disponible, incluyendo números de contactos, páginas web, números de licencia y de serie. También debe contar con la recopilación de los archivos de registros, de configuración y de comandos necesarios.

Seguridad / Control de Acceso: los usuarios deben contar con una cuenta única para el acceso al sistema, proporcionando un control y registro de las actividades que realice. Se debe además cambiar la contraseña de forma periódica. Las cuentas de usuarios que ya no pertenezcan a la empresa deben ser inmediatamente borradas.

Administración de bases de datos: documentación del mantenimiento de rutina a las bases de datos. En todas las tablas se deben revisar diariamente los datos que cambiaron y se deben generar estadísticas, esto garantiza la consistencia de la base de datos. Se debe tener la documentación de cómo realizar copias de seguridad y donde se deben guardar para evitar pérdidas de la información. Revisar el registro de errores diariamente. Documentación acerca de la nomenclatura y ubicación lógica con la cual se trabaja en el sistema, y debe realizarse de forma coherente en todos los ambientes. Se debe conocer lo que es normal para el sistema en el cual se está trabajando, lo cual hace más fácil diagnosticar un comportamiento anormal. Documentar muy bien los comandos del sistema operativo.

Aseguramiento de la Calidad: documentación sobre cómo se debe validar un nuevo módulo o ejecutable a fin de cumplir con los requisitos especificados en el sistema, incluyendo funcionalidad, rendimiento y concurrencia; estos requisitos no deben ser asumidos al azar.

Asistencia de Producción / Solución de problemas: se deben definir los SLAs para cada sistema y realizar la documentación que permita determinar lo que es aceptable para el entorno en el cual se encuentra el sistema, lo que hace falta y estimar un plan para lograr los objetivos propuestos. Documentación acerca de los eventos que generan problemas: sus indicios, como se generan, medidas de solución, resultados de pruebas, como se resolvió, causa del problema. Esta información es la que permite resolver los problemas que se presenten, sin necesidad de invertir demasiado tiempo buscando su solución. Se debe contar con documentación que le permita al

personal de apoyo identificar rápidamente los pasos a seguir para la solución adecuada del problema.

Todas y cada una de las buenas prácticas mencionadas anteriormente tienen un valor real y un valor añadido. El uso sistemático de estas buenas prácticas resultara en [98]: i) ubicación de la empresa en un alto rango dentro de la industria, ii) aprovechar y tomar ventaja de la tecnología, iii) mejorar la calidad, eficiencia y la satisfacción del cliente, iv) obtener bajos costos, v) dar más control y ejercer influencia en la gestión. Lo anterior se debe a la reducción en los errores y procesos asociados y a la minimización de las interrupciones del servicio no planificadas [96].

4.3.7. Requerimientos de resiliencia

Como se explicó en la sección 4.3.1., los requerimientos de resiliencia son el conjunto de aplicaciones, datos y entornos que deben ser protegidos; así, en caso de una eventual parada del sistema de producción estos continuarán disponibles, asegurando de esta forma la protección de la información de los operadores de telecomunicaciones y la eventual restauración de los servicios [29] [99].

A fin de identificar cuáles son los requerimientos de resiliencia necesarios para la organización, así como la técnica más adecuada para conseguir la misma, los operadores de telecomunicaciones deben plantear las siguientes preguntas al interior de la organización.

1. ¿Tiene completamente identificados cuales son los servicios críticos de su negocio y los sistemas que los soportan?
2. ¿Puede identificar los servicios críticos de su organización según el grado de importancia o el grado de riesgo que tienen para la organización?
3. ¿Salen todos los servicios de su organización mediante un único canal de comunicaciones?
4. ¿Conoce si los servicios críticos de su sistema se enrutan a través de diferentes componentes de la red, de forma que ante una eventual falla de uno de los componentes no se vean afectados esta clase de servicios?
5. ¿Son revisados periódicamente los requisitos de resiliencia a fin de mantener las prioridades de la organización al día?
6. ¿Se han discutido los planes de reacción de la organización ante una posible emergencia?

A continuación se muestran las diferentes alternativas con las que cuentan los operadores de telecomunicaciones a fin de implementar resiliencia en sus organizaciones.

Resiliencia de aplicaciones

La resiliencia de aplicaciones es clasificada según el efecto que provoca en los usuarios las paradas previstas o imprevistas del sistema. Este tipo de resiliencia controla mediante un programa de salida: el inicio, la parada, el reinicio y la conmutación de las aplicaciones a sistemas de copia de seguridad [29] [99].

Las diferentes alternativas a fin de implementar resiliencia de aplicaciones son:

Sin restauración de aplicaciones: en caso de una parada del sistema, los usuarios deben reiniciar manualmente sus aplicaciones. De acuerdo al estado de los datos los usuarios deben determinar en qué punto es necesario reiniciar el procesamiento dentro de la aplicación.

Reinicio automático de aplicaciones y el reposicionamiento manual dentro de las aplicaciones: mediante la utilización de procedimientos programados son reiniciadas las aplicaciones que estaban activas en el momento de una eventual parada. Basándose en el estado de los datos, el usuario debe determinar en qué punto se debe reanudar la aplicación.

Reinicio automático de aplicaciones y restauración semiautomática: tras una parada en el sistema, son reiniciadas las aplicaciones de forma manual, y los usuarios son devueltos a un punto de reinicio predeterminado dentro de la aplicación. Generalmente el punto de reinicio es coherente con el estado de los datos de la aplicación de resiliencia, sin embargo puede darse el caso en que el usuario necesite avanzar dentro de la aplicación a fin de hacer coincidir los datos de manera más precisa. Al iniciar sesión, la aplicación detecta el estado de cada usuario y determina si necesita restaurar la aplicación a partir del último estado salvado.

Reinicio automático de aplicaciones y restauración automática al último límite de transacción: al restaurarse el sistema se sitúa al usuario dentro de la aplicación en el punto de procesamiento que es coherente con la última transacción confirmada. Mediante esta técnica los datos de las aplicaciones y los puntos de reinicio de las aplicaciones coinciden plenamente.

Resiliencia de aplicaciones completa con reinicio automático y anomalía transparente: al ocurrir una parada, el usuario es situado en la última transacción confirmada, por lo cual este continúa en la última ventana donde ocurrió la parada. Mediante esta técnica no es necesario el reinicio de las secciones, no se pierden los datos. Es necesaria una relación cliente-servidor.

Resiliencia de datos

A continuación se presentan las técnicas para la implementación de resiliencia datos, las cuales pueden ser combinadas para obtener mejores resultados [29, 99].

Duplicación lógica: es una topología de resiliencia de datos de multisistema, la cual mediante software realiza duplicación de los objetos (archivos, miembros, áreas de datos de programas, etc.), en las copias de seguridad. La duplicación de los datos se realiza en tiempo real. Adicionalmente el uso de la duplicación lógica trae beneficios, como: visualización del estado de las duplicaciones en tiempo real, agregación automática de objetos recién creados a los objetos en proceso de duplicación, duplicación de subconjuntos de objetos en bibliotecas o directorios determinados, mediante el intercambio de roles se logra una rápida activación del entorno de producción en el entorno de copia de seguridad.

Dispositivo conmutable: un dispositivo conmutable se refiere a la recopilación de recursos hardware, como: dispositivos de cinta, adaptadores de comunicación, entre otros. Dichas recopilaciones pueden ser configuradas en agrupaciones especiales, las cuales son independientes del sistema principal, como resultado se consigue: menor tiempo en la conmutación de discos, simplicidad de funcionamiento, no existe latencia en la transmisión. Sin embargo este procedimiento también trae algunos retos como: pérdida de datos debido a la trasmisión

asíncrona, se puede presentar disminución en el rendimiento, restricciones relacionadas con hardware, complejidad en la configuración de los dispositivos de almacenamiento de acceso directo (datos y estructura de las aplicaciones).

Duplicación de sitios cruzados (XSM):

- ✓ ***Duplicación geográfica:*** los datos almacenados en copias de producción son duplicados en una segunda agrupación de discos independientes del sistema. Adicionalmente a las ventajas proporcionadas por la técnica de "dispositivo conmutable", proporciona recuperación ante desastres a una segunda copia en una ubicación diferente. Sin embargo se debe considerar degradaciones en el rendimiento del sistema primario y posibles aumentos en las sobrecargas de la unidad de procesamiento central, el cual es necesario para la duplicación geográfica.

- ✓ ***Duplicación metro:*** la transferencia de datos para la duplicación se realiza de forma simultánea, lo cual tiene limitaciones en cuanto a la distancia y a los requisitos de ancho de banda relacionada con los tiempos de transmisión.

- ✓ ***Duplicación global:*** la transmisión de los datos se realiza de forma asíncrona y es necesario el uso de software especializado para un tercer grupo de discos a fin de mantener coherencia en los datos. No existen límites en cuanto a la dispersión geográfica.

Resiliencia de entornos

La resiliencia de entornos se divide en dos secciones, el entorno físico y el entorno lógico. El entorno físico se centra en aspectos relacionados con el hardware de la red y la topología de la misma. El entorno lógico es donde se hospedan las aplicaciones y el entorno de ejecución, adicionalmente incluye aspectos como la configuración del sistema, perfiles de usuarios, entre otros [29].

Entorno físico: el entorno físico está conformado por programas necesarios para el adecuado mantenimiento de un entorno operativo del sistema. Adicionalmente incluye los servicios de programas de utilidad necesarios para la dirección de un centro de datos.

Entorno lógico: el entorno lógico es el entorno de tiempo de ejecución de aplicaciones. Está conformado por atributos del sistema, atributos de configuraciones de red, perfiles de usuarios, etc., los anteriores aspectos deben ser almacenados de manera cuidadosa, para que el entorno de aplicación funcione de forma idéntica en el sistema de copia de seguridad y en el sistema de producción primario. Se puede manejar la coherencia de estos valores a través de un dominio administrativo de clúster, la duplicación lógica o mediante procesos manuales bien definidos.

4.3.8. Presupuesto

El presupuesto es el cálculo anticipado de los ingresos y gastos de una actividad económica, durante un periodo determinado de tiempo y bajo ciertas condiciones previstas. Es un plan de acción dirigido a cumplir una meta prevista, expresada en valores y términos financieros; los presupuestos son programas en los que se les asignan cifras a las actividades e implican una

estimación de capital de los costos, de los ingresos, y de las unidades o productos requeridos para lograr los objetivos propuestos. Las finalidades principales del presupuesto consisten en determinación de la mejor forma de utilización y asignación de los recursos y a la vez controlar las actividades de la organización en términos financieros [100].

Elaborar un presupuesto permite a las empresas y/o las organizaciones establecer prioridades y evaluar la consecución de sus objetivos.

Un presupuesto debe tener las siguientes características:

- ✓ Debe ser un documento formal ordenado sistemáticamente.
- ✓ Deber presentar un plan expresado en términos cuantitativos.
- ✓ Debe ser general, porque se establece para toda la empresa.
- ✓ Debe ser específico, porque puede referirse a cada una de las áreas de la organización.

Además los presupuestos son herramientas fundamentales para un negocio ya que permiten planificar, coordinar y controlar operaciones como [101]:

- **Planeación**: los presupuestos permiten planificar actividades, objetivos, recursos, estrategias; anticipándose a los hechos y ayudando a reducir la incertidumbre y los cambios.

- **Coordinación**: los presupuestos sirven como guía para coordinar actividades, permitiendo armonizar e integrar todas las secciones o áreas del negocio con los objetivos de la empresa.

- **Control**: los presupuestos sirven como instrumento de control y evaluación, permiten comparar los resultados obtenidos con los presupuestados para saber en qué áreas o actividades existen variaciones o diferencias entre lo obtenido y lo presupuestado.

El presupuesto es muy importante para los operadores de telecomunicaciones, ya que estos hacen parte de un medio económico en el que predomina la incertidumbre, y si desean sostenerse en el mercado competitivo deben realizar un buen presupuesto, que les permita estar entre las organizaciones con mayor éxito.

En una solución de alta disponibilidad en una NGSDP, debe existir también un presupuesto; el costo de la solución debe compararse con las ventajas que se pueden obtener en las empresas u organizaciones. Pero, aunque desde el punto de vista técnico, existen soluciones excelentes de alta disponibilidad con un tiempo de inactividad igual a cero, su valor puede ser demasiado elevado. Es por esto que dependiendo también del valor del presupuesto de la organización se puede dar un estimado de la disponibilidad que teóricamente se podría alcanzar. El costo de la solución incluye, el costo inicial para proporcionar y desplegar la solución, los costos continuos del uso de la solución, y los impactos costo/rendimiento [29].

Por lo anterior, los operadores de telecomunicaciones que pretendan brindar servicios de alta disponibilidad en una plataforma NGSDP, deben contar con un presupuesto adecuado, para las soluciones que desee implementar dentro de la empresa, que implique una buena relación entre lo que se invierte y lo que se obtiene (costo/beneficio); para esto deben plantear unos

interrogantes que le permitirán conocer con qué tipo de presupuesto cuenta, y percatarse si es el más apropiado para el modelo de negocios que se maneja en la organización, o si existen otros tipos de presupuestos que pueden ser adoptados:

1. ¿En su organización existen presupuestos para cada área?
2. ¿El presupuesto puede ser modificado dependiendo de la solución?
3. ¿El presupuesto es único e inmodificable?
4. ¿El presupuesto se realiza a plazos?
5. ¿La organización tiene en cuenta el modelo de negocios para definir el presupuesto?
6. ¿Los presupuestos de periodos anteriores, han resuelto de forma satisfactoria los problemas que se afrontan en la empresa?
7. ¿Se deben ajustar a un presupuesto inmodificable cada periodo?
8. ¿Su empresa adquiere soluciones que traen beneficios a corto plazo?
9. ¿Su organización estaría dispuesta a hacer una mayor inversión, en soluciones que a largo plazo traigan mayores beneficios?

Teniendo en cuenta las anteriores respuestas, se deben considerar las siguientes clases de presupuesto [101]:

Según la flexibilidad

- ✓ **Presupuesto rígido, estático, fijo o asignado:** son aquellos que se elaboran para un único nivel de actividad y no permiten realizar ajustes necesarios cuando se presenta alguna variación en el modelo de negocio. Se basa fundamentalmente en que las estimaciones de los pronósticos son correctas. Dejan de lado el entorno de la empresa (económico, político, cultural, etc.). Pueden ser utilizados cuando los pronósticos sobre el futuro de la empresa son altamente confiables.

- ✓ **Presupuesto flexible o variable:** son los que se elaboran para diferentes niveles de actividad y se pueden adaptar a las circunstancias cambiantes del entorno. Son dinámicos adaptativos, pero complicados y costosos. Su característica es que evita la inflexibilidad del presupuesto estático, transformándolo en un instrumento dinámico con varios niveles de operación para conocer el impacto sobre los resultados pronosticados de cada rango de actividad, como consecuencia de las distintas reacciones de los costos frente a aquellos.

Según el periodo

- ✓ **Presupuesto a corto plazo:** son los que se realizan para cubrir la planeación de la organización en el ciclo de operaciones de un año.

- ✓ **Presupuesto a largo plazo:** este tipo de presupuestos corresponden a los planes de desarrollo que generalmente adoptan los estados y grandes empresas.

Según el campo de aplicabilidad

- ✓ **De operación o económicos:** incluye el presupuesto de todas las actividades para el período siguiente al cual se elabora y cuyo contenido se resume en un estado de pérdidas y ganancias proyectado, estos son:
  - Ventas
  - Producción
  - Gastos
  - Mano de Obra
- ✓ **Financieros:** incluye el cálculo de partidas y/o rubros que inciden fundamentalmente en el balance.
  - Presupuesto de Tesorería.
  - Presupuesto de Erogaciones Capitalizables.

Según el sector

- ✓ **Presupuestos del Sector Privado:** usados por las empresas particulares; se conocen también como presupuestos empresariales. Buscan planificar todas las actividades de una empresa; en este tipo de presupuesto se fijan metas específicas y se asignan los recursos necesarios para su consecución. Generalmente busca las ganancias en beneficio del negocio para incrementar la riqueza de los propietarios.
- ✓ **Presupuesto del Sector Público:** involucran planes, políticas, programas, proyectos, estrategias y objetivos del Estado. Son el medio más efectivo de control del gasto público y en ellos se contempla las diferentes alternativas de asignación de recursos para gastos e inversiones.

También existen algunas sugerencias que permiten al operador de telecomunicaciones mejorar el presupuesto; estas sugerencias son [102]:

- Ser flexible con las soluciones que impliquen un aumento en el presupuesto.
- Los objetivos deben dirigir el presupuesto; el presupuesto no debe determinar los objetivos.
- Utilizar software que le permita simular el presupuesto.
- Recordar que los presupuestos son herramientas y que las utilidades resultan de una administración inteligente.

4.3.9. Construcción en Redundancia

La construcción en redundancia de un sistema, se refiere a la duplicación de componentes hardware, software, información, etc., a fin de incrementar la fiabilidad, tolerancia a fallos y la disponibilidad de los sistemas. Para este propósito se deben identificar claramente cuáles son los sistemas que realmente requieren duplicación, esta clasificación depende de la importancia y la pérdida de dinero asociada que esta genera al no estar disponible; no es necesario tener componentes redundantes si el costo asociado es mayor al de las pérdidas que se generan por la indisponibilidad de los mismos.

A continuación, se presenta un pequeño cuestionario guía, el cual proporciona directrices en cuanto a la importancia de la redundancia y hacia las técnicas más adecuadas para la implementación de la misma.

1. ¿Actualmente está implementando redundancia en su organización?
2. ¿Tiene claramente identificada las pérdidas que generan los momentos de indisponibilidad para la organización?
3. ¿Tiene claramente identificados los benéficos que trae para la organización la implementación de redundancia en el sistema?
4. ¿Tiene claramente identificados los componentes críticos de su organización? ¿Se le realiza a estos elementos algún tipo de redundancia?
5. La redundancia implementada en su organización, ¿funciona de forma correcta? ¿se obtienen los resultados esperados?
6. ¿Estaría dispuesto a implantar, cambiar o completar la solución de redundancia de su organización a fin de incrementar la fiabilidad y disponibilidad de sus servicios?, ¿Cómo sería la relación costo/beneficio de las soluciones?

Teniendo en cuenta las anteriores respuestas, el operador de telecomunicaciones debe escoger el tipo de redundancia más adecuado para su caso específico. A continuación se presentan diversos métodos que deben ser tenidos en cuenta a fin de lograr componentes redundantes en los sistemas.

Redundancia en discos

Los discos al ser los componentes donde se almacenan los datos, son de vital importancia para todas las organizaciones. El propósito de este tipo de redundancia es la creación de un arreglo de discos, el cual permite el almacenamiento de la información aun en el evento de que uno de los discos falle y el remplazo de los mismos sin la necesidad de detener el sistema. Entre las técnicas más utilizadas se encuentra: “RAID (Redundant array of independent disks)” con la cual se crean conjuntos de discos redundantes, los cuales aumentan la velocidad y el rendimiento del sistema [103] [104].

Redundancia en componentes de red

A fin de aumentar la fiabilidad de los sistemas, se debe prestar especial atención en la identificación de los componentes críticos de la red, ya que de nada sirve poseer servidores altamente disponibles, si estos no van acompañados de componentes de red fiables; estos componentes son [103] [104]: enrutadores (routers), interruptores (switches), cables, líneas de conexión.

Para los sistemas de tiempo real, existen técnicas de redundancia especializadas, como se muestra a continuación.

Redundancia hardware

Para el caso de la redundancia hardware se pueden identificar dos tipos de técnicas redundantes, la redundancia estática y la redundancia dinámica [105].

Redundancia estática: consiste en la utilización de componentes idénticos redundantes a fin de evitar posibles fallos. Como ejemplo típico de este tipo de redundancia se puede nombrar el caso de duplicación o triplicación de determinados componentes, los cuales mediante el uso de software especializado son comparados continuamente, y en caso de presentarse una anomalía en la información de salida, es decir que una de las salidas difiera de las otras, se bloquea la salida (Figura 31) [105].

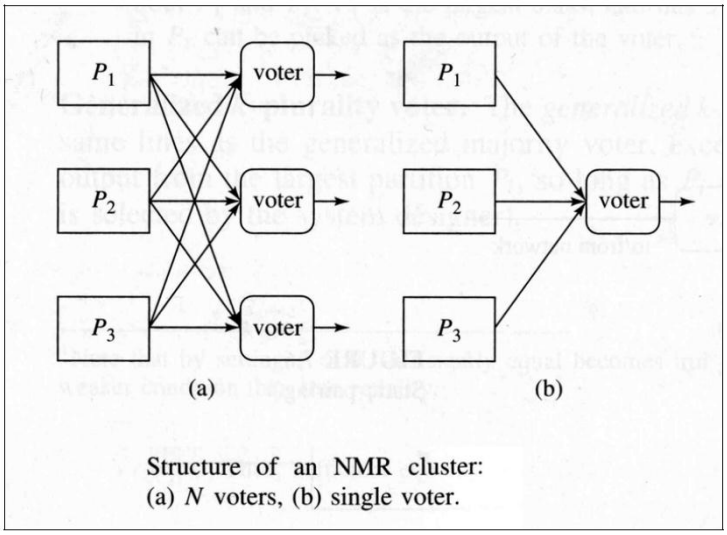

Figura 31. Redundancia estática [105]

Redundancia dinámica: en este tipo de redundancia, los propios componentes determinan si existe alguna anomalía en la información de salida (detectores de paridad), es decir que deben poseer la capacidad de detectar errores, sin embargo la recuperación de los errores está a cargo de otros sistemas.

Redundancia software

Esta clase de redundancia se refiere a la realización de software tolerante a fallos, lo cual sin duda es de gran importancia debido a que como se mostró anteriormente, es mayor la tasa de errores producidos por deficiencias del software que por el hardware. Las diferentes técnicas a fin de conseguir software redundante para tiempo real se muestra a continuación [105].

Programación de N-versiones: se basa en la generación independiente de N programas, los cuales prestan funciones equivalentes, sin embargo son construidos bajo diferentes lenguajes y entornos de programación a fin de evitar posibles fallas relacionadas con los leguajes. Al finalizar la construcción de los programas, estos son puestos en marcha acompañados de un programa director que compara los datos arrojados por los mismos, es decir que cada programa realiza sus respectivos cálculos, los cuales son enviados al director quien compara los mismos y en caso de diferir, suspende el funcionamiento (Figura 32). La forma de trabajo de este tipo de redundancia es muy parecida al de la redundancia estática hardware.

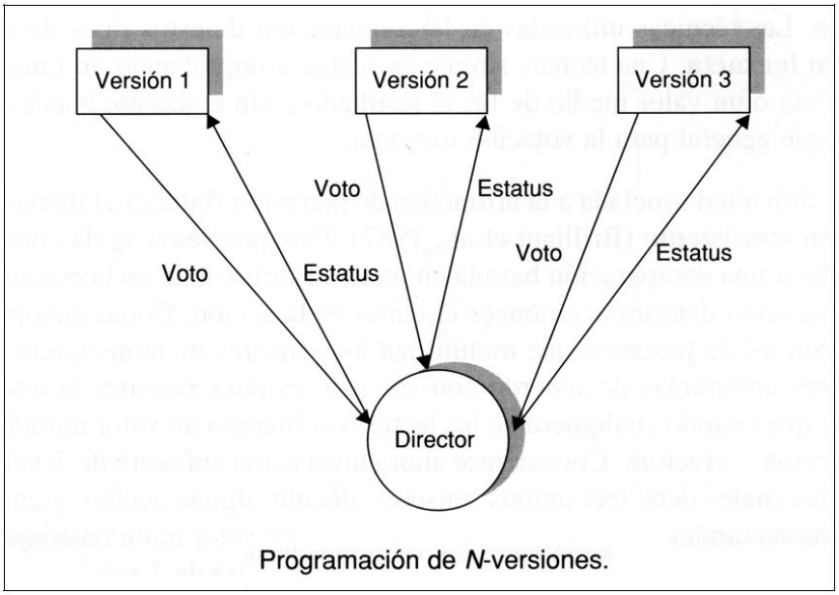

Figura 32. Redundancia por programación de N-versiones [105]

Detección y recuperación de errores: su funcionamiento es análogo a la redundancia dinámica hardware. Los bloques de recuperación poseen en su entrada un punto de recuperación automática y en su salida se encuentra un test de aceptación, el cual verifica que el sistema se encuentre en un estado aceptable después de la ejecución del bloque, en caso de no estarlo se regresa al punto de recuperación del bloque. En la Figura 33 se muestra claramente el funcionamiento de esta técnica [105].

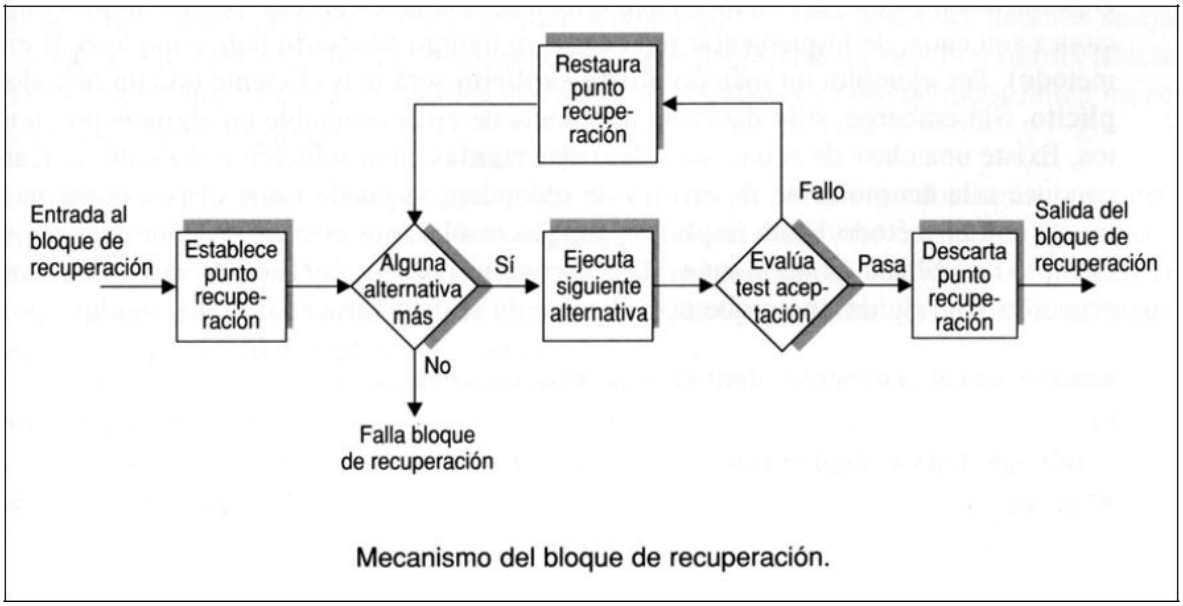

Figura 33. Mecanismo de bloque de recuperación [105]

4.3.10. Rendimiento

Es el desempeño deseado del sistema; a partir de este se trazan los requerimientos para cada una de las áreas y elementos dentro de una organización.

Dentro de un operador de telecomunicaciones que desee contar con un alto rendimiento, se deben utilizar tecnologías que garanticen una alta tasa efectiva de llamadas (throughput), baja latencia y control de sobrecarga [32]. Así, muchas organizaciones saben qué tipo de tecnología utilizar pero no saben cómo medir el rendimiento de su sistema [106].

El rendimiento es con frecuencia olvidado hasta que un cliente informa un problema. En otros casos, el rendimiento se empieza a evaluar en el despliegue inicial. Es por esto que los problemas de rendimiento no suelen ser evidentes dentro de una organización.

La medición del rendimiento del sistema cuando se coloca bajo cargas cada vez más altas determina su escalabilidad. Cuando el rendimiento empieza a caer por debajo de los requisitos mínimos establecido, se ha alcanzado el límite de la escalabilidad. Un rendimiento malo, se traducirá en una aplicación que funciona mal.

El rendimiento se puede ver afectado en el tiempo de diseño y en el tiempo de ejecución. En tiempo de diseño, se debe evitar la introducción de problemas que puedan afectar el rendimiento del sistema; esto se logra siguiendo las buenas prácticas aceptadas y al aprovechamiento de los resultados ya obtenidos. Un método común para identificar los problemas es llevar a cabo revisiones de rutina. Los obstáculos encontrados y corregidos en esta fase suelen ser más económicos y fáciles de reparar. En tiempo de ejecución, el sistema debe someterse a pruebas de rendimiento obligatorias para identificar cuellos de botella y así como la contención de los

recursos. Antes de que una solicitud se someta a extensas pruebas de rendimiento, es fundamental que todas las pruebas funcionales se hayan completado. En pocas palabras, un sistema debe trabajar antes de que pueda funcionar bien. Sin embargo, las pruebas de rendimiento deben comenzar tan pronto como sea posible, con el fin de identificar las áreas problema [107].

Existe también la Ingeniería de rendimiento, ya que es necesario integrar una cultura de rendimiento en el ciclo de vida de un sistema, y para esto se necesita un proceso a seguir, para saber exactamente por dónde empezar, cómo proceder y saber cuándo se finaliza. El modelado de rendimiento ayuda a aplicar la disciplina en los procesos. El enfoque fundamental es establecer objetivos y medir el progreso hacia esos objetivos. Se debe realizar una medición continúa durante todo el ciclo de vida para ayudar a determinar si la organización se está acercando a los objetivos de rendimiento o se encuentra muy lejos de alcanzarlos [107].

El rendimiento afecta los roles dentro de una organización de diferentes maneras:

- Un arquitecto, debe equilibrar el rendimiento y la escalabilidad con los atributos de calidad de servicio (QoS), como la administración, interoperabilidad, seguridad, y facilidad de mantenimiento.
- Un desarrollador, debe conocer por dónde empezar, cómo proceder, y saber cuando el software se ha optimizado lo suficiente.
- Un controlador de calidad, debe validar si el sistema y/o aplicación es compatible con las cargas de trabajo previstas.
- Un administrador, debe saber cuándo un sistema no cumple con los SLAs, y tiene que ser capaz de crear planes de crecimiento efectivos.
- Una organización, necesita saber cómo gestionar el rendimiento a través del ciclo de vida del software.

En la Figura 34 se muestran los elementos requeridos para la Ingeniería de rendimiento.

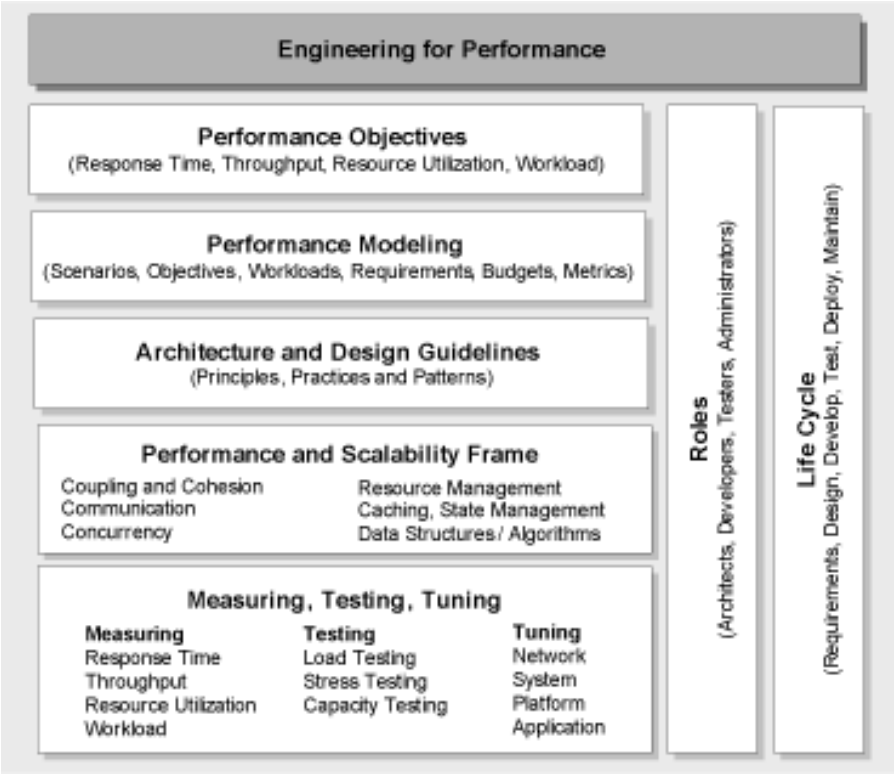

Figura 34. Ingeniería de rendimiento [105]

Así, la Ingeniería de rendimiento se divide en las siguientes categorías y áreas de responsabilidad [107]:

- Objetivos de rendimiento: permiten saber cuándo el sistema se ajusta a las metas de rendimiento.
- Modelado de rendimiento: proporciona un enfoque estructurado que se puede repetir, para cumplir con los objetivos de rendimiento.
- Arquitectura y Guías de diseño: permiten al ingeniero obtener un rendimiento desde el principio.
- Marco de rendimiento y escalabilidad: permiten organizar y priorizar los problemas de rendimiento.
- Medición: permite ver si el sistema está alejado de los objetivos de rendimiento.
- Roles y Ciclo de vida: Proporcionar una segmentación clara de roles para los arquitectos, desarrolladores, controladores de calidad y administradores, que les permitan comprender sus responsabilidades dentro del ciclo de vida de una aplicación y/o sistema.

Existen varias recomendaciones desde el punto de vista de la Ingeniería de rendimiento, que se deben seguir, para que cualquier tipo de organización pueda contar con un alto rendimiento en el sistema y/o aplicaciones.

Por esto los operadores de telecomunicaciones que deseen tener un alto rendimiento deben plantearse las siguientes preguntas:

1. ¿Conoce el rendimiento de su sistema?
2. ¿Este es el rendimiento deseado?
3. ¿Con qué rapidez es necesario que la aplicación se ejecute?
4. ¿En qué momento el rendimiento de su aplicación y/o sistema se vuelve inaceptable?
5. ¿Qué tanta memoria, CPU y disco duro puede consumir la aplicación?

Teniendo en cuenta las respuestas a las anteriores preguntas, y considerando que para un operador de telecomunicaciones que ya tenga o vaya a implementar una NGSDP, es muy importante contar con un alto rendimiento; se le recomienda seguir los siguientes consejos en cada una de las áreas de la Ingeniería de rendimiento; estos se muestran a continuación [107]:

1. Establecer objetivos de rendimiento

Los objetivos de cualquier tipo de proyecto deben incluir objetivos de rendimiento que se puedan medir. Los objetivos de rendimiento se suelen especificar en términos de:

- Tiempo de respuesta: cantidad de tiempo que le toma a un servidor responder a una solicitud.
- Rendimiento: número de solicitudes que pueden ser atendidas por la aplicación por unidad de tiempo. El rendimiento se mide frecuentemente como las solicitudes u operaciones lógicas por segundo o minuto.
- Utilización de recursos: medida de los recursos hardware, software y de red, que son consumidos por la aplicación. Los recursos incluyen la CPU, memoria, disco duro, y la red.
- Carga de trabajo: incluye el número total de usuarios y usuarios activos concurrentes, volúmenes de datos, y los volúmenes de intercambio.

2. Métricas

Las métricas son los criterios que se utilizan para medir los escenarios con los objetivos de rendimiento. Por ejemplo, una métrica podría ser el tiempo de respuesta, el rendimiento y la utilización de recursos. El objetivo de rendimiento para cada métrica es un valor aceptable. Se deben comparar el valor actual de las métricas para verificar que los objetivos de rendimiento se están cumpliendo, excediendo o fallando.

3. Diseño para mejorar el rendimiento

Se debe contar con un buen diseño desde el inicio de cualquier proyecto, ya que muchos de los problemas de rendimiento se introducen en las elecciones de arquitecturas, diseños, y tecnologías, que se realizan en el ciclo de desarrollo, en la etapa del diseño.

4. Marco de Rendimiento y Escalabilidad

Se recomienda seguir el siguiente marco de rendimiento y escalabilidad, el cual ayuda a organizar y priorizar los problemas de rendimiento y escalabilidad. Las categorías dentro del marco son un conjunto de prioridades, que son generalizadas en todas las aplicaciones (Tabla 7).

Categoría	Consideraciones clave
Acoplamiento y cohesión	Perdida de acoplamiento y alta cohesión.
Comunicación	Mecanismo de transporte, limites, diseño de la interfaz remota, round trips, serialización, ancho de banda.
Concurrencia	Transacciones, bloqueos, enlaces, colas.
Gestión de los recursos	Asignar, crear, destruir, poner en común.
Almacenamiento en caché	Por usuario, aplicación a escala, volatilidad de los datos.
Estado de gestión	Por usuario, aplicación a escala, persistencia, ubicación.
Estructuras de datos y algoritmos	Elección del algoritmo. Arreglos vs colecciones.

Tabla 7. Categorías de rendimiento

5. Validar las hipótesis

Es necesario validar las suposiciones que se hayan realizado. Cuanto más lejos se esté en el ciclo de vida del proyecto, mayor será la precisión de la validación. Al principio, la validación se basa en los puntos de referencia disponibles, pero después se puede medir en lo que ya se encuentra desplegado.

6. Escenarios

Los escenarios son importantes desde la perspectiva del rendimiento, ya que ayudan a identificar las prioridades y a definir y aplicar las cargas de trabajo. Si se cuenta con documentación de casos de uso o historias de usuarios, esto puede ayudar a definir el escenario que se necesita. Los escenarios críticos pueden tener objetivos específicos de rendimiento, o pueden afectar a otros escenarios críticos.

7. Ciclo de Vida

Se presenta un enfoque de ciclo de vida basado en el rendimiento, y provee una guía que aplica a todos los roles que intervienen en el ciclo de vida, incluyendo arquitectos, diseñadores, desarrolladores, controlador de calidad y administradores. La Figura 35 muestra cómo la guía se aplica a las categorías generales asociados con el ciclo de vida de una aplicación.

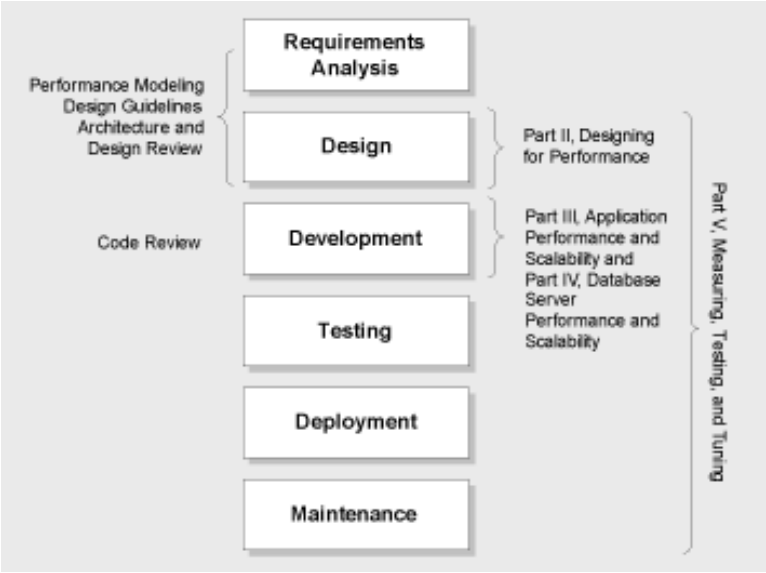

Figura 35. Ciclo de vida [107]

El rendimiento es integrado en estas etapas de la siguiente manera [107]:

Reunir los requisitos: se empieza a definir objetivos de rendimiento, flujos de trabajo y escenarios clave. Se consideran las cargas de trabajo y los volúmenes estimados para cada escenario. Si es necesario se puede empezar a evaluar el rendimiento en esta etapa.

Diseño: las decisiones de diseño deben basarse en principios probados y patrones. El diseño debe ser revisado desde una perspectiva de rendimiento. La medición debe continuar durante todo el ciclo de vida.

Desarrollo: se debe empezar a capturar métricas reales para validar las hipótesis formuladas en la fase de diseño. Se debe tener cuidado de mantener un enfoque equilibrado durante el desarrollo.

Pruebas: las pruebas de tensión y de carga se utilizan para generar indicadores y para verificar el comportamiento de la aplicación y rendimiento bajo condiciones de carga normal y pico.

Despliegue: durante la fase de despliegue, validar el modelo mediante el uso de métricas de producción. Se pueden validar las estimaciones de carga de trabajo, los niveles de utilización de recursos, el tiempo de respuesta y el rendimiento.

Mantenimiento: se debe continuar midiendo y monitoreando cuando una aplicación se implementa en el entorno de producción. Los cambios que pueden afectar el rendimiento del sistema incluyen el aumento de cargas de usuarios, el despliegue de nuevas aplicaciones sobre la infraestructura compartida, las revisiones del software del sistema, y actualizaciones de la aplicación para ofrecer mayores o nuevas funcionalidades. `

Capítulo 5

Prototipo y Evaluación

5. Introducción

En el presente capítulo se resumen los resultados del proceso de evaluación de los lineamientos planteados en el capítulo anterior, de acuerdo al diagnóstico realizado para el operador de telecomunicaciones EMCALI en el contexto de un servicio de prueba; en este sentido y con propósitos de comparación, se generaron dos escenarios: el primero, sin la implementación de criterios y lineamientos, y el segundo con la implementación de los mismos. En ambos casos fue calculado el número de peticiones que se pueden resolver por segundo, tiempos de servicio y los porcentajes de disponibilidad.

5.1. Entrevista de diagnóstico

Con el fin de determinar los requerimientos para alcanzar alta disponibilidad en un ambiente de telecomunicaciones real, fue diseñada la entrevista para evaluar los requerimientos de la alta disponibilidad; dicha entrevista fue realizada a algunos de los funcionarios expertos de la empresa de telecomunicaciones EMCALI; como resultado de la entrevista se generó un diagnóstico, el cual consiste en: el conjunto de aspectos técnicos (criterios) que deben ser mejorados por la organización y los lineamientos que se consideran deben ser implementados a fin de mejorar los mismos. Los detalles de este proceso pueden ser consultados en el Anexo B.

A continuación se presenta de forma resumida el diagnóstico obtenido a partir de la información capturada durante la entrevista:

- La empresa cuenta con las características hardware necesarias para atender el número de usuarios con los que cuenta actualmente. Sin embargo, cuando se genera un alto número de peticiones al sistema, se presentan considerables periodos de inactividad en la prestación de los servicios o periodos en los cuales la calidad en los servicios prestados no es la mejor.

- Existe una fuerte falencia en cuanto a la documentación de los procedimientos necesarios para la resolución de problemas propios de la organización; sólo se cuenta con los procedimientos proporcionados por las empresas de las cuales EMCALI adquiere equipos o servicios.

- La empresa realiza copias de seguridad, pero sólo a las bases de datos y a algunos servidores. Sin embargo, elementos de gran importancia para la organización como el servidor de medios, no cuenta con copias de respaldo.

- La empresa cuenta con clúster hardware, sin embargo cuando se presentan periodos en los que se generan múltiples peticiones al sistema, el rendimiento del mismo no es el indicado.

De acuerdo al análisis anterior, se sugiere el siguiente plan de mejoramiento en el marco de los criterios y lineamientos planteados en este trabajo de grado:

- Es necesario mejorar el manejo de buenas prácticas al interior de la organización. Para esto, deben ser generados documentos propios en los cuales se explique de forma detallada los problemas técnicos que se han presentado y la forma en la que han sido resueltos, manuales de instalación (equipos, software, aplicaciones, servicios, entre otros.) y de despliegue de aplicaciones y servicios, detalles técnicos propios de la organización a tener en cuenta, etc.

- A fin de mejorar el desempeño del balanceador de cargas y por consiguiente la disponibilidad del sistema, se sugiere la implementación de un balanceador de cargas software que complemente el balanceador de cargas hardware existente.

- Con el fin de mejorar la fiabilidad del sistema y en especial la del servidor de medios se recomienda la implementación de copias de seguridad automáticas e incrementales a dicho servidor.

- Dado el hecho que EMCALI posee potentes medios hardware y a fin de mejorar la fiabilidad y disponibilidad del sistema, se recomienda complementar el sistema de clúster hardware que posee la compañía con sistemas de clúster software (virtualización).

Por lo tanto los criterios y lineamientos que deben ser implementados por el operador de telecomunicaciones EMCALI son (Tabla 8) (Anexo B):

Criterio	Lineamiento
Buenas prácticas	- Recomendaciones sección 4.3.6.
Copias de seguridad	- Automáticas e incrementales
Balanceo de Cargas	- Balanceo de cargas SIP. - Balanceo de cargas DNS – Round Robin – HTTP.
Clúster	- Clúster de alta disponibilidad en configuración activo - activo (virtualización de nodos)

Tabla 8. Criterios y lineamientos a implementar

Con el fin de evaluar los anteriores lineamientos y de acuerdo al contexto del presente trabajo de grado, es necesaria la implementación de un prototipo de laboratorio, que permita simular una red de telecomunicaciones y una NGSDP que posibilite la interacción con la Web; en dicho prototipo será desplegado un servicio convergente, el cual será sometido a diferentes pruebas a fin de determinar la disponibilidad del mismo antes y después de la implementación de los lineamientos.

No obstante, para el caso práctico del prototipo de laboratorio, no se cuentan con las herramientas necesarias para evaluar el correcto funcionamiento de las copias de seguridad y de

las buenas prácticas; por esta razón los criterios y lineamientos que se implementarán y evaluarán en el prototipo de laboratorio son: clúster de alta disponibilidad en configuración activo - activo y balanceo de carga software en las modalidades: SIP y HTTP.

5.2. Descripción del prototipo de laboratorio

Con el fin de implementar y evaluar los lineamientos anteriormente seleccionados, se construyó un prototipo de laboratorio que permite simular una red de telecomunicaciones (capacidades de control IMS) y una NGSDP que posibilite la interacción con la Web y el despliegue de servicios convergentes.

A continuación se muestra, la arquitectura de referencia que se utilizará, las herramientas utilizadas para la implementación de la arquitectura y el servicio convergente que se desplegará sobre la misma.

5.2.1. Arquitectura de Referencia

La arquitectura a utilizar para la implementación del prototipo, el despliegue del servicio convergente y las posteriores pruebas de rendimiento, consta de los siguientes elementos:

Capa de aplicación: conformada por un servidor de aplicaciones JAIN SLEE.

Capa de control: conformada por los elementos básicos de la arquitectura IMS/NGN (CSCF-HSS).

Capa de acceso: conformada por terminales computacionales desde donde se generan las peticiones Web y por teléfonos SIP para la recepción y finalización de las llamadas.

Los componentes de la arquitectura, se proporcionan en la siguiente sección.

5.2.2. Implementación de la arquitectura

A continuación se presentan los elementos que serán utilizados para la implementación de cada una de las capas de la arquitectura de referencia y que permitirán el despliegue del servicio convergente

5.2.2.1. Plataforma de comunicaciones Mobicents - Capa de aplicación

La capa de aplicación es conformada por la plataforma de comunicaciones Mobicents, debido a que es un servidor de aplicaciones de libre distribución. Es una plataforma orientada a objetos, además de estar conformada por diversos componentes posee un entorno de ejecución tolerante a fallos. La arquitectura de esta plataforma fue diseñada para posibilitar la creación, despliegue y gestión de servicios y aplicaciones que involucran: voz, video y datos en redes IP y de comunicaciones. Los componentes que permiten lograr la convergencia en la plataforma de comunicaciones Mobicents son (Figura 36) [108] [31] [109]:

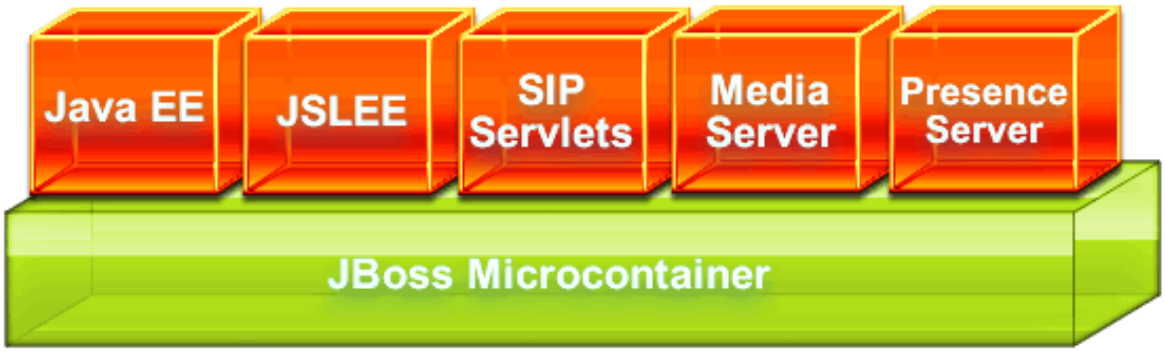

Figura 36. Servidor de aplicaciones Mobicents

- **Jain Slee:** Mobicents mediante su contenedor JSLEE, realiza una implementación de la especificación JAIN SLEE v1.1 (JSR 240) de libre distribución; brinda un modelo de componentes y un entorno de ejecución robusto para aplicaciones de telecomunicaciones, el cual es un complemento de J2EE para permitir la convergencia de datos, voz y video en las aplicaciones de nueva generación.

- **Sip Servlets:** es una distribución libre de la especificación SIP Servlets v1.1 (JSR 289) que permite desarrollar y desplegar servicios SIP y JEE (Java Enterprise Edition) portables y convergentes; se caracteriza por brindar ejecuciones eficientes y seguras.

- **Media Server:** es un componente de libre distribución que provee una funcionalidad completa y competitiva de una pasarela de contenido que satisface las necesidades de las redes inalámbricas, cableadas, fijas y móviles IP convergentes, desde una única plataforma de contenido.

- **Presence Server:** es una implementación de libre distribución que permite la gestión de los documentos XML de los usuarios, tales como reglas de autorización de presencia, listas de contactos y de grupos, información de presencia estática, entre otros.

- **JBoss Microcontainer:** es el entorno de almacenamiento en el cual residen los componentes anteriormente nombrados, adicionalmente provee servicios de configuración, registro, empaquetamiento, entre otros.

En el entorno de las telecomunicaciones Mobicents está posicionado como un núcleo de alto rendimiento para las plataformas SDP; permite la composición de bloques SBB (Service Building Block), como: control de llamada, facturación, administración, capacidades de detección de presencia, entre otros. Sin embargo la plataforma de comunicaciones de Mobicents puede ser aplicada a entornos diferentes como: transacciones financieras, juegos online, control distribuido, etc., que requieren alto grado de señalización y baja latencia.

Para la implementación del prototipo, será utilizado el servidor JAIN SLEE de la plataforma Mobicents, así como el servidor de medios de la misma, esto debido a las razones presentadas en la sección 3.4.1.3 del presente documento.

5.2.2.2. OpenIMSCore - Capa de control

Para la capa de control fue seleccionado OpenIMSCore, debido a que es una implementación de libre distribución de los elementos CSCF (Call Control Session Control Functions) de la arquitectura IMS y una versión ligera del componente HSS (Home Subscriber Server) de la misma, los cuales en conjunto conforman los elementos básicos de la arquitectura IMS/NGN (Figura 37), según las especificaciones del 3GPP, 3GPP2, ETSI TISPAN. Los cuatro componentes del OpenIMSCore (P-CSCF, S-CSCF, I-CSCF y HSS) estan basados en código abierto (MySQL, SER (SIP Express Router)) [110].

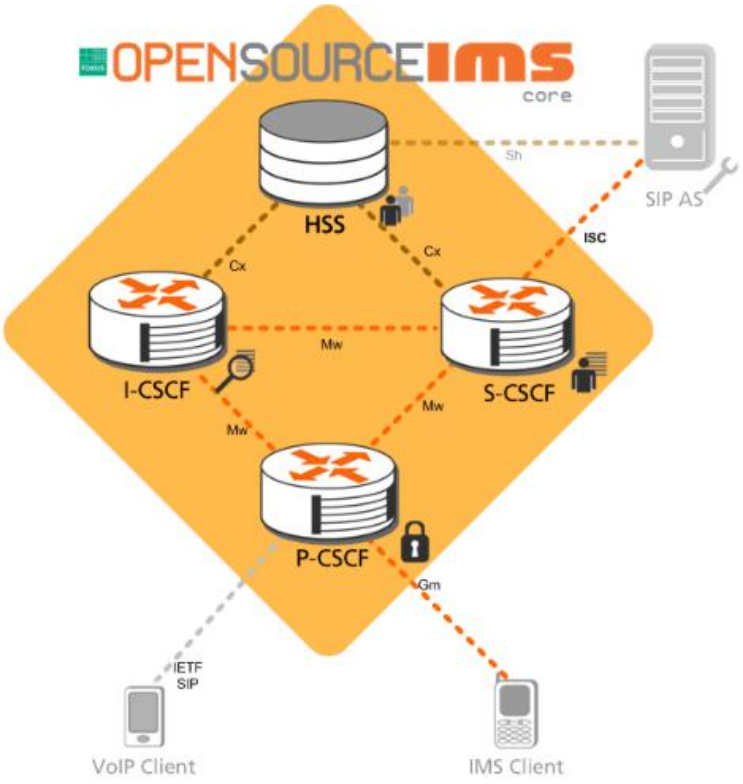

Figura 37. OpenImsCore

5.2.2.3. Capa de acceso

Los elementos utilizados en la capa de acceso son: un terminal computacional genérico (PC), desde el cual se generan las peticiones HTTP (compra de los productos). Los teléfonos SIP serán implementados mediante el softphone de libre distribución “X-lite” (recepción y finalización de llamadas).

Por lo tanto la arquitectura del prototipo queda conformada de la siguiente forma (Figura 38)

- La capa de aplicación estará conformada por un entorno JAIN SLEE, el cual hará las veces de servidor de aplicaciones del OpenIMSCore y adicionalmente proporcionará los adaptadores de recursos necesarios.
- La capa de control estará conformada por los elementos básicos de la arquitectura IMS/NGN.
- La capa de acceso será conformada por teléfonos SIP y terminales computacionales (PC).

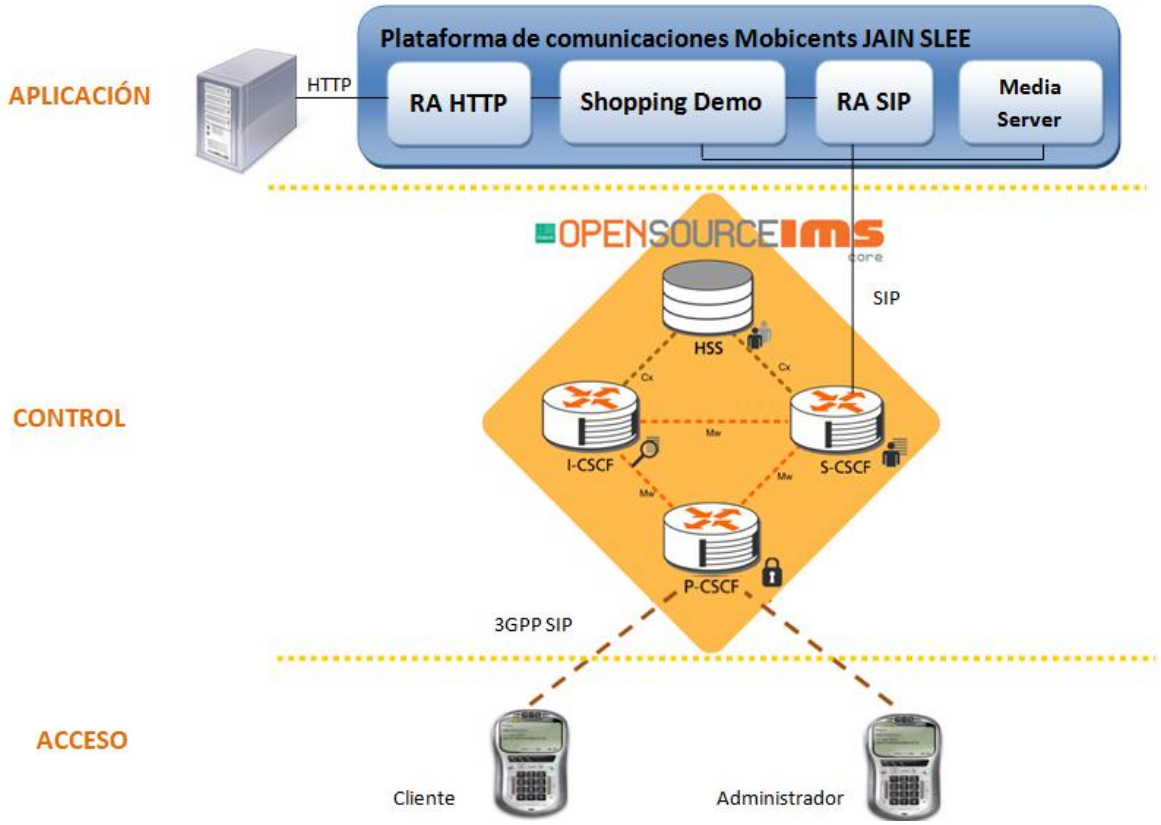

Figura 38. Arquitectura del Prototipo

5.2.3. Descripción del servicio

Con el fin de realizar las pruebas de laboratorio en un ambiente de convergencia, será desplegado en la arquitectura anteriormente expuesta el servicio convergente "Shopping Demo"; cabe aclarar que el único criterio que se tuvo en cuenta para la selección del servicio fue la convergencia, lo cual indica que podría ser utilizado cualquier otro servicio convergente.

"Shopping Demo" es un servicio de compra vía Internet, en el cual un usuario inscrito en el servicio de compra, selecciona uno o más productos a comprar, después de esta selección el usuario recibirá una llamada de confirmación. El flujo detallado del servicio se muestra a continuación:

5.2.3.1. Flujo del servicio

A continuación se realiza la descripción del flujo del servicio "Shopping Demo" (

Figura **39**):

- Una vez el usuario se encuentre inscrito puede realizar compras.
- Cuando se ordena la compra de algún producto, el cliente recibirá una llamada de confirmación, en la cual se presentan las siguientes opciones:
    - Si el usuario presiona la tecla "1", confirmará el pedido y el flujo del servicio continúa.
    - Si el usuario presiona la tecla "2", cancelará el pedido y el flujo del servicio terminará.
- Si el pedido es inferior a $ 100, la orden queda aprobada sin la necesidad de la confirmación del administrador y automáticamente el usuario recibirá una llamada para establecer la fecha y hora de entrega del producto.
- Si el pedido es superior a $ 100, el administrador tiene que aprobar o rechazar el pedido.
    - Si el administrador presiona la tecla "1", aprueba la compra y el flujo del servicio continúa.
    - Si el administrador presiona la tecla "2", la compra no es aprobada y finaliza el flujo del servicio.
- Una vez la compra ha sido aprobada por el administrador, se genera una llamada al usuario para confirmar la hora y la fecha de entrega.
- Cuando el usuario ha fijado la hora y fecha de entrega mediante el sistema DTMF (Dual Tone Multi Frequency), dichos datos pueden verificarse en la pestaña "My orders" dando "click" en la pestaña "Show Details".
- Después el administrador debe iniciar sesión en portal web y marcar la orden para la entrega.
- Tan pronto como el administrador ha marcado la orden para ser enviada, el usuario recibe una llamada recordándole la fecha y hora de envío.

El diagrama de flujo completo del servicio se puede apreciar en la Figura 39:

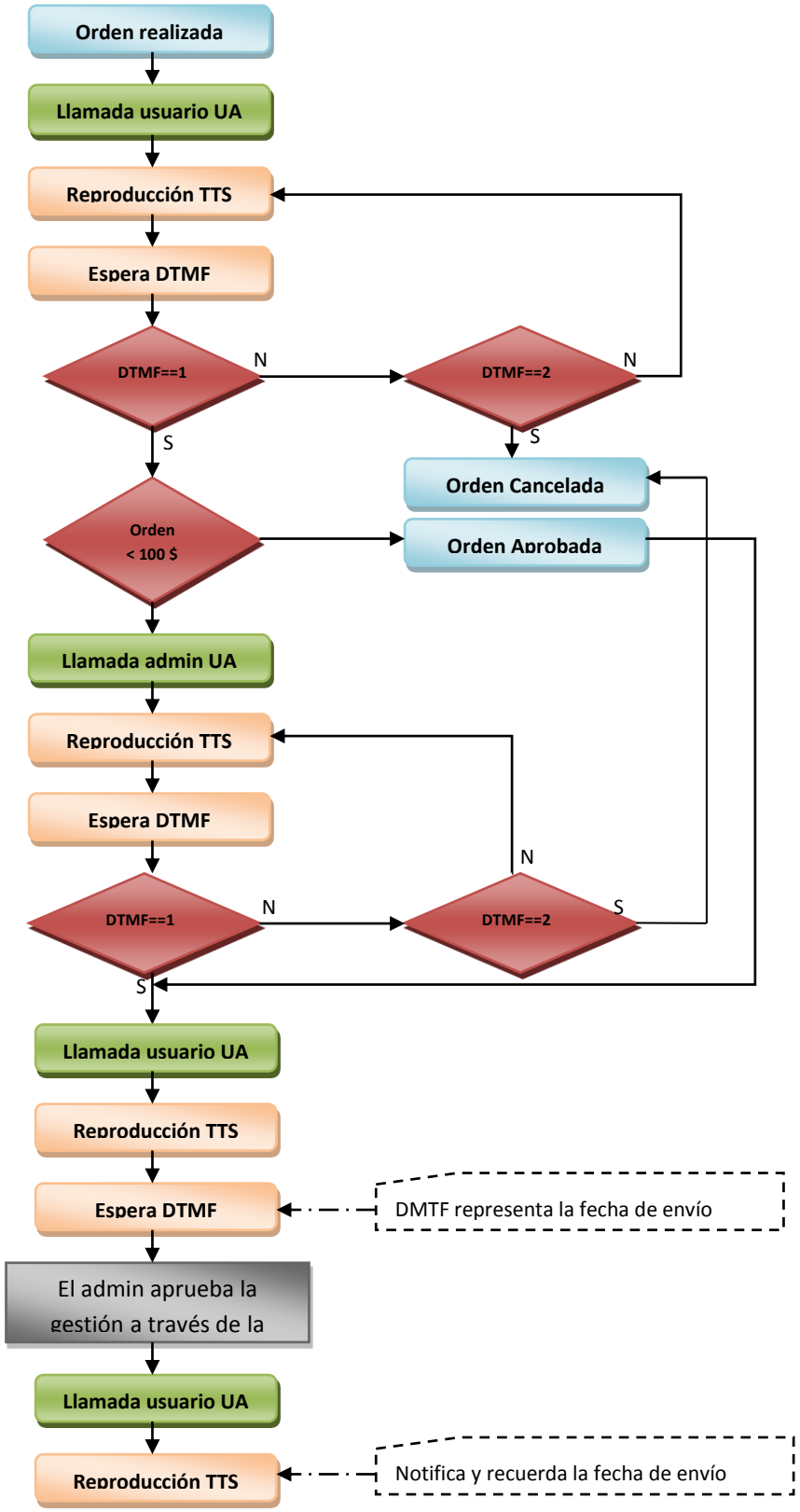

Figura 39. Diagrama de flujo del servicio "Shopping Demo"

5.3. Plan de pruebas y resultados obtenidos

El plan de pruebas al que será sometido el prototipo de laboratorio y con el cual se calcularán los porcentajes de disponibilidad del mismo, se basa en la realización de pruebas de estrés en dos escenarios: i) sin la implementación de lineamientos asociados al balanceo de cargas y clúster ii) con la posterior implementación de los mismos.

5.3.1. Especificaciones técnicas

Los parámetros técnicos para la realización de las pruebas se muestran a continuación (Tabla 9):

Servidor central	Procesador Intel(R) Core (TM)2 Duo P7450 2.13 GHz, Memoria RAM 4 GB
Sistema Operativo	Ubuntu v11.04
Servidor JAIN SLEE	Servidor JAINSLEE Mobicents v2.4.0
Servidor de aplicaciones	Servidor de medios v2
Sistema de control de telecomunicaciones	OpenIMSCore v2
Generador de trafico SIP	SIPp v3.2
Generador de trafico Web	JMeter v2.4

Tabla 9. Especificaciones técnicas

5.3.2. Proceso de pruebas

El proceso de pruebas consiste en la emulación de un usuario que realiza una compra vía Internet y que posteriormente recibe una llamada de confirmación; dicho proceso de pruebas será simulado a grandes escalas, como se muestra a continuación.

Mediante la utilización del generador de peticiones HTTP “JMeter” serán generadas múltiples peticiones de compra al servicio convergente “Shopping Demo”, el cual al finalizar un proceso de compra realizará una llamada (SIP) de confirmación al cliente. Las llamadas SIP serán realizadas por el generador de tráfico “SIPp” en el cual se encuentra registrado el usuario que realiza la compra. El objetivo del anterior proceso de pruebas es el determinar el número de peticiones y llamadas que alcanza el sistema antes de dejar de funcionar.

Los dos escenarios en los que se realizarán las anteriores pruebas son descritos a continuación.

5.3.3. Escenarios de pruebas

Como se mencionó anteriormente, el prototipo será sometido a dos diferentes escenarios de pruebas, adicionalmente cada escenario será sometido a cinco diferentes contextos de evaluación; los escenarios y los contextos de evaluación son explicados a continuación.

- **Escenario 1:** consiste en realización de pruebas de desempeño, sin la implementación de lineamientos relacionados con clúster y balanceo de cargas.

- **Escenario 2**: consiste en la realización de las pruebas de desempeño después de la implementación de un clúster en la configuración activo-activo y de los balanceadores de cargas SIP y HTTP.

Los contextos de evaluación son los diferentes niveles de estrés a los que será sometido el prototipo, y mediante los cuales se determinara la capacidad de servicio del prototipo, dichos contextos fueron construidos de la siguiente forma:

- **Contexto 1:** se generan 50 hilos (cada hilo está conformado por 20 peticiones http; un hilo exitoso genera una llamada SIP), un hilo es enviado cada segundo para su resolución, es decir que cada segundo se envían 20 peticiones al servicio.

- **Contexto 2:** se generan 100 hilos, enviando dos hilos cada segundo; por lo tanto se envían 40 peticiones por segundo al servicio.

- **Contexto 3:** se generan 500 hilos, enviando 50 hilos cada segundo; es decir que se envían 1000 peticiones por segundo al servicio.

- **Contexto 4:** se genera 1000 hilos, enviando 100 hilos por segundo; es decir 2000 peticiones por segundo al servicio.

- **Contexto 5:** se generan 1500 hilos, enviando 150 hilos por segundo; lo cual es 3000 peticiones por segundo al servicio.

Los anteriores contextos de pruebas fueron formados con la herramienta JMeter; en la Figura 40 se muestra la forma en que se genera uno de los contextos anteriormente nombrados.

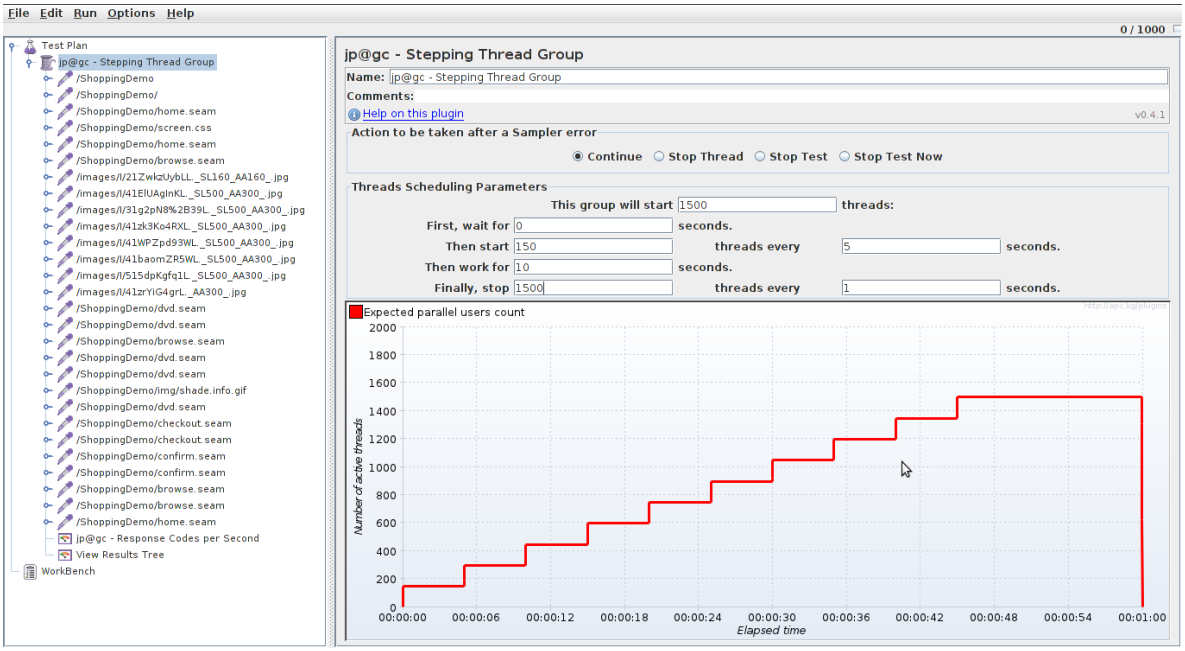

Figura 40. Generación de un contexto de pruebas mediante la herramienta JMeter

5.3.4. Resultados obtenidos

A continuación se presentan los resultados obtenidos en las diferentes pruebas de rendimiento, para dicho objetivo se generó para cada escenario: una gráfica en la que se muestra el número de peticiones generadas por segundo vs las peticiones erróneas, se cálcula el tiempo de indisponibilidad del sistema, se indica el número de llamadas SIP generadas y se realiza el cálculo de la disponibilidad.

5.3.4.1. Resultados obtenidos – Escenario 1.

A fin de explicar de forma general cómo se generó la gráfica de peticiones por segundo vs peticiones erróneas, a continuación se muestra y explica el resultado obtenido de una de las pruebas (contexto 5).

En la Figura 41 se puede apreciar en color azul las peticiones exitosas, en rojo las peticiones que para su resolución son redireccionadas a otros sitos (resolución de imágenes) y en color fucsia las peticiones erróneas; también se puede apreciar los periodos de inactividad del sistema o "downtime" (periodos llanos; no se resuelve ninguna petición).

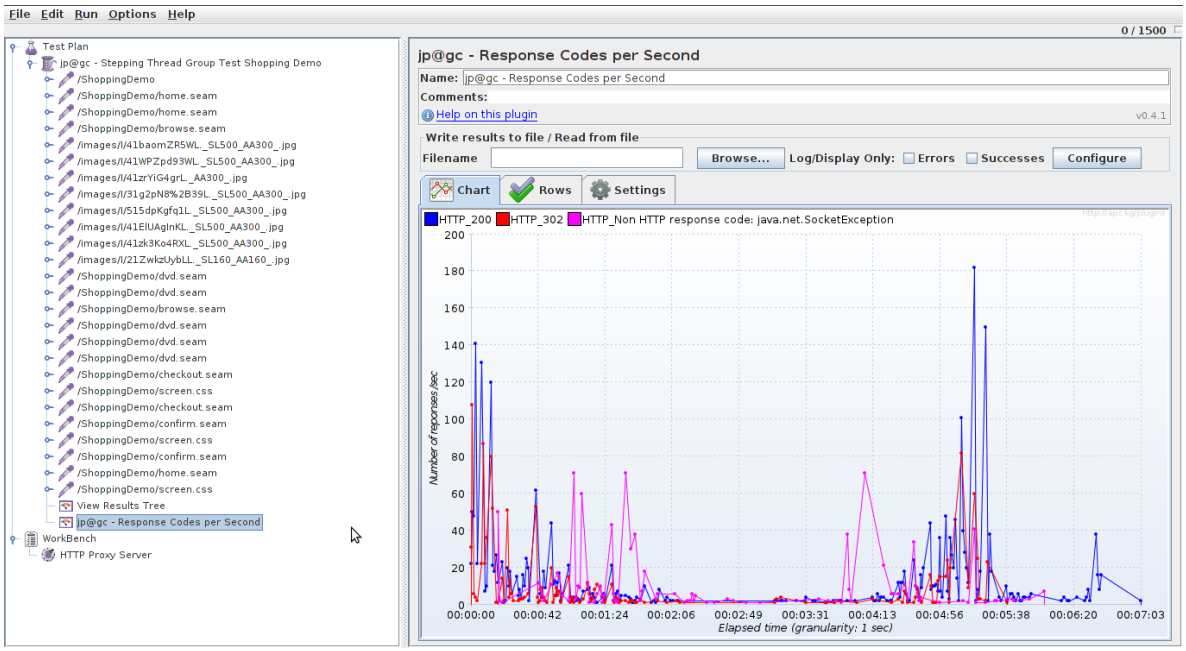

Figura 41. Resultados obtenidos para el contexto 5, escenario 1

De la anterior gráfica y de las pruebas realizadas en los otros contextos de evaluación se pueden abstraer los siguientes datos (Tabla 10):

Peticiones por segundo	Tiempo de funcionamiento	Tiempo de inactividad	Tiempo total de actividad	Peticiones perdidas	Llamadas SIP generadas por segundo
20	1 min 37 seg	0 seg	1 min 37 seg	0	1
40	2 min 6 seg	29 seg	2 min 35 seg	0	2
1000	3 min 58 seg	2 min 1 seg	5 min 59 seg	75	45
2000	4 min 52 seg	2 min 2 seg	6 min 54 seg	750	60
3000	5 min 1 seg	2 min 2 seg	7 min 3 seg	2000	50

Tabla 10. Resultados pruebas escenario 1

A continuación se presenta la gráfica de peticiones por segundo vs peticiones perdidas para el escenario de pruebas 1 y todos sus contextos (Figura 42).

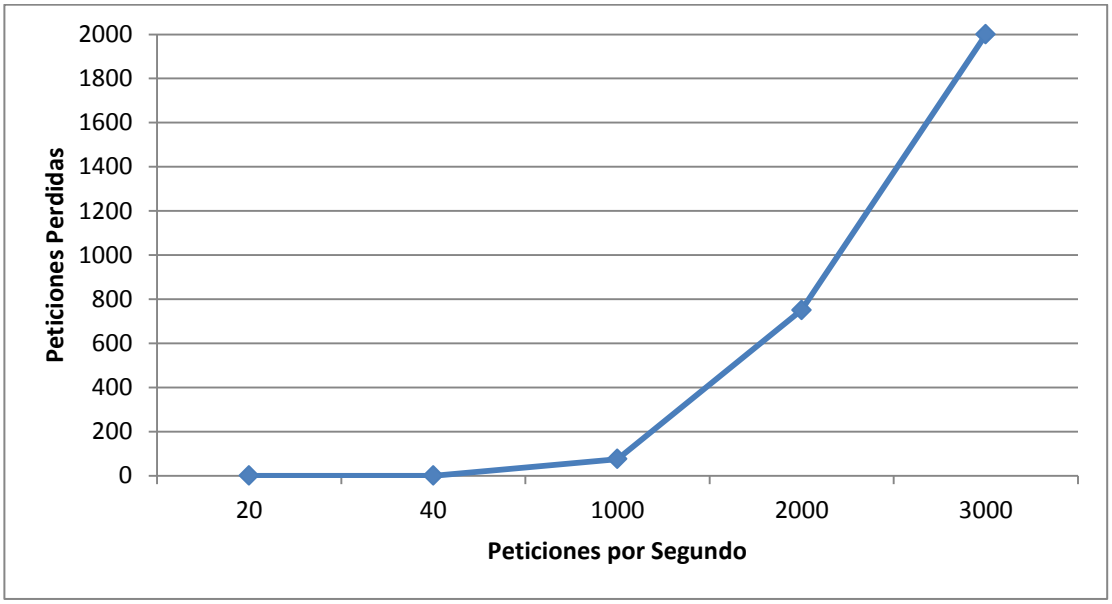

Figura 42. Peticiones por segundo vs peticiones perdidas escenario 1

Para el cálculo de la disponibilidad del sistema se utilizó la ecuación 2 (EC 2) mostrada en el capítulo 2, sección 2.1.4., y se consideraron los siguientes aspectos.

- De las pruebas realizadas al prototipo, se calcularon los periodos de inactividad del sistema; para el presente escenario y teniendo en cuenta los contextos en que se generan mayores peticiones por segundo, el promedio de tiempo interrupción o "downtime" es de 2 minutos y 2 segundos (2,03 minutos) (Tabla 10);

- Los servicios 24/7 se encuentran expuestos a errores y a interrupciones; en promedio se producen errores o interrupciones con una taza de 8,33 por mes, sin embargo para efectos de cálculo este valor se aproximará a 8 [22, 99].

Considerando la anterior información el tiempo de inactividad del sistema o "downtime" para el presente escenario es:

DT = 8 [veces/mes] x 2,03 [minutos] = 16,24 [minutos/mes]

Así, el cálculo de la disponibilidad del servicio para este escenario de pruebas es:

%Disponibilidad= ((AST-DT)/AST)*100

Dónde:

AST = 24 [horas/día] * 30 [días] = 43200 [minutos], tiempo acordado del servicio en un mes.

DT = 16,24 [minutos/mes], tiempo de caída o interrupción del servicio.

%DisponibilidadEscenario1= ((43200-16,24)/43200)*100 = 99.9624%

5.3.4.2. Resultados obtenidos – Escenario 2.

Como se mencionó anteriormente, en el escenario de pruebas dos serán implementados los criterios i) balanceo de cargas, en las modalidades: balanceo de cargas SIP y balanceo de cargas HTTP, y ii) clúster en la modalidad de alta disponibilidad y configuración activo-activo. Para la implementación de los criterios, se mantuvieron los mismos componentes hardware.

La implementación del criterio clúster, en la modalidad de alta disponibilidad y configuración activo – activo, se recreó mediante la replicación del servidor JAIN SLEE Mobicents en dos nodos virtuales.

El balanceador de cargas SIP fue implementado mediante la herramienta "Mobicents SIP Load Balancer", mientras que el balanceo de cargas HTTP fue implementado mediante el servidor "Apache" en el módulo "mod_jk". Los aspectos técnicos acerca de la implementación de los lineamientos son expuestos detalladamente en el Anexo D.

A continuación se presentan los resultados obtenidos en el presente escenario.

En la Figura 43 se puede apreciar el escenario de pruebas dos bajo el contexto 5.

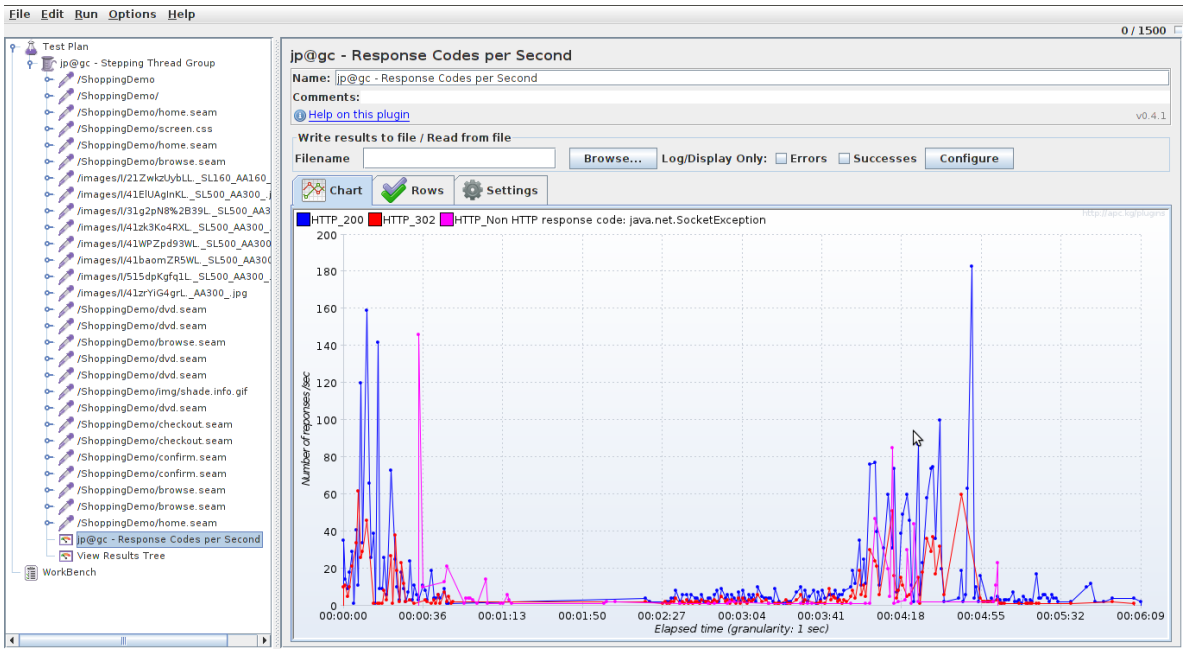

Figura 43. Realización pruebas bajo el contexto 5, escenario 2

De la anterior gráfica y de los resultados obtenidos en los otros contextos de evaluación se generó la siguiente la siguiente información (Tabla 11):

Peticiones por segundo	Tiempo de funcionamiento	Tiempo de inactividad	Tiempo total de actividad	Peticiones perdidas	Llamadas SIP generadas por segundo
20	1 min 40 seg	0 seg	1 min 40 seg	0	1
40	2 min 26 seg	16 seg	2 min 42 seg	0	2
1000	4 min 32 seg	1 min 32 seg	6 min 4 seg	0	50
2000	4 min 34 seg	1 min 31 seg	6 min 5 seg	300	85
3000	4 min 39 seg	1 min 30 seg	6 min 9 seg	1000	100

Tabla 11. Resultado pruebas escenario 2

A continuación se presenta la gráfica de peticiones por segundo vs peticiones perdidas para el escenario de pruebas 2 y todos sus contextos (Figura 44).

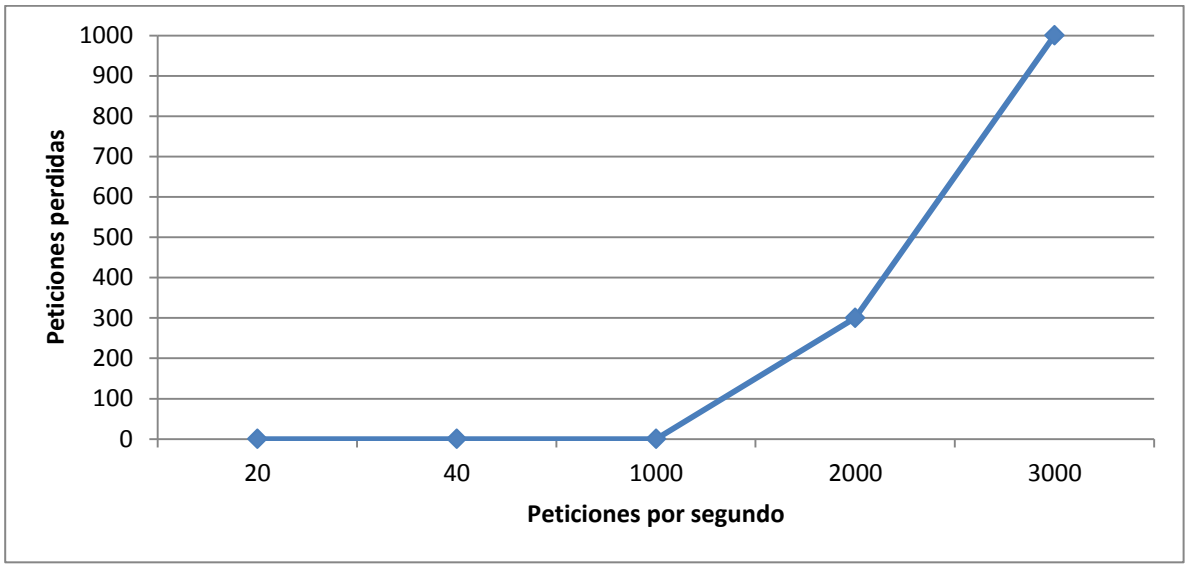

Figura 44. Peticiones por segundo vs peticiones perdidas escenario 2

De las pruebas realizadas al prototipo, se calcularon los periodos de inactividad del sistema; para el presente escenario y teniendo en cuenta los contextos en que se generan mayores peticiones por segundo, el promedio de tiempo interrupción o "downtime" es de 1 minuto y 32 segundos (1,52 minutos) (Tabla 11);

Como se mencionó anteriormente el promedio de paradas de los servicio 24/7 es de 8 paradas por mes [22, 99].

Por lo tanto el tiempo de interrupción o "downtime" mensual para el escenario dos es de:

DT = 8 [veces/mes] x 1,52 [minuto] = 12,16 [minutos/mes]

Así, el cálculo de la disponibilidad del servicio bajo el presente escenario de pruebas es:

%Disponibilidad= ((AST-DT)/AST)*100

Dónde:

AST = 24 [horas/día] * 30 [días] = 43200 [minutos], tiempo acordado del servicio en un mes.

DT = 12,16 [minutos/mes], tiempo de caída o interrupción del servicio.

%DisponibilidadEscenario2= ((43200-12,16)/43200)*100 = 99.9720%

5.3.4.3. Comparación de resultados

De los resultados obtenidos en las dos secciones anteriores, se puede concluir que mediante la aplicación de los lineamientos propuestos en el capítulo 4 sección 4.3, se logra un aumento considerable en el porcentaje de disponibilidad del prototipo, a pesar de que no se logra el objetivo de disponibilidad de 99.999%. Sin embargo, cabe recordar que los porcentajes de disponibilidad clásicos para los servicios Web son del 99% y en el presente trabajo se logró un porcentaje de 99.9720% para un servicio convergente, el cual combina el mundo Web y el mundo Telco y que los porcentajes de inactividad se generaron por las fallas producidas en la parte Web del servicio.

Adicionalmente, es importante tener en cuenta que otros factores que limitan el alcance de los porcentajes de alta disponibilidad en condiciones de laboratorio tales como el acceso a equipos hardware de tipo carrier class[20] disponibles normalmente en la infraestructura de un operador de telecomunicaciones y las limitantes a nivel de implementación para reunir la totalidad de los criterios y lineamientos necesarios como fue analizado en el capítulo 4. Es posible concluir que la alta disponibilidad es un concepto que se alcanza de forma integral, es decir mediante la utilización de hardware carrier class, el software requerido y la infraestructura de soporte adecuada para la implementación de los criterios y lineamientos requeridos.

Por otro lado, es importante resaltar el alto rendimiento de la plataforma de comunicaciones Mobicents, con la cual se logró un porcentaje de disponibilidad de 99.9624%. No obstante, este no es el porcentaje, que según la documentación, puede lograr la herramienta, puesto que se establece que con el entorno JAIN SLEE se logra el carrier grade exigido por los operadores de telecomunicaciones, es decir 99.999%.

De la anterior informacion se puede observar que el prototipo en el escenario de pruebas uno, es decir sin la implementacion de lineamientos asociados al balanceo de cargas y clúster, soporta una carga de aproximadamente 40 peticiones HTTP por segundo, sin presentar errores, mientras que el prototipo en el escenario de pruebas dos (con la implementacion de los lineamientos), soporta aproximadamente 200 peticiones sin presentar error. Aunque no se alcanza el porcentaje de alta disponibilidad deseable por las razones expuestas anteriormente, se logra un incremento importante en el número de peticiones que soporta el servicio.

[20] Carrier class: en telecomunicaciones se refiere a un sistema o a un hardware o software que es extremadamente fiable.

Capítulo 6

Aportes, Conclusiones y Trabajos futuros

6.1. Aportes

Se generó una guía para los operadores de telecomunicaciones, donde se incluyen los criterios y lineamientos más relevantes, que deben ser tenidos en cuenta a fin de alcanzar servicios altamente disponibles en el contexto de plataformas NGSDP.

Para la selección de criterios asociados a la alta disponibilidad en el contexto de las NGSDP, se contó con la participación del operador de telecomunicaciones EMCALI, logrando de esta forma capturar los requerimientos y necesidades en un contexto práctico.

Se generó un modelo de entrevista para capturar los requerimientos de alta disponibilidad en el contexto de un operador de telecomunicaciones.

Se construyó una base de conocimiento alrededor de los criterios asociados a la alta disponibilidad, definidos por empresas líderes en el mundo de las tecnologías de la información y las telecomunicaciones.

Fueron documentados los valores de desempeño y de disponibilidad de un servicio convergente bajo el contexto de la plataforma de comunicaciones Mobicents JAIN SLEE en la capa de aplicación, OpenIMSCore en la capa de control y softphones X-lite y PCs genéricos en la capa de acceso, las cuales son herramientas de libre distribución y ofrecen una contribución importante como punto de comparación para aquellos trabajos relacionados que consideren la utilización de herramientas propietarias.

El artículo "*Propuesta de Lineamientos Técnicos para Proporcionar Alta Disponibilidad de Servicio en el contexto de una NGSDP*" generado a partir del trabajo de grado fue presentado en el Congreso Ibero Americano de Telemática (CITA) permitiendo la socialización del trabajo en el marco de un evento internacional. Adicionalmente el artículo fue publicado en la revista para la divulgación de trabajos de investigación del Instituto de Informática y Ciencias de la Computación de la UFRGS (*Universidade Federal do Rio Grande do Sul*) "*Cadernos de Informática*" (Edición v.6, n.1 (2011)).

6.2. Conclusiones

En el marco de un servicio de un servicio convergente se debe tener en cuenta que: la disponibilidad de los servicios tradicionales de telecomunicaciones es de 99.999% mientras que el porcentaje de disponibilidad de los servicios web es tan solo de 99%. En este sentido, la alta disponibilidad busca acercar los porcentajes de disponibilidad de los servicios convergentes a los establecidos para los servicios con características carrier grade en los Telco.

A partir del análisis conceptual realizado de los criterios planteados por las diferentes empresas líderes en las tecnologías de la información, se puede observar que existe consenso en algunos de los criterios técnicos planteados. Sin embargo, algunos de estos se centran en aspectos inherentes de sus propios modelos de negocio, por lo cual no es posible tomar una única referencia; por esta razón, el presente trabajo de grado enfocó sus esfuerzos en realizar un análisis más profundo de estos planteamientos con el apoyo de un operador de telecomunicaciones nacional, logrando unificar dichos criterios en un contexto más genérico y acorde al entorno regional.

La implementación de un prototipo de laboratorio, demostró la eficacia de la aplicación de los criterios y lineamientos planteados a través de un diagnóstico inicial, al pasar de un porcentaje de disponibilidad del 99.962% a 99.972%, lo cual se tradujo en un aumento de 200 peticiones resueltas por segundo y de 10 llamadas por segundo realizadas. Este valor es muy superior al porcentaje de disponibilidad tradicional de los servicios en la Web (99%) y más cercano al porcentaje de alta disponibilidad de los servicios de telecomunicaciones (99.999%), lo cual es bastante aceptable dadas las condiciones usadas en la implementación del prototipo que no considera los componentes hardware “carrier class” disponibles en la infraestructura Telco.

En este contexto, es posible establecer que mediante la implementación de criterios y lineamientos técnicos usando técnicas software, se produce un incremento considerable en la disponibilidad de los sistemas de telecomunicaciones de nueva generación. Esta es una alternativa interesante para que los operadores de telecomunicaciones puedan reducir de alguna manera sus costos de inversión.

A través de la participación de otros operadores de telecomunicaciones y el análisis de sus contextos específicos, la entrevista modelo diseñada para establecer un diagnóstico sobre los requerimientos de alta disponibilidad generada en el presente trabajo de grado, se constituye en un punto de partida importante para la evolución del marco de trabajo planteado.

6.3. Trabajos Futuros

Realizar una comparación del rendimiento alcanzado al implementar soluciones SDP propietarias como la de OpenCloud y las soluciones de libre distribución, como la implementada en el presente trabajo de grado.

Realizar la implementación de un piloto sobre la infraestructura de un operador de telecomunicaciones, a fin de verificar los resultados obtenidos en el ambiente de laboratorio contando con la infraestructura necesaria para abordar la alta disponibilidad de manera integral.

Buscar estrategias que permitan la realimentación de los operadores de telecomunicaciones, con el ánimo de enriquecer los aportes al conjunto de criterios y lineamientos proporcionados en el presente trabajo de grado.

Referencias

[1] ITU-D. (2009, 5 - 9 Octubre). The world in 2009: ICT FACTS AND FIGURES. *ITU TELECOM WORLD 2009*. Available: http://www.itu.int/ITU-D/ict/material/Telecom09_flyer.pdf [Citado Marzo 2011]

[2] ENTER and IDATE. (2009, Febrero). Mobile 2009: Markets & Trends - Facts & Figures. Available: http://www.enter.ie.edu/enter/mybox/cms/9812.pdf [Citado Marzo 2010]

[3] E. 4Gamericas, "Global Mobile Market Shares," ed, 2010, p. [Documento en linea] http://www.3gamericas.org/index.cfm?fuseaction=page&pageid=565 [Citado Marzo 2011].

[4] ColombiaDigital. (2009, Marzo). *"Internet en el Mundo"*. Available: http://www.colombiadigital.net/index.php?option=com_content&view=article&id=174&Itemid=218 [Citado Abril 2011]

[5] Y. Zheng*, et al.*, "An Intelligent and Cognitive Service Delivery Platform Model," in *Intelligent Information Technology Application, 2008.IITA '08. Second International Symposium on*, 2008, pp. 137-140.

[6] CiscoSystems, "Designing High-Availability Services," ed, pp. 173-201.

[7] CarlosSerrano, *Modelo de Construcción de Soluciones*. Popayán: Universidad del Cauca, 2005.

[8] R. Christian and H. Hanrahan, "Structuring the Next Generation Network Using a Standards-Based Service Delivery Platform," *Innovations in NGN: Future Network and Services, 2008. K-INGN 2008. First ITU-T Kaleidoscope Academic Conference,* pp. 33-40, 12-13 May 2008 [Citado Febrero 2010].

[9] H. Lu*, et al.*, "The Next Generation SDP Architecture: Based on SOA and Integrated with IMS," in *Intelligent Information Technology Application, 2008. IITA '08. Second International Symposium on*, 2008 [Citado Mayo 2010], pp. 141-145.

[10] S. H. Maes, "Service delivery platforms as IT Realization of OMA service environment: service oriented architectures for telecommunications," in *WCNC 2007- IEEE Wireless Communications and Networking Conference*, 2007 [Citado Mayo 2010], pp. 2883-2888.

[11] Moriana, "Section A Moriana Analysis SDP in the Web 2.0 Era," in *SDP 2.O Service Delivery Platforms in the Web 2.0 Era. Free Operator Guide*, K. Kimbler, Ed., ed, 2008, pp. 15-36.

[12] Moriana, "Section B Thought Leadership White Papers," in *SDP 2.O Service Delivery Platforms in the Web 2.0 Era. Free Operator Guide*, K. Kimbler, Ed., ed, 2008, pp. 87-96.

[13] OASIS. *SOA*. Available: [Documento en línea]http://www.oasis-open.org/committees/tc_cat.php?cat=soa [Citado Enero 2011]

[14] H. Packard, "Service Delivery Platform 2.0," in *SDP 2.O Service Delivery Platforms in the Web 2.0 Era. Free Operator Guide*, D. Isaacson, Ed., ed, 2008, pp. 197-206.

[15] R. a. Markets. (2007). *Business Models and Drivers for Next-Generation IMS Services*. Available: http://www.researchandmarkets.com/reports/571538/business_models_and_drivers_for_next_generation [Citado Junio 2010]

[16] r. G. P. P. (3GPP). *IP-Multimedia Subsystem*. Available: http://www.3gpp.org/article/ims [Citado Junio 2010]

[17] C. d. I. e. I. e. T. (CINIT). (2007). *IMS*. Available: http://www.cinit.org.mx/articulo.php?idArticulo=50 [Citado Junio 2010]

[18] J.-L. D. Simón ZNATY, Roland GELDWERTH EFORT, "IP Multimedia Subsystem :Principios y Arquitectura," *http://www.efort.com/media_pdf/IMS_ESP.pdf [Citado Enero 2011]*.

[19] T. I. e. Telecomunicaciones). (Junio). *Arquitectura IMS: Definición general de los componentes*. Available: http://www.teleco.com.br/es/tutoriais/es_tutorialims/pagina_3.asp [Citado Junio 2010]

[20] Normalización(UIT-T). *Disponibilidad*. Available: http://www.itu.int/ITU-R/asp/terminology-definition.asp?lang=es&rlink={476DA5CB-FA18-4A35-83D1-421ADB9B7989} [Citado Enero 2011]

[21] S. A. Forum. *Providing Open Architecture High Availability Solutions*. Available: http://www.saforum.org/ [Citado Julio 2010]

[22] S. M. P. Cárdenas and J. J. I. Patiño, "PROPUESTA DE SOLUCIÓN DE ALTA DISPONIBILIDAD DE LOS SERVICIOS CRÍTICOS DEL CENTRO DE DATOS DE LA UNIVERSIDAD DEL CAUCA," Investigación, Telecomunicaciones, Universidad del Cauca, Popayán, 2010.

[23] M. M. D. Network). (2003). *Introducción a la disponibilidad*. Available: http://msdn.microsoft.com/es-es/library/aa291543 [Citado julio 2010]

[24] Kioskea. (2008). *Alta disponibilidad-Introducción a la fiabilidad*. Available: http://es.kioskea.net/contents/surete-fonctionnement/haute-disponibilite.php3 [Citado Julio 2010]

[25] I. T. I. Library(ITIL). (2009). *Gestión de la Disponibilidad Métodos y Técnicas*. Available: http://itil.osiatis.es/Curso_ITIL/ [Citado Julio 2010]

[26] Oracle, "Increasing Application Availability Using Oracle VM Server for SPARC: An Oracle Database Example," in *Increasing Application Availability Using Oracle VM Server for SPARC: An Oracle Database Example*, ed, April 2010, pp. 1-4.

[27] E. Ciurana. (2010). *Scalability & High Availability*. Available: http://library.dzone.com/sites/all/files/refcardz/rc043-010d-scalability_3.pdf [Citado Julio 2010]

[28] J. Hernández. (2006). *Alta disponibilidad, solución invaluable*. Available: http://esemanal.mx/2006/06/alta_disponibilidad_solucion_invaluable/ [Citado Julio 2010]

[29] IBM-i, "Visión general de la alta disponibilidad," in *Visión general de la alta disponibilidad*, ed, pp. 3-7. http://publib.boulder.ibm.com/infocenter/iseries/v7r1m0/index.jsp?topic=/rzarj/rzarjprint.htm [Citado Agosto 2010].

[30] OpenCloud, "RHINO 2.1: OVERVIEW AND CONCEPTS," in *RHINO 2.1: OVERVIEW AND CONCEPTS*, ed, 2009, pp. 29-30. Available: http://www.opencloud.com/ [Citado Julio 2010].

[31] J. Rojas and D. Ramirez, "Lineamientos para composición de servicios de telecomunicaciones en un entorno JAIN SLEE basado en software de libre distribución.," Pregrado, Telemática, Universidad del Cauca, Popayán, 2010.

[32] C. F. E. Solano and J. A. C. Muñoz, "Criterios Técnicos Para El Aprovisionamiento De Vas En Una Ngn Dentro Del Contexto Colombiano," Pregrado, Telemática, Universidad del Cauca, Popayán, 2010.

[33] T. Anderson, "Providing open architecture high availability solutions," 2004.

[34] S. c. L. C. d. Conocimiento. (2001). *Servicios de telecomunicaciones*. Available: http://www.sappiens.com/castellano/glosario.nsf/Telecomunicaciones/Servicios_de_telecomunicaciones/BE8B79F2C2415128002569FA00407AA4!opendocument [Citado Enero 2011]

[35] U. I. d. T. ITU. (2007). *SERIE F: SERVICIOS DE TELECOMUNICACIÓN NO TELEFÓNICOS*. Available: http://www.itu.int/rec/T-REC-F.500-198811-S/en [Citado Enero 2011]

[36] C. d. Colombia. *Ley 80*. Available: www.ing.unal.edu.co/site/htm/facultad/normatividad/nac/ley_80_1993.doc [Citado Enero 2011]

[37] R. d. Colombia. (1990). *Decreto 1900*. Available: http://www.presidencia.gov.co/prensa_new/decretoslinea/1990/agosto/19/dec1900191990.pdf [Citado Enero 2011]

[38] i-Uris. (2003). *DECRETO 600 DE MARZO 14 DE 2003*. Available: http://www.i-uris.com/leyes/dec/600_03.htm [Citado Enero 2011]

[39] C. N. d. T. CNTV. (2007). *Decreto 2870*. Available: http://www.cntv.org.co/cntv_bop/basedoc/decreto/2007/decreto_2870_2007.html [Citado Enero 2011]

[40] W. W. W. Consortium. (2004). *Web Services Glossary*. Available: http://www.w3.org/ [Citado Enero 2011]

[41] T. O reilly, "What is Web 2.0: Design patterns and business models for the next generation of software," *Communications and Strategies,* vol. 65, p. 17., 2007. http://oreilly.com/pub/a/web2/archive/what-is-web-20.html?page=5 [Citado Enero 2011].

[42] T. Goodmanson, "Web 2.0 and its Benefits," *Calabrio,* 2007.

[43] S. P. Crespo. (2007). *Cómo será la web 3.0*. Available: http://sociedadinformacion.fundacion.telefonica.com/DYC/SHI/Articulos_Tecnologias_-_Como_sera_la_web_30/seccion=1188&idioma=es_ES&id=2009100116310011&activo=4.do [Citado Enero 2011]

[44] K. Leins, "The Wheel Of Convergence How operators can realise the benefits of real time convergent charging and billing," *Ericsson Australia,* Agoust 2009.

[45] S. SECTOR and O. ITU, "SERIES Y: GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS AND NEXT-GENERATION NETWORKS."

[46] M. Camelo*, et al.*, "Convergencia de servicios en redes de próxima generación."

[47] M. d. T. d. l. I. y. l. Comunicaciones, "Plan Vive Digital Colombia," in *Plan Vive Digital Colombia*, ed Bogota, 2010, pp. 104-105.

[48] R. Orduz. (2010). *Convergencia o Empaquetamiento*. Available: http://www.ccdboletin.net/index.php?option=com_content&view=article&id=1254:iconvergencia-o-empaquetamiento&catid=308:blogs-el-heraldo [Citado Enero 2011]

[49] M. d. R. Guerra. (2007). *La Convergencia En Las Telecomunicaciones Y Sus Desafíos Para El Ministerio De Comunicaciones Decreto 2870 De 2007* Available: http://www.estatalescolombiaisp.org.co/apc-aa-files/fe67ef471db3e7c125d2e86aefbb87e5/DECRETO_SOBRE_CONVERGENCIA.pdf [Citado Enero 2011]

[50] J. C. M. Mendoza and E. M. Múnera, "Servicios convergentes en redes de próxima generación."

[51] D. Eror, "mobile service delivery platform " pp. 1-2. [Citado Enero 2011], 2010.

[52] D. Ferry and D. Page, "JAIN SLEE – a new standard for industry competitiveness."

[53] N. Kryvinska*, et al.*, "Next Generation Applications Mobility Management with SOA-A Scenario-Based Analysis," 2010, pp. 415-420.

[54] C. Yoon and H. Lee, "Service delivery platform for convergence service creation and management," 2010, pp. 1335-1338.
[55] Y. C. Zhou*, et al.*, "Service Storm: A Self-Service Telecommunication Service Delivery Platform with Platform-as-a-Service Technology," 2010, pp. 8-15.
[56] Y. Hu*, et al.*, "Global Service Delivery Platform: Deployment architecture and performance analysis," pp. 1-6.
[57] CiscoSystems. (2010, Enero). *About Cisco*. Available: http://www.cisco.com/web/about/index.html
[58] OpenCloud. (2000). *About OpenCloud*. Available: http://www.opencloud.com/about/ [Citado Enero 2011]
[59] I. B. M. (IBM). (2001). *our history of progress*. Available: http://www-03.ibm.com/ibm/history/interactive/ibm_ohe_pdf_13.pdf [Citado Enero 2011]
[60] Oracle. (2005). *About Oracle*. Available: http://www.oracle.com/us/corporate/index.html [Citado Enero 2011]
[61] Oracle, "Oracle DatabaseHigh Availability Overview 11g Release 2 (11.2)," in *Oracle DatabaseHigh Availability Overview 11g Release 2 (11.2)*

ed, 2010, pp. 13-26. Citado Febrero 2011.
[62] loadbalancing-org. (2010). *Load Balancing FAQ*. Available: http://www.loadbalancing.org/ [Citado noviembre 2010]
[63] J. G. Arco, A. Carral, JA Ibañez, G. (2007). *BSO algoritmo de reparto de tráfico para MPLS-TE*. Available: https://portal.uah.es/portal/page/portal/epd2_profesores/prof28259/publicaciones/jitel%2007.pdf [Citado noviembre 2010]
[64] O. L. B. S. Guide. *How Does Layer 7 Load Balancing Work?* Available: http://www.loadbalancingswitches.com/how-does-layer-7-load-balancing-work/ [Citado Noviembre 2010]
[65] C. U. O. d. C. Mateu. (2004). *Desarrollo de aplicaciones web*. Available: http://ocw.uoc.edu/computer-science-technology-and-multimedia/development-of-web-applications/development-of-web-applications/XP06_M2108_01497.pdf [Citado Diciembre 2010]
[66] G. Díaz*, et al.*, "Adaptación de Clusters de Linux para Servicios de Redes," *VI jornadas Científico Técnicas de la facultada de ingenierías, Universidad de los Andes,* 2007.
[67] V. Martín-Rubio Pascual and A. Fuentes Bermejo, "Estudio e Implantación de un Sistema de alta disponibilidad en RedIRIS," *RedIRIS,* p. 27, 2009.
[68] S. IT. *Introducción a Network Load Balancing (NLB)* Available: http://sitioit.com/introduccionNLB.aspx [Citado Diciembre 2010]
[69] P. A. Chitriv. *Overview of SQL Server 2000 Database Clustering using MSCS*. Available: http://www.codeproject.com/kb/database/SqlServerDBClusterMSCS.aspx [Citado Diciembre 2010]
[70] radware. *SIP Load Balancing*. Available: http://www.radware.com/Resources/sip_load_balancing.aspx [Citado Diciembre 2010]
[71] G. Kambourakis*, et al.*, "High Availability for SIP: Solutions and Real-Time Measurement Performance Evaluation."
[72] S. I. Inc. *Scalability and Performance Community* Available: http://support.sas.com/rnd/scalability/index.html [Citado Diciembre 2010]

[73] M. msdn. *Escalabilidad*. Available: http://msdn.microsoft.com/es-es/library/aa292172%28v=VS.71%29.aspx [Citado Diciembre 2010]
[74] D. W. Duma, *"Systems Engineering for Mission Success" RELIABILITY, AVAILABILITY, AND MAINTAINABILITY*, 2005.
[75] A. September, "IEEE Standard Glossary of Software Engineering Terminology," *Office,* vol. 121990.
[76] A. Birolini, *Reliability Engineering: Theory and Practice*, 2010.
[77] C. Friedman*, et al.*, "GENIES: a natural-language processing system for the extraction of molecular pathways from journal articles," *Bioinformatics,* vol. 17, p. S74, 2001.
[78] I. Sommerville, *Ingeniería del software*: Pearson Educación, 2005.
[79] Weibull. (2003). *Reliability Basics*. Available: http://www.weibull.com/hotwire/issue26/relbasics26.htm [Citado Enero 2011]
[80] J. Pan, "Software testing," *Retrieved January,* vol. 5, p. 2006.
[81] T. D. A. C. f. Software. *Software Reliability*. Available: https://www.thedacs.com/databases/url/key/2 [Citado Enero 2011]
[82] L. Prasad*, et al.*, "Measurement of Software Reliability Using Sequential Bayesian Technique," 2009.
[83] ReliaSoft. *Reliability Prediction Methods for Electronic Products*. Available: http://www.reliasoft.com/newsletter/v9i1/prediction_methods.htm [Citado Enero 2011]
[84] S. Storage. *Definition Backup*. Available: http://searchstorage.techtarget.com/definition/backup [Citado Enero 2011]
[85] Commvault. *Back Up Data*. Available: http://documentation.commvault.com/commvault/release_7_0_0/books_online_1/english_us/features/backup/backups.htm [Citado Enero 2011]
[86] S. Storage. *Full, incremental or differential: How to choose the correct backup type*. Available: http://searchdatabackup.techtarget.com/feature/Full-incremental-or-differential-How-to-choose-the-correct-backup-type[Citado Enero 2011]
[87] G. E. Cuestas Flores and S. J. Totoy Buitrón, "Análisis, diseño e implementación de las soluciones de alta disponibilidad de base de datos para pequeñas y medianas empresas (PYMES) utilizando tecnología MICROSOFT," 2010.
[88] J. S. Cubas, "Diseño en Alta Disponibilidad," Universidad Autonoma de Madrid: Escuela Politécnica Superior Informática.
[89] S. Finder. *Computacion en Alta Disponibilidad*. Available: http://www.slidefinder.net/c/computaci%C3%B3n_alta_disponibilidad/3605643 [Citado Enero 2011]
[90] A. G. B. Burbano, "Configuración de un cluster de alta disponibilidad y balanceo de carga en linux para satisfacer gran demanda web y servicios de resolución de nombres," *EPN,* 2007.
[91] J. C. Castillo and R. B. O. Acosta, "Implementacion de un cluster OPENMOSIX para computo cientifico en el instituto de ingenieria," Universidad Nacional Autonoma de Mexico, 2006.
[92] R. Buyya, "High Performance Cluster Computing: Architectures and Systems, Volume 1," *Prentice Hall PTR,* vol. 82, pp. 327-350, 1999.
[93] F. J. P. Rondón and E. Valencia, "CLUSTER MANGOSTA: Implementación y evaluación," *Faraute de Ciencias y tecnología,* 2006.

[94] O. M. G. Prieto, "Cluster de Alta Disponibilidad y Alto Desempeño para Servidores Web (ADAD-SW)," *UNIVERSIDAD NACIONAL DEL ESTE FACULTAD POLITÉCNICA,* p. 35.
[95] S. S. Quality. *Best Practice*. Available: http://searchsoftwarequality.techtarget.com/definition/best-practice [Citado Enero 2011]
[96] C. SOLUTIONS, "Utilizing proven practices is key to efficiency, usability, extensibility, and asset protection in IT.," *ORIGINALLY AN Internal Training Document,* 2008.
[97] A. P. Q. Center. *Make Best Practices Your Practices*. Available: http://www.apqc.org/ [Citado Febrero 2011]
[98] T. H. Group. (2010). *Business Best Practices*. Available: http://www.hackettbenchmarking.de/ [Citado Febrero 2011]
[99] N. I. S. C. C. N. I. S. C.-O. Centre, "Good Practice Guide To Telecommunications Resilience," 2006.
[100] A. I. d. P. Público, "Seminario Internacional de Presupuesto Publico," 2011.
[101] I. Sanchez., "Presupuesto y Control Presupuestario," Universidad de Carabobo.
[102] S. P. Robbins, "Herramientas y Técnicas de planeacion," 2009.
[103] Solusan. *Sistemas Informaticos Redundantes*. Available: http://www.solusan.com/sistemas-informaticos-redundantes.html [Citado Febrero 2011]
[104] IBM, "Redundancia Incorporada."
[105] M. A. O. López, "Fiabilidad y Tolerancia a Fallos en Sistemas en Tiempo Real," 2006.
[106] Accenture. *High Performance IT A Road Map for Technology Investments*. Available: http://www.accenture.com/us-en/technology/high-performance-information-technology/Pages/index.aspx [Citado Febrero 2011]
[107] M. Msdn, "Improving Applications Perfromance and Scalability," 2004.
[108] Mobicents. *Mobicents*. Available: http://www.mobicents.org/products.html [Citado Marzo 2011]
[109] J. Deruelle, "Mobicents Communications Platform."
[110] O. S. IMS. *Open Ims Core*. Available: http://www.openimscore.org/ [Citado Marzo 2011]