

UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

Facultad 1, Centro CESOL



SOLUCIÓN PARA LA GESTIÓN DEL ALMACENAMIENTO DE DATOS EN LAS INSTITUCIONES CUBANAS

Tesis presentada en opción al título académico de
Máster en Informática Avanzada

Autor: Yasiel Pérez Villazón

Tutores: Dr. Ramón Santana Fernández

MsC. Yoandy Pérez Villazón

Ciudad de La Habana, noviembre de 2018

Declaración de autoría

Declaro por este medio que yo Yasiel Pérez Villazón, con carné de identidad 89031525305, soy el autor principal del trabajo final de maestría "Solución para la gestión del almacenamiento de datos en las instituciones cubanas", desarrollada como parte de la Maestría en Informática Avanzada y que autorizo a la Universidad de las Ciencias Informáticas a hacer uso de la misma en su beneficio, así como los derechos patrimoniales con carácter exclusivo.

Y para que así conste, firmo la presente declaración jurada de autoría en La Habana a los ____ días del mes de _____ del año _____.

Ing. Yasiel Pérez Villazón

Dedicatoria

A mis padres, mis hermanos por ser mis guías cada día y alentarme para lograr mis objetivos. A mi esposa e hijo por apoyarme incondicionalmente en todos estos años. En general a toda mi familia y amigos...

Agradecimientos

- A toda mi familia por apoyarme siempre y sobre todo en mi superación profesional.
- A mis tutores por apoyarme durante el desarrollo de la maestría.
- A todos mis compañeros de trabajo del centro CESOL, que de una forma u otra colaboraron a este gran día, y en especial a Dayli, Nurisel, Gladys, Yadiel, Yasiel, Yosel y Gustavo por los momentos de alegría y de apoyo.
- A mis profesores y compañeros de los cursos de la Maestría y al comité científico de la Facultad1.
- A mi oponente y al presente tribunal por formar parte de este día
- A la UCI y a la revolución cubana por haberme regalado esta posibilidad de superación de forma gratuita

Resumen

Como parte del proceso de informatización de la sociedad cubana se realizó el diagnóstico de migración a código abierto a diez ministerios de la administración central del estado. En el análisis de los resultados referente al despliegue de los servicios telemáticos se constató que existen un conjunto de servicios telemáticos críticos para las instituciones que tenían el almacenamiento de sus datos en el propio servidor. El almacenamiento en los servidores implica que además del manejo de la concurrencia de los usuarios al servidor tenga que dedicar recursos a la lectura y escritura de datos, también disminuye la capacidad de intercambio de información con otros servicios lo que trae como consecuencia la duplicidad y la descentralización de los datos. Para dar solución al problema se crea una solución de almacenamiento basado en la arquitectura NAS teniendo en cuenta los principales tecnologías, protocolos y estándares existentes para lograr la interoperabilidad entre el sistema de almacenamiento y los servidores de las instituciones. Esta infraestructura está estructurada por diferentes componentes que tienen como objetivo la gestión del almacenamiento, la disponibilidad, la integridad, el respaldo, la seguridad y el monitoreo de los datos. Para validar la solución se emplean métodos para medir el criterio de experto y la satisfacción de los administradores de redes con la solución. También se seleccionan a partir de la Norma Cubana de Ingeniería de Software y Sistemas los indicadores para medir el intercambio de formatos de datos y la suficiencia del protocolo de intercambio de datos para evaluar a través del despliegue de la solución en la empresa ECOAIND la interoperabilidad que existe entre el sistema de almacenamiento y los servidores.

Palabras claves: Almacenamiento, centralización, interoperabilidad.

Abstract

As part of the information process of Cuban society, the diagnosis of open source migration was made to ten ministries of the central administration of the state. The analysis of the results refers to the deployment of telematic services found that there is a set of critical telematic services for institutions that had the storage of their data on the server itself. The storage in the servers also implies the handling of the management, the concurrence, the server, the service, the reading and writing of the data, the exchange of information, and the service. Decentralization of data. To obtain a solution to the problem, a storage solution based on the NAS architecture is created taking into account the main technologies, protocols and deadlines to achieve interoperability between the storage system and the institutions' servers. This infrastructure is structured by different components that have the objective of storage management, availability, integrity, backup, security and data monitoring. To validate the solution, methods are used to measure expert criteria and the satisfaction of network administrators with the solution. It can also be selected from the Cuban Standard for Software and Systems Engineering to measure the exchange of data formats and the experience of the Data Exchange Protocol to evaluate through the deployment of the solution in the company. It exists between the storage system and the servers.

Keywords: Storage, centralization, interoperability.

Índice de contenido

1 - Introducción.....	1
2 - Capítulo 1 - Almacenamiento de sistemas informáticos.....	8
2.1 - Conceptos asociados al almacenamiento de datos en sistemas informáticos.....	8
2.1.1 - Almacenamiento.....	8
2.2 - Arquitecturas de almacenamiento.....	9
2.2.1 - Definición de la arquitectura.....	11
2.3 - Análisis de herramientas que implementan la arquitectura NAS.....	13
2.3.1 - FreeNAS.....	13
2.3.2 - NAS4Free.....	14
2.3.3 - Openmediavault.....	14
2.3.4 - Elementos que complementan la gestión del almacenamiento.....	15
2.4 - Especificaciones para la gestión del almacenamiento propuestas por SNIA.....	16
2.4.1 - Perfiles para la Especificación de la Iniciativa de gestión de almacenamiento en NAS.....	16
2.5 - Análisis de interoperabilidad.....	18
2.5.1 - Aplicaciones integradas a la Distribución Cubana GNU/Linux Nova.....	18
2.5.2 - Productos desarrollados bajo la alianza UCI-GEDEME.....	20
2.6 - Conclusiones parciales.....	21
3 - Capítulo 2: Solución para el almacenamiento en las instituciones cubanas.....	23
3.1 - Componentes para la gestión del almacenamiento.....	23
3.1.1 - Distribución de los componentes.....	24
3.1.2 - Componente para la gestión del almacenamiento (CGA).....	25
3.1.3 - Componente para la gestión de la disponibilidad (CGD).....	29
3.1.4 - Componente para la gestión de seguridad (CGS).....	32
3.1.5 - Componente para la gestión de respaldos (CGR).....	34
3.1.6 - Componente para la gestión de integridad (CGI).....	35
3.1.7 - Componente para la gestión de monitoreo (CGM).....	37
3.2 - Conclusiones parciales.....	38
4 - Validación de la solución de almacenamiento de datos informáticos.....	39

4.1 - Valoración de los expertos sobre la solución de almacenamiento.....	39
4.1.1 - Resultados de la valoración de los expertos.....	41
4.2 - Satisfacción de los administradores de redes con la solución de almacenamiento.....	45
4.3 - Aplicación de la solución de almacenamiento en la empresa ECOIND.....	47
4.4 - Triangulación metodológica de los métodos aplicados.....	50
Conclusiones parciales.....	51
5 - Conclusiones.....	52
Recomendaciones.....	53
6 - Referencias bibliográficas.....	54
7 - Anexos.....	57
8 - Glosario de términos.....	69

Índice de figuras

Figura 1: Interacción entre componentes en el despliegue de las arquitecturas de almacenamiento. Fuente: elaboración propia.....	11
Figura 2: Distribución de componentes en el sistema de almacenamiento. Fuente: elaboración propia.....	24
Figura 3: Despliegue del componente de almacenamiento. Fuente: elaboración propia.....	29
Figura 4: Despliegue del componente de disponibilidad. Fuente: elaboración propia.....	31
Figura 5: Despliegue del componente de seguridad.. Fuente: elaboración propia.....	33
Figura 6: Despliegue del componente de gestión de respaldos.. Fuente: elaboración propia.....	35
Figura 7: Despliegue del componente de integridad. Fuente: elaboración propia.....	37

Introducción

Como parte del proceso de informatización de la sociedad cubana en el 2004 se publica por consenso del Consejo de Ministros de la República de Cuba el acuerdo 084, el cual orienta la migración paulatina de los Organismos de la Administración Central del Estado hacia aplicaciones de código abierto (Villazón, Vitier 2015a). Para organizar la migración se creó una estructura que se compone de cuatro grupos de trabajo con tareas y acciones específicas, bajo la dirección del Grupo Ejecutivo (Villazón, Villazón 2013). Este grupo ejecutivo se compone de cuatro subgrupos:

- **Legal:** Tiene la responsabilidad de redactar los mecanismos legales en la aplicación de los procesos de migración
- **Capacitación:** Tiene la responsabilidad de organizar y brindar los cursos de capacitación basados en las nuevas tecnologías que se van a desplegar en las instituciones a los usuarios que se van a beneficiar
- **Divulgación:** Promueve el uso de tecnologías libres en las instituciones. Tiene la responsabilidad de brindar charlas de sensibilización con el objetivo de minimizar el rechazo al cambio.
- **Técnico:** Tiene la responsabilidad de redactar mecanismos para aumentar la eficiencia en la aplicación de los procesos de migración.

La Universidad de las Ciencias Informáticas (UCI) forma parte del grupo técnico nacional y ha sido uno de los principales protagonistas en este proceso de migración. La UCI ha contribuido al desarrollo de aplicaciones y estrategias para hacer más eficiente el despliegue de las tecnologías libres en las empresas cubanas.

Entre los resultados más evidentes es el desarrollo de la distribución cubana de GNU/Linux Nova, un sistema operativo que se despliega en cuatro variantes: Nova Escritorio, Nova Ligero, NovaDroid y Nova Servidor, este último destinado a la administración de servicios telemáticos (UCI 2018). Esta distribución esta basada en cuatro principios:

- **Soberanía tecnológica:** Se basa en el principio de un desarrollo de forma autónoma, así como la capacidad decisonal sobre su uso y desarrollo (Fuentes 2011).
- **Socio-adaptabilidad:** Un sistema operativo hecho por cubanos para cubanos, alineado a las políticas que orienta la informatización nacional y optimizado para las condiciones tecnológicas del país (Fuentes 2011).

- **Sostenibilidad:** Se basa en el principio de mantener un proceso flexible y versátil en la construcción de la distribución. Promueve un desarrollo en constante innovación y consonancia con las nuevas tendencias tecnológicas para garantizar modelos de comercialización que permitan el ingreso de divisas por el concepto de exportación de productos y servicios (Fuentes 2011).
- **Seguridad:** Este principio propone un modelo de desarrollo colaborativo que permita el acceso al código fuente y el exhaustivo proceso de revisión y auditoría de código para garantizar un sistema seguro (Fuentes 2011).

El libro “Buenas prácticas para migración a código abierto” publicado en el 2015, es una actualización de la guía cubana de migración a código abierto. Este libro describe procedimientos y mecanismos a tener en cuenta para efectuar la migración a código abierto en una institución. Para implementar un proceso de migración se tienen en cuenta las personas, las tecnologías y la infraestructura de red. En el libro está reflejado cuales son las aplicaciones y tecnologías que están certificadas y avaladas para ser desplegadas como alternativas libres en las instituciones durante el proceso de migración. Este registro de aplicaciones contempla, entre otras, aquellas que están autorizadas a instalarse en los servidores, para la administración de los servicios telemáticos de las empresas (Villazón, Vitier 2015b).

La primera etapa del proceso de migración es realizar la Consultoría en Migración a Código Abierto, esta actividad consiste en planificar todas las actividades que serán realizadas durante el diagnóstico a la entidad. Entre las actividades más importantes en el diagnóstico de la entidad están:

- **Levantamiento de activos informáticos de la entidad:** Consiste en registrar cuales son las principales herramientas que utilizan los usuarios de la empresa. También se registran las propiedades de las estaciones de trabajo y de los dispositivos de hardware.
- **Levantamiento de los servicios telemáticos:** Consiste en identificar cual es la distribución física de los servidores, cuales son los servicios telemáticos que tiene la empresa y que tecnologías lo implementan.
- **Entrevistas a usuarios y administradores de redes:** Consiste en realizar entrevistas a los usuarios que interactúan con las estaciones de trabajo y servicios telemáticos.

El autor de la presente investigación analizó la documentación generada en el diagnóstico realizado a diez ministerios de la administración central del estado. Durante este proceso se llevó a cabo un levantamiento informático de las tecnologías que estaban desplegadas en las estaciones de trabajos y en los servidores.

Del análisis referente al despliegue de los servicios telemáticos en las instituciones se detectó que los

servicios más críticos para las empresas debido a la concurrencia de los usuarios en el acceso son:

- **Servicio de correo electrónico:** El correo electrónico (también conocido como e-mail, un término inglés derivado de electronic mail) es un servicio que permite el intercambio de mensajes a través de sistemas de comunicación electrónicos. El concepto se utiliza principalmente para denominar al sistema que brinda este servicio vía Internet mediante el protocolo SMTP, pero también permite nombrar a otros sistemas similares que utilicen distintas tecnologías. Los mensajes de correo electrónico posibilitan el envío, además de texto, de cualquier tipo de documento digital (imágenes, vídeos, audio, etc.)(Villazón, Villazón 2013).

Las empresas tienen un servidor de correo electrónico que permite brindar el servicio de comunicación entre los usuarios dentro y fuera de la empresa. La información que almacena este servicio son los buzones de correo electrónico. La capacidad de esta información depende de la capacidad de buzón de cada usuario y de la cantidad de usuarios que en la empresa tiene habilitado este servicio.

- **Servicios de compartición de archivos y directorios:** La compartición de archivos es la práctica de distribuir o proporcionar acceso a medios digitales, como programas de computadora, multimedia (audio, imágenes y video), documentos o libros electrónicos. El intercambio de archivos se puede lograr de varias maneras. Los métodos comunes de almacenamiento, transmisión y dispersión incluyen el intercambio manual utilizando medios extraíbles y servidores centralizados en redes informáticas (Chan, Baril 2017). Los servidores de transferencia y persistencia de archivos le brindan a las empresas la posibilidad a los usuarios de tener sus datos alojados en un centro de datos y no en sus estaciones de trabajo. La información que almacena este servicio son los archivos asociados a cada usuario.
- **Servicio de mensajería instantánea:** Servidores que implementan mecanismos de comunicación rápida a los usuarios de la empresa. La información que almacena este servicio son los registros de las conversaciones de los usuarios (Reina 2016).
- **Servicio de aplicaciones web:** Los servidores web desempeñan un papel dominante en la infraestructura de estos servicios. Su tarea principal es recibir y procesar solicitudes de clientes que exigen objetos específicos de los servidores y devolverlos mediante respuestas relacionadas. Los mensajes de solicitud y respuesta asociados se transportan mediante el Protocolo de Transferencia de Hipertexto (HTTP) o el Protocolo de Transferencia de Hipertexto (HTTPS) mediante conexiones TCP entre los clientes y el servidor (Tien Van Do, Udo R. Krieger 2008). La

información que almacena este servicio son los metadatos de las aplicaciones y de los usuarios que la utilizan.

- **Servicios de internet:** Servidores que proveen la capacidad de conexión a internet a los usuarios. Este servicios implementa mecanismos de caché¹ para aumentar la rapidez de conexión de los usuarios. La información que almacena este servicio son los datos temporales de la caché (Nishio 2010).
- **Controladores de dominio:** Provee mecanismos de autenticación centralizada de los usuarios y de las estaciones de trabajo de la empresa. La información que almacena este servicio son los objetos registrados (Rueda Nieto, María Lucia, Rodríguez Reyes, Rossana 2017).

En la distribución física de los servicios, cada servidor almacena la información referente a su servicio en el propio almacenamiento que dispone, lo que implica que además del manejo de la concurrencia de los usuarios al servidor tenga que dedicar recursos a la lectura y escritura de datos. A partir de las propiedades de capacidad de almacenamiento de los servidores que están desplegados en las empresas, estos no poseen un almacenamiento dedicado con grandes capacidades lo que provoca que los discos superen sus capacidades con frecuencia trayendo como consecuencia:

- Pérdida de la información
- Interrupción del servicio telemático
- Falta de disponibilidad de la información
- Errores en la integridad de los datos

El almacenamiento local en los servidores trae como consecuencia la duplicidad de la información ya que en los servidores se pueden almacenar los mismos datos debido a que no existen mecanismos de comunicación entre ellos para garantizar en un determinado servicio existan información de otros servicios.

Teniendo en cuenta los anteriormente descrito de plantea como problema científico:

¿Como facilitar la interoperabilidad en la comunicación con el sistemas de almacenamiento de los datos informáticos en las instituciones cubanas?

Para dar respuesta al problema científico se plantea como objetivo general:

Desarrollar una solución para la gestión del almacenamiento de datos informáticos, basado en tecnologías

1 Memoria de acceso rápido que guarda temporalmente los datos recientes de los procesados.

libres, para facilitar la interoperabilidad en la comunicación con el sistemas de almacenamiento de los datos informáticos en las instituciones cubanas.

A partir del objetivo general se definen los siguientes objetivos específicos:

- Elaborar el marco teórico referencial de la investigación asociado a las tecnologías asociadas a la gestión del almacenamiento de datos informáticos
- Fundamentar los componentes que integran la solución de almacenamiento de datos informáticos.
- Desarrollar la solución de almacenamiento de datos informáticos que centralice la gestión del almacenamiento de datos informáticos
- Validar la solución para la gestión del almacenamiento de datos informáticos .

Para guiar la investigación se propone como objeto de estudio: “El proceso de gestión del almacenamiento de datos” enmarcado en el campo de acción: “La interoperabilidad en la gestión del almacenamiento de datos en tecnologías libres”.

La investigación se sustenta en la siguiente hipótesis:

El desarrollo de una solución para la gestión del almacenamiento de datos informáticos, basado en tecnologías libres, contribuirá a la interoperabilidad en la comunicación con el sistemas de almacenamiento de los datos informáticos en las instituciones cubanas.

En la presente investigación se define interoperabilidad como:

La capacidad de comunicación entre el sistema de almacenamiento y las tecnologías que requieran integración con este, teniendo en cuenta las principales tecnologías, protocolo y estándares existentes para el intercambio de datos.

Para la operacionalización de las variables se toma como variable independiente “Sistema para la gestión del almacenamiento de datos” y como variable dependiente “Interoperabilidad”. La variable independiente se va a medir teniendo en cuenta los indicadores descritos en el Anexo 1. Los indicadores de medición fueron obtenidos a partir de la Norma Cubana de Ingeniería de Software y Sistemas que describe requisitos de la calidad y evaluación de software y sistemas, así como indicadores de medición de la calidad del producto de software o de sistema (ISO/IEC 25023:2017 2016)

La investigación realizada se sustenta en el empleo de los métodos científicos siguientes:

Analítico-Sintético: La aplicación de este método permitió la extracción de los elementos más importantes relacionados con las tecnologías de almacenamiento de datos y el funcionamiento de los

sistemas de gestión de almacenamientos de datos. Se analizaron documentos, bibliografías y estándares relacionadas con el almacenamiento de datos durante el proceso de escritura y lectura de información mediante el servicio de almacenamiento y a partir de esto se concretaron resultados que contribuyen a obtener cuales son los elementos más importantes en la gestión del almacenamiento.

Análisis Histórico-Lógico: Permitió constatar teóricamente cómo ha evolucionado los sistemas de gestión de almacenamientos de datos desde la creación de este término y las tecnologías asociadas al almacenamiento, lo que permitió inferir elementos para proponer una mejor solución.

Hipotético-Deductivo: Permitió verificar la hipótesis planteada y establecer nuevas predicciones a partir del sistema de conocimientos que se tiene.

Observación: Desde un punto de vista contemplativo, la aplicación de esta técnica permitió observar el funcionamiento de todas las tecnologías que complementan la solución para el almacenamiento de datos informáticos.

Encuesta: Utilizada en la validación de la solución tecnológica para el almacenamiento de datos mediante el método Delphi, aplicada en la selección de los expertos y recopilación de sus criterios; y en la técnica de ladov, aplicada a los administradores de servicios telemáticos que utilizan la solución del almacenamiento para determinar su grado de satisfacción con el mecanismo de seguimiento.

Como **aporte práctico** se obtiene una infraestructura tecnológica para la gestión del almacenamiento de datos en las instituciones cubanas.

La presente investigación se estructura en tres capítulos, conclusiones, recomendaciones, bibliografía y anexos.

A continuación se describen los capítulos

Capítulo 1: Descripción de los conceptos asociados al almacenamiento de sistemas informáticos. Se realiza un estudio acerca de las arquitecturas de almacenamientos que existen para determinar cual se puede aplicar a la solución de almacenamiento en las instituciones cubanas. Se describen las herramientas que administran la arquitectura seleccionada a desplegar con el objetivo de identificar tecnologías y estándares que se utilizan en el despliegue de sistemas de almacenamientos. Un punto importante de este capítulo es el análisis de las especificaciones que propone la Asociación de la Industria de Redes de Almacenamiento y la Alianza para las Tecnologías de Almacenamiento con el objetivo de determinar cuales son las mejores prácticas para el despliegue de este tipo de tecnologías.

Capítulo 3: Validación de la solución propuesta. Se realiza la validación de la solución de almacenamiento

de las instituciones cubanas teniendo en cuenta el cumplimiento de los indicadores propuestos en la hipótesis de la investigación, el criterio de los expertos la satisfacción de los usuarios.

Capítulo 1 - Almacenamiento de sistemas informáticos

En este capítulo se describen las arquitecturas de almacenamientos que existen, así como las principales tecnologías, estándares y protocolos existentes para la gestión del almacenamiento y las herramientas que lo implementan. También se estudia el estándar SMI-S como iniciativa para la gestión de sistemas de almacenamientos.

2.1 - Conceptos asociados al almacenamiento de datos en sistemas informáticos

Un sistema de almacenamiento centralizado, en un servidor dedicado a la persistencia de los datos que provee mayor capacidad de almacenamiento (Pradhan, Cai 2015). El acceso mediante una única interfaz, a través los principales protocolos de comunicación permite que los servidores interactúen contra un sólo servicio de almacenamiento.

La centralización de los datos permiten desplegar sistemas de alta disponibilidad, donde la redundancia y la replicación desempeñan un papel fundamental. Implementar mecanismos que implementen estas funciones, sólo es posible en servidores dedicados al almacenamiento donde existen un conjunto de arreglos de discos físicos que permiten esta gestión.

La gestión de un almacenamiento centralizado provee mayor control en la protección de datos, control de versiones y seguridad (Xu, Zhao, Zhang 2017). Proporciona un conjunto único y coherente de datos. Facilita el control de la configuración del hardware, la capacidad y el rendimiento.

2.1.1 - Almacenamiento

Datos informáticos: Datos que se interpretan dentro de un contexto de almacenamiento o comunicación utilizado por una una aplicación o un proceso (Agrawal, Choudhary 2016).

Un **dispositivo de almacenamiento** de datos es un conjunto de componentes utilizados para leer o grabar datos en el soporte de almacenamiento de datos, en forma temporal o permanente (Wang, Lu 2014).

Un **sistema de almacenamiento:** Una colección de dispositivos de almacenamiento de uno o más subsistemas de almacenamiento comúnmente accesibles, combinados con un cuerpo de software de control (SNIA 2018a).

Almacenamiento de archivos: También llamado almacenamiento basado en archivos o a nivel de archivos, archiva los datos en una estructura jerárquica. Los datos se guardan en archivos y carpetas y se presentan tanto al sistema que lo almacena como al sistema que lo recupera en el mismo formato (M. Rouse 2017a).

Almacenamiento en bloques: Es un tipo de almacenamiento de datos que se suele utilizar en entornos de red de área de almacenamiento (SAN) donde los datos se almacenan en volúmenes, también denominados bloques (M. Rouse 2017b).

Almacenamiento conectado a la red: El almacenamiento conectado a la red es un tipo de dispositivo de almacenamiento de archivos dedicado, que proporciona nodos de red de área local (LAN) con almacenamiento compartido basado en archivos a través de una conexión Ethernet estándar (M. Rouse 2017c).

2.2 - Arquitecturas de almacenamiento

Las arquitecturas de almacenamiento muestran como se distribuye físicamente los componentes que interactúan en una infraestructura que implementa almacenamiento conectado a la red. Existen tres tipos de arquitectura: Almacenamiento de Conexión Directa (DAS, por sus siglas en inglés), Área de almacenamiento en la red (SAN, por sus siglas en inglés) y Almacenamiento Conectado a la Red (NAS, por sus siglas en inglés).

DAS: Es el método de almacenamiento tradicional y consiste en conectar el dispositivo de almacenamiento directamente al ordenador o servidor. Su desventaja es que no suele ser un almacenamiento compartido por otros equipos (Lipinski 2017).

- No es una tecnología costosa. Si se conecta directamente al servidor, el costo puede ser tan simple como comprar discos duros.
- No se puede implementar mecanismos de compartición de archivos. Está dedicado al servidor al que está conectado. Los recursos no se pueden compartir fácilmente con otro servidor, especialmente si los discos duros se conectan directamente al chasis del servidor.
- No es escalable. La conectividad al servidor estará limitada por la cantidad de ranuras de expansión disponibles para acomodar las tarjetas SCSI y SAS.
- No tiene el rendimiento de las soluciones NAS y SAN.

SAN: Es un almacenamiento en el que los equipos cliente tienen la capacidad de leer y escribir

directamente sobre el volumen compartido, como si fuera un almacenamiento local. Por otro lado, su conexión por Fibre Channel le aporta mucha menor latencia y un mejor promedio de ancho de banda sostenido (John Wiley 2010).

- Se pueden implementar mecanismos de compartición. Los recursos se pueden compartir fácilmente entre múltiples dispositivos de hardware de servidor virtual. Esto es importante en entornos de servidores virtuales agrupados.
- Son altamente escalables, tanto desde una perspectiva de capacidad como de rendimiento. Para grandes despliegues de servidores virtuales, la escalabilidad es extremadamente importante y la premisa completa de SAN es proporcionar altos niveles de escala.
- El almacenamiento SAN proporciona replicación sincrónica. La replicación sincrónica no suele estar disponible en las implementaciones de NAS, y para muchos entornos, este puede ser un requisito clave para una estrategia de recuperación de desastres.

NAS: Es un almacenamiento al que se accede por red, donde un ordenador actúa de servidor y comparte el volumen a los equipos que lo solicitan. El servidor actúa como intermediario y es quien lee y escribe en el volumen compartido (John Wiley 2010).

- Presenta mecanismos de compartición de archivos. Los recursos pueden ser compartidos por múltiples protocolos.
- Es escalable en términos de capacidad y rendimiento. La concentración de la capacidad en arreglos de discos provee la funcionalidad de incrementar el almacenamiento en cualquier momento del ciclo de vida del servidor. El rendimiento depende de la aplicación de las buenas prácticas asociadas al despliegue del servidor. Una eficiente configuración de las tecnologías de almacenamiento proporciona buen rendimiento en el servicio.
- Proporciona funciones avanzadas para la replicación, la toma de instantáneas y restricción de almacenamiento por cliente. La restricción de almacenamiento se puede usar para limitar la cantidad de almacenamiento real proporcionado a los clientes, que no usará todo el almacenamiento con el que se presentan. Las instantáneas permiten capturar sistemas de archivos para copias de seguridad de datos o para clonar, donde se crean múltiples sistemas de archivos virtuales a partir de una única toma.
- Abstrae la administración de almacenamiento del servidor. Esto significa que los volúmenes de almacenamiento virtuales pueden administrarse y posteriormente respaldarse como un sistema de

archivos.

- Arquitectura parcialmente costosa. A pesar del costo, la gama de dispositivos NAS en el mercado es extremadamente amplia, desde pequeños dispositivos de cuatro discos hasta configuraciones de hardware de varios petabytes.

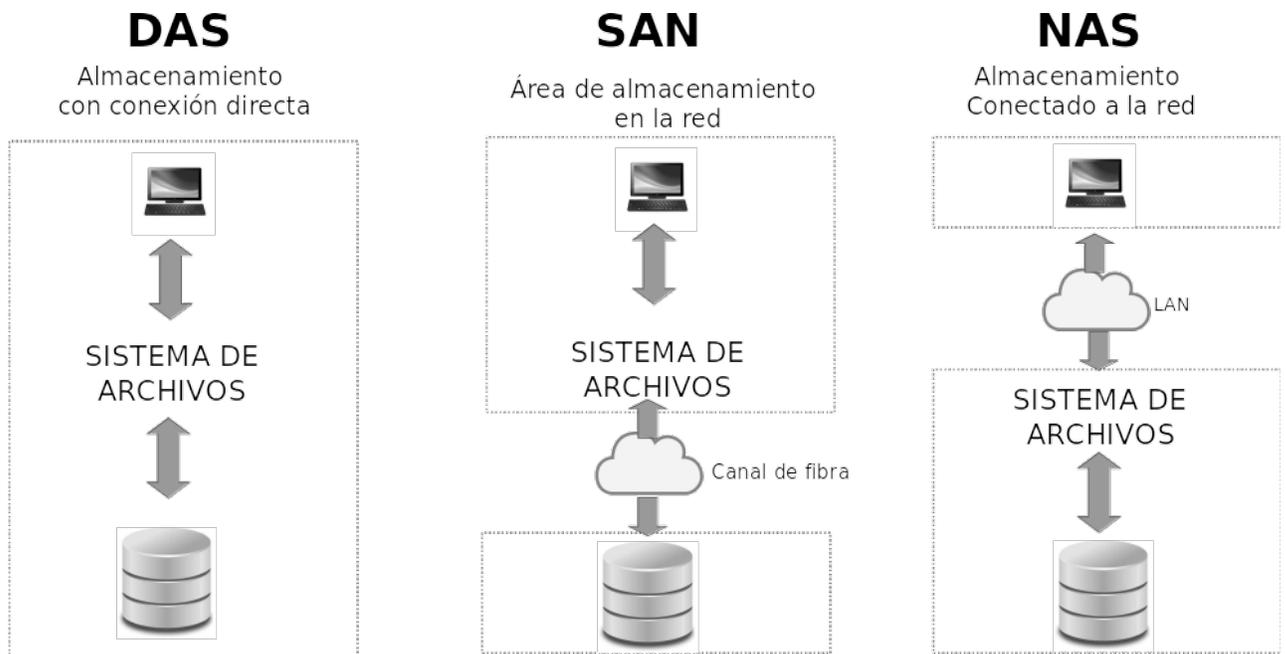


Figura 1: Interacción entre componentes en el despliegue de las arquitecturas de almacenamiento. Fuente: elaboración propia

2.2.1 - Definición de la arquitectura

A continuación se realiza una comparación entre las arquitecturas de almacenamiento para definir cual es la arquitectura a utilizar para el despliegue de la solución de almacenamiento en las instituciones cubanas. Para realizar la comparación se van a tener en cuenta a partir de los análisis escritos en (John Wiley 2010) y (Daehee Kim, Sejun Song, Baek-Young Choi 2015) los siguientes aspectos:

- **Transmisión de datos:** Tecnología usada por los clientes para establecer la comunicación con el sistema de almacenamiento.
- **Compartición:** Esquema de compartición de los datos.
- **Capacidad de conexión:** Cantidad de clientes que se pueden conectar al sistema de almacenamiento.

- **Velocidad de acceso:** Velocidad en las operaciones de lectura y escritura en el sistema de almacenamiento.
- **Expansión del almacenamiento:** Capacidad de expandir o aumentar el almacenamiento con el sistema funcionando.
- **Administración por software:** Capacidad de administración por software.
- **Administración centralizada:** Capacidad de administrar las tecnologías que implementan el sistema de almacenamiento.
- **Costo:** Costo de implementación del sistema de almacenamiento.

Tabla 1: Comparación entre las arquitecturas de almacenamiento. Fuente: elaboración propia

Aspectos a comparar	SAN	DAS	NAS
Transmisión de datos	Canal de fibra	IDE/SCSI	TCP/IP
Compartición	Compartición de datos a nivel de bloque	Sistema de archivo físicamente conectado a un cliente	Compartición de datos en red
Capacidad de conexión	Sin límites	Limitado	Sin límites
Velocidad de acceso	Muy rápido	Rápido	Rápido
Expansión del almacenamiento	Expansión en caliente ²	Requiere apagar el sistema de almacenamiento	Expansión en caliente
Administración por software	No	No	Si
Administración centralizada	Si	No	Si

² Ejecutar operaciones en un servidor de almacenamiento sin tener que apagarlo.

Costo	Caro	Relativamente barato	Barato
-------	------	----------------------	--------

Teniendo en cuenta los aspectos a comparar las diferentes arquitecturas existentes para la gestión de un sistema de almacenamiento se puede concluir que la arquitectura NAS es la seleccionada para el desarrollo de la solución de almacenamiento en las instituciones cubanas. Esta arquitectura tiene un costo de implementación barato, permite la administración por software lo que propicia la instalación de la distribución cubana GNU/Linux Nova en su variante para servidores. Permite la administración centralizada de las tecnologías que configuren un sistema de almacenamiento. Otro de los aspectos importantes es la forma de compartición de archivos en red mediante el protocolo TCP/IP punto importante porque permite la lectura y escritura en el sistema de almacenamiento de múltiples clientes incluyendo otros servicios que requieran guardar información en centros de datos aislados.

2.3 - Análisis de herramientas que implementan la arquitectura NAS

Las tecnologías que intervienen en el despliegue de una arquitectura NAS tienen como sistema base un sistema operativo en donde el cual se ejecutan y brindan los diferentes servicios de compartición de archivos dando la posibilidad de que los clientes puedan leer y escribir información en el sistema de almacenamiento. A continuación se analizan las principales herramientas de código abierto que centralizan la administración de un sistema de almacenamiento basado en la arquitectura NAS para determinar cuales son los componentes y tecnologías que los conforman. Se escogen las siguientes herramientas basado en las tendencias de búsquedas de Google de los usuarios respecto a estos términos, la calificación basada en la multitud que brinda el sitio *alternativeto.net* y el análisis de las herramientas descrito en la tesis “*Análisis, diseño y propuesta de un servidor nas creado a base de raspberry pi en la mayorista de turismo alltravel através de un sistema operativo open source*” (Durán 2017).

2.3.1 - FreeNAS

FreeNAS es un sistema operativo que se puede instalar en prácticamente cualquier plataforma de hardware para compartir datos a través de una red. Es la forma más sencilla de crear un lugar centralizado y de fácil acceso para sus datos. Mediante la tecnología ZFS se puede proteger, almacenar y realizar copias de seguridad a los datos (FreeNAS 2018a).

Características:

- Permite la compartición de archivos implementando compatibilidad con archivos compartidos a Windows mediante el protocolo SMB, recursos compartidos a sistemas Unix mediante el protocolo NFS y recursos compartidos de archivos a Apple mediante el protocolo AFP, así como FTP y iSCSI para el uso compartido de bloques.
- Implementa funcionalidades de protección de datos mediante la tecnología ZFS la cual está orientada a garantizar la integridad de la información mediante técnicas de sumas de verificación al sistema de archivos.
- Permite la creación de instantáneas al sistema de archivos para en caso de pérdidas de información de puedan restablecer las copias de seguridad. Estas copias de seguridad no son solo locales también permite realizar copias de seguridad remotas.
- Permite el cifrado de sistema de archivos mediante la encriptación AES-XTS estándar lo cual ofrece seguridad a los datos. Estos sistemas solo pueden leerse por sistemas que implementen el sistema operativo FreeNAS.

2.3.2 - NAS4Free

El sistema operativo NAS4Free se puede instalar prácticamente en cualquier plataforma de hardware para compartir el almacenamiento de datos de la computadora a través de una red informática. Es la forma más simple y rápida de crear un servidor centralizado y de fácil acceso para todo tipo de datos de fácil acceso con todo tipo de protocolos de red y desde cualquier red (iXsystems 2018).

Características:

- Permite la gestión de unidades de disco y gestión de volúmenes y tarjeta aceleradora criptográfica si está presente en el hardware.
- Permite soporte para los protocolos de red como SMB/CIFS (Samba), Samba AD, AFP, NFS, FTP (ProFTPD), TFTP (tftp-hpa), RSYNC (Client/Server), Unison, SCP (SSH), iSCSI target.
- Permite el monitoreo de los recursos de hardware mediante un servidor de estadísticas históricas. También implementa el seguimiento en tiempo real del CPU y los recursos de red.

2.3.3 - Openmediavault

Openmediavault es una solución de almacenamiento conectado en red basada en Debian Linux. Contiene servicios como SSH, FTP, SMB/CIFS, servidor de medios DAAP, Rsync y cliente BitTorrent (V. Theile 2018).

Características:

- Gestión de discos, volúmenes lógicos y sistemas de archivos.
- Compartición de archivos mediante los principales protocolos existentes: FTP, SMB/CIFS, servidor de medios DAAP, Rsync.
- Posee un sistema de notificaciones por correo electrónico
- Permite la administración de algunos recursos del sistema operativo como son las tareas programadas, la generación de certificados autofirmados y la configuración de la red.
- Ejecuta tareas de monitoreo de discos y guarda el historial del consumo de los recursos de hardware asociados al CPU, SWAP, las interfaces de red y volúmenes lógicos.

2.3.4 - Elementos que complementan la gestión del almacenamiento

El análisis anteriormente descrito sobre las herramientas que implementan arquitecturas NAS ofrece una panorámica sobre el comportamiento de las tecnologías que lo complementan. Estas tienen como objetivo ofrecer mecanismos para poder establecer la comunicación entre los clientes y el sistema de almacenamiento para ejecutar operaciones de lectura y escritura de datos.

Estas herramientas implementan una infraestructura donde interactúan tecnologías, protocolos y aplicaciones que alinean la gestión del almacenamiento creando mecanismos para optimizar el uso de dispositivos de almacenamiento, proteger la integridad de los datos y garantizar el acceso a los recursos por parte de los clientes.

Teniendo en cuenta las características de la herramientas existen cuatro elementos que hay que tener en cuenta para el desarrollo de una solución de almacenamiento y que contribuyen a la eficiencia de esta gestión:

- **Capacidad de almacenamiento:** Consiste en proporcionar el suficiente almacenamiento para garantizar el alojamiento de la información y cual va a ser la distribución física y lógica de los volúmenes.

- **Monitoreo del rendimiento:** Consiste en conseguir la mejor configuración para el rendimiento óptimo del sistema, incluye tener mecanismos de balanceo de carga en las interfaces de redes y monitoreo de los recursos de hardware.
- **Disponibilidad para el acceso:** El almacenamiento tienen que garantizar el acceso a los recursos en cualquier momento mediante los principales protocolos de comunicación que existen. Tiene que tener mecanismo que aumenten la tolerancia a fallos.
- **Seguridad:** Consiste en tener mecanismos para minimizar la pérdida de la información, aplicando métodos para realizar copias de seguridad y aplicar reglas para regir el control de acceso de los clientes al sistema de almacenamiento.

2.4 - Especificaciones para la gestión del almacenamiento propuestas por SNIA

La Asociación de la Industria de Redes de Almacenamiento (*SNIA por sus siglas en inglés*) es una asociación de productores y consumidores de productos de redes de almacenamiento de datos informáticos. Su principal objetivo es garantizar que las redes de almacenamiento se conviertan en soluciones completas y de confianza en toda la comunidad de la TI (SNIA 2018a).

Uno de los resultados más importantes es el desarrollo y mantenimiento de la Especificación de la Iniciativa de gestión de almacenamiento, comúnmente denominada SMI-S (*por sus siglas en inglés*), la cual es un estándar de almacenamiento de datos informáticos (Wang, Kung 2018).

2.4.1 - Perfiles para la Especificación de la Iniciativa de gestión de almacenamiento en NAS

SMI-S define perfiles de gestión CIM³ para sistemas de almacenamiento. La especificación SMI completa se categoriza en perfiles y subperfiles. Un perfil describe los aspectos de comportamiento de un dominio de gestión autónomo e independiente. SMI incluye perfiles para matrices de discos⁴, *Switches*⁵, gestores de volúmenes y otros dominios dentro de la gestión del almacenamiento (SNIA 2018b).

En un nivel muy básico, las entidades SMI-S se dividen en dos categorías:

- Los clientes son aplicaciones de software de administración que pueden residir virtualmente en

3 Modelo de información Común (CIM) es un estándar abierto que define cómo se representan los elementos gestionados en un entorno de TI como un conjunto común de objetos y relaciones entre ellos.

4 Proporciona mayor disponibilidad, flexibilidad y capacidad de mantenimiento mediante el uso de componentes existentes

5 Dispositivo de red informática que conecta dispositivos en una red informática

cualquier lugar dentro de una red, siempre que tengan un enlace de comunicaciones (ya sea dentro de la ruta de datos o fuera de la ruta de datos) a los proveedores.

- Los servidores son los dispositivos bajo administración. Los servidores pueden ser matrices de discos, motores de virtualización, conmutadores, unidades de cinta, entre otros.

Un sistema de almacenamiento se debe desplegar en un servidor de almacenamiento para poder proveer un servicio de almacenamiento a los clientes que requieran leer y escribir información en el sistema de almacenamiento. A continuación se listan cuales son los perfiles que SMI-S propone para el correcto despliegue de un sistema de almacenamiento basado en el sistema de archivos:

- **Perfil principal de NAS (sistemas de archivos):** Ese perfil propone que todo sistema de almacenamiento tiene que proveer mecanismos de administración del almacenamiento subyacente y exportar los archivos como archivos compartidos.
- **Perfil de almacenamiento de archivos:** Un recurso compartido tiene que estar asociado al sistema de archivos subyacente.
- **Perfil de exportación de archivos (sistemas de archivos):** Este perfil tiene como propósito proponer mecanismos para enumerar recursos compartidos existentes en un servidor de archivos.
- **Perfil de autorización basado en roles:** Para enumerar y configurar permisos de acceso a un archivo compartido.
- **Perfil de manipulación de exportación de archivos:** Propone que todo sistema de almacenamiento tiene que tener la capacidad para crear, eliminar y modificar los recursos compartidos.
- **Perfil de manipulación del sistema de archivos:** Propone que deben existir mecanismos para la gestión del sistema de archivo.
- **Perfil de red:** Perfil que establece técnicas para enumerar puertos de red asociados a un recurso compartido con el objetivo de asociar puertos de salida a un recurso compartido.
- **Perfil del sistema de archivos recuperable:** Tiene que existir técnicas de recuperación de la información sobre el sistema de archivos subyacente de un archivo compartido.
- **Perfil de rendimiento del sistema de archivos** Este perfil se basa en el monitoreo y la supervisión del rendimiento del sistema de archivos.

Los perfiles que propone la SMI-S para la gestión de sistemas de almacenamientos en arquitecturas NAS

están basados principalmente en la manipulación del sistema de archivos. Estos ofrecen como se debe gestionar las carpetas compartidas para brindar servicios de acceso a la información donde tiene que existir mecanismos de exportación de carpetas, control de acceso basado en la autorización por roles y técnicas para la recuperación de información y así aumentar la disponibilidad de los datos.

2.5 - Análisis de interoperabilidad

En este epígrafe se describen las aplicaciones que están autorizadas en la guía cubana de migración a código abierto que requieren para su funcionamiento la disposición de un sistema de almacenamiento para almacenar sus datos y están disponibles en el repositorio de la Distribución Cubana GNU/Linux Nova o forman parte de la propia distribución. También se describen cuales son los sistemas que se desarrollan bajo la alianza UCI-GEDEME que tienen que ser compatibles con la solución de almacenamiento.

La interoperabilidad es la capacidad de dos o más sistemas o componentes para intercambiar información y utilizar la información intercambiada (IEEE, 1990) y (Cretella, Esposito 2015). En el despliegue de la solución de almacenamiento de datos se va a tener en cuenta:

- La integración de las aplicaciones existentes en el repositorio de la Distribución Cubana GNU/Linux Nova que necesiten un servicio de almacenamiento para interactuar y que están avaladas en la Guía Cubana de Migración a Código Abierto.
- La compatibilidad de los sistemas desarrollados bajo la alianza UCI-GEDEME que requieran de un servicio de almacenamiento para su correcto funcionamiento como son:
 - NovaLTSP⁶ : Almacenamiento de las imágenes de sistemas operativos.
 - Nova360⁷ : Almacenamiento de archivos de los usuarios.
- La capacidad de despliegue en la Distribución Cubana GNU/Linux Nova en su variante para servidores.

2.5.1 - Aplicaciones integradas a la Distribución Cubana GNU/Linux Nova

En la Distribución Cubana GNU/Linux Nova existen un conjunto de aplicaciones que tienen mecanismos que les permite la integración con un sistema de almacenamiento. La explotación de estas funcionalidades ofrece que se pueda aprovechar todo el potencial de la aplicación además que en muchas

⁶ Plataforma de administración de cliente ligeros.

⁷ Solución corporativa para la sincronización de estaciones de trabajo, teléfonos móviles y la nube.

ocaciones de ahorre el espacio en el disco local de las estaciones de trabajo. Algunas de estas aplicaciones son:

Navegador de archivos en Nova

Nautilus: Es el administrador del sistema de archivos de Nova, permite la gestión de ficheros y carpetas locales del sistema. Con esta herramienta se puede conectar a un servidor o recurso compartido de red para explorar y ver archivos en ese servidor, exactamente como si estuvieran en su propia computadora. Esta es una forma conveniente de descargar o cargar archivos en Internet, o de compartir archivos con otras personas en su red local. Es compatible con:

- SSH: Es un protocolo de conexión seguro a un servidor. Muchos servidores ofrecen cuentas de SSH a los miembros para que puedan cargar archivos de manera segura. Los servidores SSH siempre requieren que inicies sesión. Una url de conexión ssh tiene la siguiente estructura:

`ssh://usuario@nombre_servidor.ejemplo.cu/carpeta`

- FTP (con inicio de sesión): FTP es una forma popular de intercambiar archivos en Internet. Debido a que los datos no están encriptados a través de FTP, muchos servidores ahora brindan acceso a través de SSH. Sin embargo, algunos servidores aún permiten o requieren que use FTP para cargar o descargar archivos. Los sitios FTP con inicios de sesión generalmente le permitirán eliminar y cargar archivos. Una url de conexión FTP tiene la siguiente estructura:

`ftp://usuario@nombre_servidor.ejemplo.cu/carpeta`

- FTP público: Los sitios que le permiten descargar archivos a veces brindan acceso FTP público o anónimo. Estos servidores no requieren un nombre de usuario y contraseña, y generalmente no le permitirán eliminar o cargar archivos. Una url de conexión FTP tiene la siguiente estructura:

`ftp://nombre_servidor.ejemplo.cu/carpeta`

- Compartidos en Windows: Las computadoras con Windows usan un protocolo propietario para compartir archivos a través de una red de área local. Las computadoras en una red de Windows a veces se agrupan en dominios para la organización y para controlar mejor el acceso. Si tiene los permisos correctos en la computadora remota, puede conectarse a un recurso compartido de Windows desde el administrador de archivos. Una url de conexión tiene la siguiente estructura:

`smb://nombre_servidor.ejemplo.cu/carpeta`

- WebDAV: Basado en el protocolo HTTP utilizado en la web, WebDAV a veces se usa para

compartir archivos en una red local y para almacenar archivos en Internet. Una url de conexión tiene la siguiente estructura:

`dav://nombre_servidor.ejemplo.cu/carpeta`

- Compartido NFS: Las computadoras UNIX tradicionalmente usan el protocolo del Sistema de archivos de red para compartir archivos a través de una red local. Con NFS, la seguridad se basa en el UID del usuario que accede al recurso compartido, por lo que no se necesitan credenciales de autenticación al conectarse. Una url de conexión tiene la siguiente estructura:

`nfs://nombre_servidor.ejemplo.cu/carpeta`

Reproductores de audio

- DAAP: iTunes de Apple popularizó el Protocolo de Acceso de Audio Digital (DAAP) para el intercambio simple de listas de reproducción de música en red. En la Distribución Cubana GNU/Linux Nova existen varios reproductores de audio que brindan soporte para reproducir listas de reproducción de audio mediante el protocolo DAAP:
- Rhythmbox: es un reproductor de audio que reproduce y ayuda a organizar el audio digital. Rhythmbox es un software gratuito, diseñado para funcionar bien bajo el escritorio GNOME utilizando el framework de medios GStreamer. Es compatible con el protocolo de red DAAP.
- Banshee: es un reproductor de medios de código abierto multiplataforma, llamado Sonance hasta 2005. Basado en Mono y Gtk, utiliza la plataforma multimedia GStreamer para codificar y decodificar varios formatos de medios, incluidos Ogg Vorbis, MP3 y FLAC.

2.5.2 - Productos desarrollados bajo la alianza UCI-GEDEME

A continuación se describen las principales funcionalidades que tienen los sistemas desarrollados bajo la alianza UCI-GEDEME que responden a la necesidad de comunicación son una solución de almacenamiento y que son desplegados bajo la Distribución Cubana GNU/Linux Nova en su variante para servidores.

- Nova-Unificado: Es una solución desarrollada para los servidores que comercializa GEDEME orientada a la administración de servicios telemáticos para pequeñas empresas. Tiene funcionalidades para configurarlo como puerta de enlace, administrador de infraestructura, administrador de amenazas unificadas, servidor de oficina, servidor de comunicaciones unificadas o una combinación de ambos. Desde esta herramienta se pueden administrar muchos servicios

que requieren la centralización del almacenamiento en un sistema de gestión del almacenamiento. A continuación se listan por cada servicio telemático cuales son los requerimientos que asocian esta tecnología a una infraestructura NAS:

Servicio de correo electrónico: Es una aplicación de red de computadoras ubicada en un servidor de Internet, para prestar servicio de correo electrónico. En este caso al aumentar el número de usuarios que consumen este servicio aumenta la capacidad de almacenaje para los buzones de correo asociado a cada usuario. En este sentido se necesita almacenar los buzones de correo en un servidor NAS mediante los protocolo NFS o FTP.

Servicio FTP: El protocolo de transferencia de archivos es un protocolo de red estándar utilizado para la transferencia de archivos informáticos entre un cliente y un servidor en una red informática. En este caso se compartiría una carpeta en el servidor NAS mediante el protocolo NFS para la lectura y escritura de datos en el servidor FTP.

- Nova-LTSP: Es una solución desarrollada para los servidores que comercializa GEDEME orientada a la administración de clientes ligeros. Los clientes ligeros requieren de una imagen de sistema operativo que en ocasiones puede consumir mucho espacio en el servidor. También cuando aumentan el número de usuario que interactúan con el servidor de clientes ligeros aumenta la cantidad de carpetas personales alojadas en el servidor. Los dos factores descritos anteriormente evidencian que es necesario que tanto las imágenes de sistema operativos como las carpetas personales de los usuarios se alojen en el servidor NAS:

Las carpetas personales se usar el protocolo NFS.

Las imágenes de sistema operativo pueden usar el servicio TFTP.

- Nova360: Es una tecnología desarrollada para los servidores que comercializa GEDEME basada en una solución corporativa la sincronización entre las estaciones de trabajo personales, los teléfonos móviles y la nube. Esta solución permite la sincronización de archivos y carpetas, calendarios, notas y tareas. Teniendo en cuenta los grandes volúmenes de datos que puede almacenar esta solución se requiere de una solución de almacenamiento para centralizar los datos de los usuarios.

2.6 - Conclusiones parciales

- Mediante la aplicación de métodos de comparación se definió NAS como arquitectura para el

despliegue de la solución de almacenamiento de datos en las instituciones cubanas.

- El estudio de las herramientas OpenmediaVault, FreeNAS y Free4NAS permitieron definir que los elementos a tener en cuenta para el despliegue de la solución del almacenamiento son la gestión de la capacidad de almacenamiento, la seguridad, el acceso a los datos y el monitoreo de los recursos.
- Mediante el análisis de la Especificación de la Iniciativa de Gestión de Almacenamiento y de la Alianza para la Gestión del almacenamiento se obtuvo un conjunto de buenas prácticas para el despliegue de la solución tecnológica.
- Durante el análisis de las herramientas que contienen funcionalidades para establecer comunicación con un sistema de almacenamiento se definieron cuales son los principales protocolos de comunicación que existen y los formatos de datos necesario para el intercambio.

Capítulo 2: Solución para el almacenamiento en las instituciones cubanas

En este capítulo se define la solución tecnológica para el despliegue de un sistema de gestión del almacenamiento en las instituciones cubanas. La solución esta basada en un conjunto de componentes con objetivos bien definidos que enmarcan funcionalidades en la gestión del almacenamiento. Se proponen como se relacionan los componentes y que tecnologías los componen.

3.1 - Componentes para la gestión del almacenamiento

En la gestión del almacenamiento intervienen un conjunto de tecnologías que propician el correcto funcionamiento de este servicio permitiendo que clientes puedan escribir y leer información en cualquier momento garantizando la seguridad, disponibilidad e integridad de la información. A continuación se describen los componentes a tener en cuenta para el despliegue de la solución de almacenamiento en las instituciones cubanas.

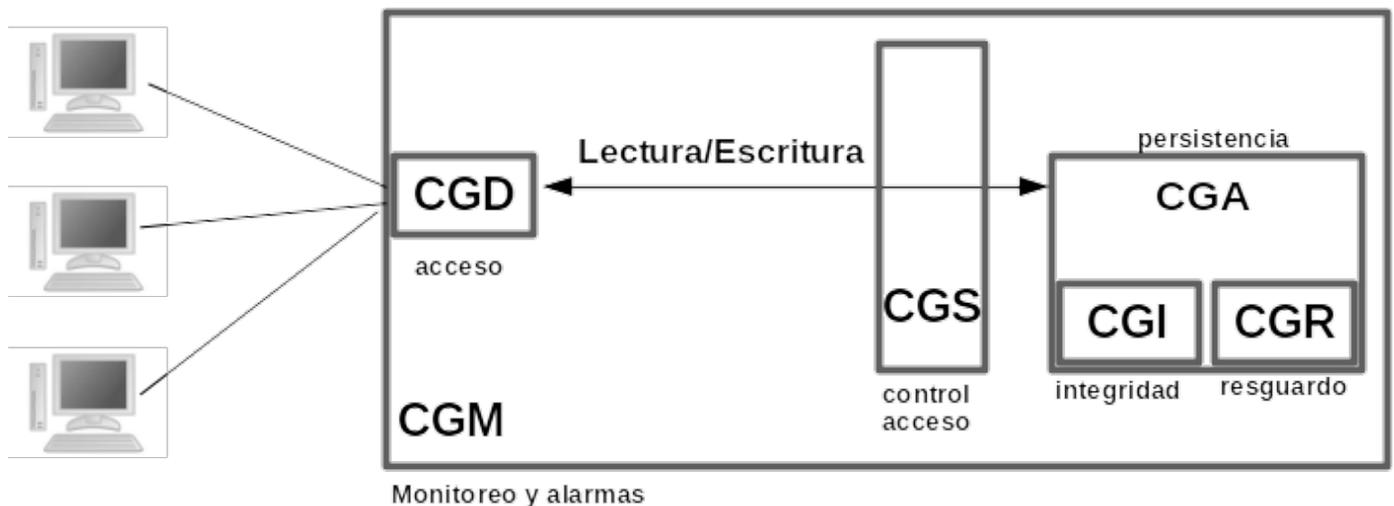
- **Componente para la gestión del almacenamiento (CGA):** Agrupa un conjunto de herramientas y tecnologías que garantizan la persistencia de la información en el sistema de almacenamiento. Es el responsable de la administración de los discos físicos y volúmenes lógicos destinados a la escritura de los datos. Crea y elimina los sistemas de archivos así como las carpetas compartidas.
- **Componente para la gestión de la disponibilidad (CGD):** Contiene la configuración y administración de las tecnologías que permiten la compartición de archivos y carpetas en la red. Implementa los principales protocolos de comunicación entre los clientes y el sistema de almacenamiento. Tiene la responsabilidad de garantizar la escritura y lectura que establecen los clientes con el sistema de almacenamiento.
- **Componente para la gestión de la seguridad (CGS):** Este componente tiene la responsabilidad de garantizar la seguridad en el sistema de almacenamiento. Implementa tecnologías que limitan el control de acceso a los recursos compartidos. Garantiza sistemas de prevención y detección de intrusos, así como mecanismos antihackeo.
- **Componente para la gestión de la integridad (CGI):** Este componente implementa tecnologías para la detección y eliminación de programas malignos que se puedan alojar en el sistema de almacenamiento. Para lograr la integridad de la información se despliegan mecanismos para el chequeo de sumas de verificación a los archivos para la detección de cambios. Este componente

implementa encriptación de sistemas de archivo.

- **Componente para la gestión de respaldos (CGR):** Este componente tiene la responsabilidad de implementar un sistema de respaldo para la información escrita en el sistema de almacenamiento. Está diseñado para automatizar tareas de copia de seguridad que a menudo requieren la intervención de un administrador de sistemas o un operador de computadora.
- **Componente de gestión del monitoreo (CGM):** Este componente tiene la responsabilidad de realizar el control y seguimiento del sistemas de almacenamiento, los servicios que garantizan la disponibilidad de la información y los recursos del servidor. Implementa tecnología para la emisión de notificaciones y alarmas.

3.1.1 - Distribución de los componentes

Figura 2: Distribución de componentes en el sistema de almacenamiento. Fuente: elaboración propia



Cuando se despliega el sistema de almacenamiento los clientes acceden al componente de gestión de la disponibilidad para realizar operaciones de lectura y escritura en el servidor de almacenamiento mediante los protocolos existentes para establecer la comunicación entre los clientes y el servidor. El componente de gestión de la seguridad regula el control de acceso a los recursos y valida la autenticidad de los clientes verificando si tienen permiso para realizar operaciones de escritura y lectura en el sistema de almacenamiento. El componente de gestión del almacenamiento garantiza la persistencia de la información y la distribución de los datos en los volúmenes físicos y lógicos, por otra parte el componente

de gestión de la integridad se mantiene monitoreando la información en búsqueda de programas malignos. Las acciones de resguardo están a cargo del componente de gestión de resguardo el cual tiene la función de realizar salvallas periódicas de la información. Finalmente el componente de monitoreo está en constante vigilancia sobre los recursos de hardware y velando por el buen funcionamiento de los servicios de disponibilidad de los datos. En la Figura 2 se ilustra la interacción entre los componentes.

3.1.2 - Componente para la gestión del almacenamiento (CGA)

El almacenamiento constituye la persistencia en el tiempo de los datos. Es un conjunto de componentes utilizados para leer o grabar datos en el soporte de almacenamiento de datos, en forma temporal o permanente. Está constituido por un grupo de tecnologías que administran la disposición del almacenamiento físico.

Tecnologías, estándares y protocolos aplicados

- **Logical Volume Management** es una tecnología de administración de dispositivos de almacenamiento que brinda a los usuarios la capacidad de agrupar y abstraer el diseño físico de los dispositivos de almacenamiento de componentes para una administración más fácil y flexible. Utilizando el mapeador de dispositivos *Linux kernel framework*, la iteración actual, LVM2, se puede usar para reunir dispositivos de almacenamiento existentes en grupos y asignar unidades lógicas del espacio combinado según sea necesario (Matotek, Turnbull, Lieverdink 2017).
- **RAID** es una tecnología de virtualización de almacenamiento de datos que combina múltiples componentes de discos físicos en una o más unidades lógicas para fines de redundancia de datos, mejora del rendimiento o ambos (Vadala 2002) y (Bartlett, Fairhurst 2017). Los datos se distribuyen a través de las unidades de discos de varias maneras, conocidos como niveles de RAID, dependiendo del nivel requerido de redundancia y rendimiento.
 - **Niveles de RAID:**
 - RAID-0: El controlador divide los datos en bloques y los escribe en diferentes discos en modo round-robin⁸.
 - RAID-1: El controlador escribe los datos de forma duplicada bloque por bloque en cada disco miembro del arreglo.

⁸ Disposición de elegir todos los elementos en un grupo igualmente en un orden racional, generalmente desde la parte superior a la parte inferior de una lista y luego comenzar de nuevo en la parte superior de la lista y así sucesivamente.

- RAID-4: El controlador corta bloques de datos de tamaño de bloque en cada unidad de la matriz marcada como unidad de datos. Además, una unidad se designa como una unidad de paridad dedicada.
- RAID-5: El controlador elimina el uso de una unidad de paridad dedicada y escribe la información de paridad en cada disco de la matriz, usando el mismo algoritmo XOR encontrado en RAID-4.
- **Calculo de la capacidad a partir del nivel aplicado**

Tabla 2: Capacidad obtenida al aplicar niveles de RAID. Fuente: elaboración propia

Nivel de RAID	Capacidad obtenida
RAID-0	Total de discos * Capacidad de los discos
RAID-1	Capacidad de los discos
RAID-4	(Total de discos - 1) * Capacidad de los discos
RAID-5	(Total de discos - 1) * Capacidad de los discos

- **ZFS** es un sistema de archivos combinado y un administrador de volúmenes lógicos diseñado por Sun Microsystems. Las características de ZFS incluyen protección contra corrupción de datos, soporte para altas capacidades de almacenamiento, compresión de datos, integración de los conceptos de administración de volumen y sistema de archivos, verificación de integridad continua y reparación automática de volúmenes (FreeNAS 2018b).
- **Unionfs** es un servicio de sistema de archivos que implementa mecanismos para ejecutar montajes unificados entre sistemas de archivos separados. Permite que los archivos y directorios de los sistemas de archivos separados se superpongan de forma transparente, formando un único sistema de archivos coherente. Los contenidos de los directorios que tienen la misma ruta dentro de las ramas fusionadas se verán juntos en un solo directorio fusionado, dentro del nuevo sistema de archivos virtual (ZADOK 2017).
- **Cuotas de disco** es una característica de Linux que permite que el administrador del sistema asigne una cantidad máxima de espacio en disco que un usuario. Puede ser flexible en su

cumplimiento de las reglas asignadas y se aplica por sistema de archivos.

Funcionalidades

- Soporte para el almacenamiento a largo plazo de datos y tamaños de almacén de datos de escala indefinida con cero pérdida de datos y alta capacidad de configuración.
- Cheques de sumas de verificación jerárquica de todos los datos y metadatos, asegurando que todo el sistema de almacenamiento se pueda verificar en el uso, y se confirme que se almacena correctamente o se corrige si está dañado.
- Puede almacenar una cantidad especificada por el usuario de copias de datos o metadatos, o tipos seleccionados de datos, para mejorar la capacidad de recuperación de daños en los datos de archivos y estructuras importantes.
- Retroceso automático de los cambios recientes en el sistema de archivos y los datos, en algunas circunstancias, en caso de error o incoherencia.
- Soporte para la restricción de cuotas de almacenamiento por usuarios.
- Soporte para la gestión de volúmenes lógicos.
- Auto-corrección automatizada y (por lo general) silenciosa de las inconsistencias de datos y errores de escritura cuando se detectan, para todos los errores donde los datos son capaces de reconstrucción.
- Auto-reconstrucción de datos mediante sumas de verificación de detección y corrección de errores almacenadas en el bloque primario de cada bloque y copias múltiples de datos (incluidas sumas de comprobación) en el disco.
- Soporte para la gestión de carpetas compartidas.
- Soporte para la gestión de volúmenes y formateo en los formatos EXT3, EXT4 y ReiserFS.
- Soporte para el manejo nativo de niveles de RAID estándar.
- Soporte para el manejo nativo de almacenamiento en niveles y dispositivos de almacenamiento en caché, que generalmente es una tarea relacionada con el volumen. Debido a que también comprende el sistema de archivos, puede usar el conocimiento relacionado con archivos para informar, integrar y optimizar su manejo de almacenamiento escalonado que un dispositivo separado no puede.
- Soporte para el manejo nativo de instantáneas del sistema de archivos de manera eficiente y

rápida.

- Soporte para la compresión de datos nativos.
- Reconstrucción eficiente de matrices RAID: una controladora RAID a menudo tiene que reconstruir un disco completo, pero ZFS puede combinar el conocimiento de archivos y discos para limitar cualquier reconstrucción a datos que realmente faltan o están dañados, acelerando enormemente la reconstrucción.
- Capacidad para identificar datos que se habrían encontrado en un caché pero que en su lugar se han descartado recientemente.

Despliegue

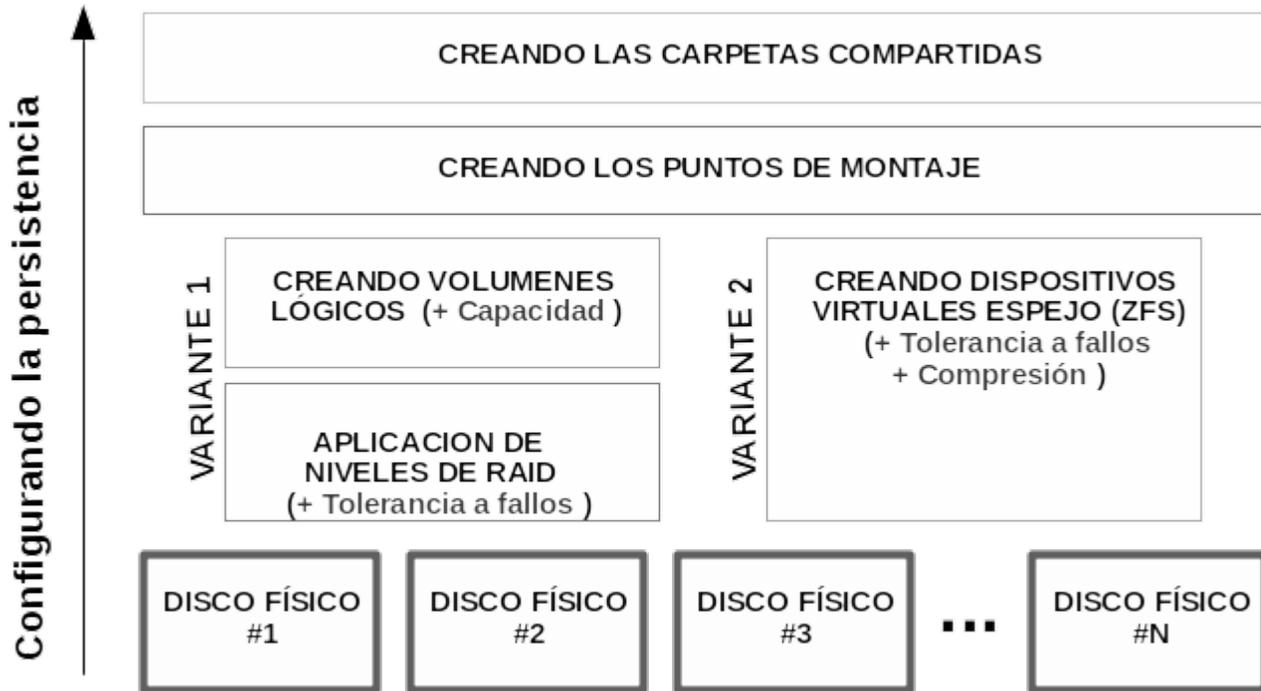
En la Figura 3 se muestra como se despliega el componente de gestión de almacenamiento. El componente de almacenamiento tiene la responsabilidad de administrar la persistencia de la información. El flujo comienza a partir de la disponibilidad de dispositivos físicos existentes en el servidor. Teniendo en cuenta la cuenta la cantidad de discos se pueden aplicar dos variantes:

1. **RAID+LVM:** Se seleccionan los niveles de RAID a aplicando en dependencia de la cantidad de los disco y su capacidad como se muestra en la Tabla 2. A continuación se aplica un mecanismo de volumen lógico para la agrupación de los sistemas de archivos en un sólo volumen virtual.
2. **XFS:** Se crea un volumen de datos que puede tener múltiples puntos de montaje, abarcando múltiples discos. Los discos se pueden combinar en grupos virtuales para permitir varias opciones de redundancia.

Independientemente de la variante seleccionada se crean los diferentes puntos de montaje asociados a los volúmenes creados. Se formatean bajo los formatos ext3, ext4 o reiserFS. Cuando se montan los volúmenes se crean las carpetas compartidas que van a contener los elementos públicos para ejecutar las tareas de escritura y lectura de información.

La interoperabilidad se maneja a través del soporte de los formatos de datos que se escriban en los diferentes dispositivos de almacenamiento durante las operaciones de escrituras.

Figura 3: Despliegue del componente de almacenamiento. Fuente: elaboración propia



3.1.3 - Componente para la gestión de la disponibilidad (CGD)

La gestión de la disponibilidad tiene la responsabilidad de proveer mecanismos de acceso a los datos que persisten en el sistema de almacenamiento. El acceso a la información se realiza a partir de la compartición de datos basada en los principales protocolos de comunicación. Compartir archivos es la práctica de distribuir o proporcionar acceso a medios digitales, como programas de computadora, multimedia (audio, imágenes y video), documentos o libros electrónicos. El intercambio de archivos se puede lograr de varias maneras. Los métodos comunes de almacenamiento, transmisión y dispersión incluyen el intercambio manual utilizando medios extraíbles y servidores centralizados de almacenamiento en redes informáticas.

Tecnologías, estándares y protocolos aplicados

- El **Protocolo de transferencia de archivos (FTP)** es un protocolo de red estándar utilizado para la transferencia de archivos informáticos entre un cliente y un servidor en una red informática (Reynolds 1985). El FTP puede ejecutarse en modo activo o pasivo, lo que determina cómo se establece la conexión de datos (Ko 2014). En ambos casos, el cliente crea una conexión de control TCP usualmente sin privilegios a través del puerto 21.

- El **Protocolo de transferencia de archivos SSH (SFTP)** es un protocolo de red que proporciona acceso a archivos, transferencia de archivos y administración de archivos a través de cualquier flujo de datos confiable. Fue diseñado por el Grupo de trabajo de ingeniería de Internet (IETF) como una extensión del protocolo Secure Shell (SSH) versión 2.0 para proporcionar capacidades seguras de transferencia de archivos. IETF Internet Draft afirma que, aunque este protocolo se describe en el contexto del protocolo SSH-2, podría utilizarse en una serie de aplicaciones diferentes, como la transferencia segura de archivos a través de *Transport Layer Security* (TLS) y la transferencia de gestión. información en aplicaciones VPN (Barrett, Silverman 2001) y (Bartok, Turner, Skehan 2018).
- El **bloque de mensajes del servidor (SMB)**, una versión del cual también se conoce como sistema común de archivos de Internet (CIFS), funciona como un protocolo de red de capa de aplicación utilizado principalmente para proporcionar acceso compartido a archivos, impresoras y puertos serie y comunicaciones entre nodos en una red. También proporciona un mecanismo de comunicación entre procesos autenticado (TechNet 2013).
- El **sistema de archivos de red (NFS)** es un protocolo de nivel de aplicación, según el Modelo OSI⁹. Es utilizado para sistemas de archivos distribuido en un entorno de red de computadoras de área local. Posibilita que distintos sistemas conectados a una misma red accedan a ficheros remotos como si se tratara de locales (Sun Microsystems 1989).
- El **protocolo de acceso a audio digital (DAAP)** es un protocolo ideado por la compañía Apple. Actualmente se utiliza mediante iTunes para intercambiar música a través de una red de trabajo o bien a través de Internet. Aunque todavía no existe una descripción oficial de este protocolo, se ha llevado a cabo ingeniería inversa que hace posible que éste pueda implementarse fuera de la plataforma iTunes. De hecho, un servidor DAAP es simplemente un servidor HTTP especializado, que es capaz de enviar y solicitar una lista de ficheros de audio (Barz, Bassett 2015).
- **SSH** es un protocolo de red criptográfica para operar servicios de red de forma segura a través de una red no segura. La aplicación de ejemplo más conocida es para el inicio de sesión remoto de los sistemas informáticos por parte de los usuarios (Lonvick 2006) y (Bartok, Turner, Skehan 2018).

Proporciona un canal seguro a través de una red no segura en una arquitectura cliente-servidor, conectando una aplicación de cliente SSH con un servidor SSH. Las aplicaciones comunes

⁹ El modelo de interconexión de sistemas abiertos (ISO/IEC 7498-1), más conocido como “modelo OSI”, (en inglés, Open System Interconnection) es un modelo de referencia para los protocolos de la red de arquitectura en capas.

incluyen el inicio de sesión remoto de la línea de comandos y la ejecución remota de comandos, pero cualquier servicio de red puede protegerse con SSH (Lonvick 2006).

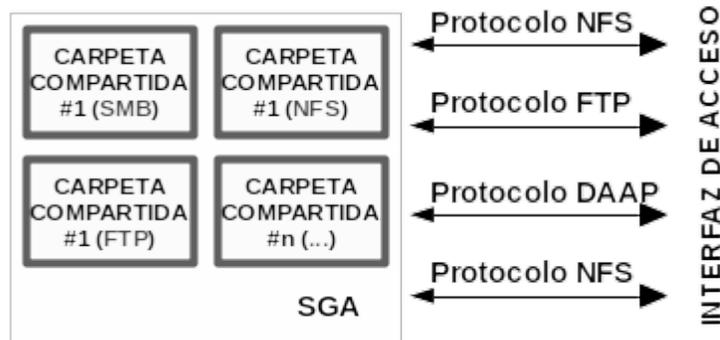
Funcionalidades

- Soporte para la compartición de archivos, compatibles con implementaciones descritas en sistemas operativos privativos mediante el protocolo SMB.
- Mecanismos para la administración remota segura del servidor mediante protocolo SSH.
- Soporte para la compartición de archivos de audio compatibles con reproductores de audio disponibles en distribuciones GNU/Linux mediante el protocolo DAAP.
- Soporte para la compartición de carpetas como si fueran sistemas de archivos locales mediante el protocolo NFS.
- Soporte para implementar copias seguras mediante el protocolo SFTP.
- Soporte para implementar comparticiones de archivos mediante FTP y TFTP.

Despliegue

En la Figura 4 se ilustra el despliegue del componente de disponibilidad dentro de la solución de almacenamiento de datos. La arquitectura permite la compartición de carpeta mediante los principales protocolos de comunicación que requieran o no algún tipo de autenticación con usuarios locales o con directorio activo. La solución muestra como una misma carpeta se puede compartir por diferentes protocolos de compartición de archivos para proveer un acceso más diverso.

Figura 4: Despliegue del componente de disponibilidad. Fuente: elaboración propia



La interoperabilidad se maneja a través del soporte de los diferentes protocolos de comunicación que se integren a la solución como mecanismo de comunicación con los diferentes dispositivos de

almacenamiento a través de las operaciones de escritura y lectura en las carpetas compartidas.

3.1.4 - Componente para la gestión de seguridad (CGS)

El proceso de gestión de la seguridad informática requiere el establecimiento de gran cantidad de controles, la implementación de variados sistemas de seguridad con mecanismos de gestión independientes, y una elevada capacidad de respuesta ante los diversos ataques y vulnerabilidades existentes (Perurena 2012).

Tecnologías, estándares y protocolos aplicados

- **Un Sistema de Prevención de Intrusos (IPS)** es un dispositivo de seguridad de red que monitorea el tráfico de red y/o las actividades de un sistema, en busca de actividad maliciosa. Entre sus principales funciones, se encuentran no sólo la de identificar la actividad maliciosa, sino la de intentar detener esta actividad.
- **Un sistema de detección de intrusiones (IDS)** es un programa de detección de accesos no autorizados a un computador o a una red.
- **Un cortafuegos (firewall)** es una parte de un sistema o una red que está diseñada para bloquear el acceso no autorizado, permitiendo al mismo tiempo comunicaciones autorizadas. Se trata de un dispositivo o conjunto de dispositivos configurados para permitir, limitar, cifrar o descifrar el tráfico entre los diferentes ámbitos sobre la base de un conjunto de normas y otros criterios.
- **Una lista de control de acceso (ACL)** es un concepto de seguridad informática usado para fomentar la separación de privilegios. Es una forma de determinar los permisos de acceso apropiados a un determinado objeto, dependiendo de ciertos aspectos del proceso que hace el pedido.
- **Los bloqueadores de las conexiones remotas** son tecnologías que bloquean el acceso al servidor a clientes que intentan hacer ataques de fuerza bruta.

Funcionalidades

- Soporte para la gestión de reglas de control de acceso a los recursos del servidor de almacenamiento con limitaciones personalizadas a cada cliente.
- Soporte para el monitoreo del comportamiento de la red. Para detectar e informar sobre posibles intrusiones no autorizadas, con lo cual se puede prevenir que se vea afectada la integridad de los

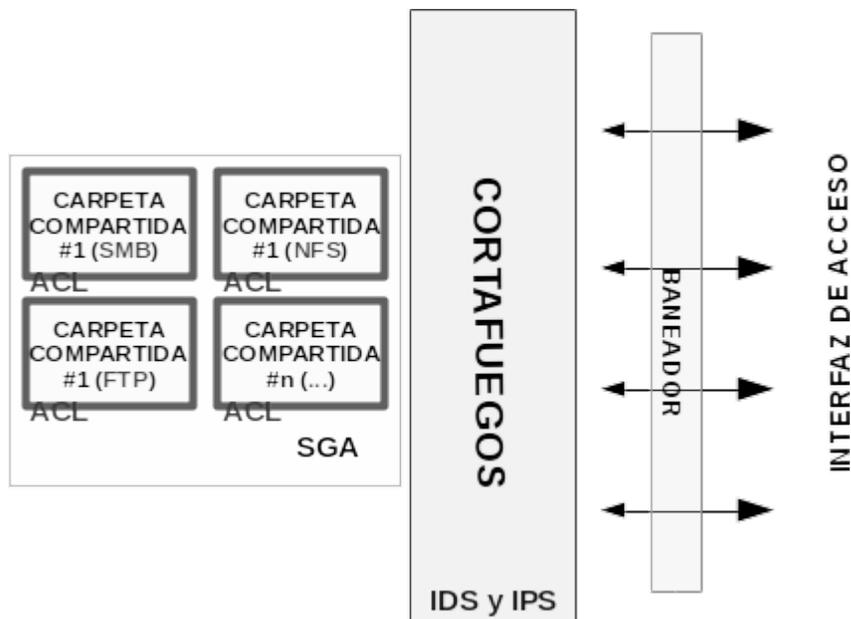
datos almacenados en el servidor.

- Soporte para el control del tráfico de la red que está asociada al servidor de almacenamiento. Cumple con la función de filtrar el tráfico de red entre Internet y el servidor, y puede funcionar de dos maneras diferentes: permitiendo todos los paquetes de red y solo bloqueando algunos considerados sospechosos; o bien denegando todos los paquetes y solo permitiendo aquellos que sean considerados como necesarios.

Despliegue

En la Figura 5 se muestra como se despliega el componente de seguridad de la solución de almacenamiento. El componente crea una zona desmilitarizada para asegurar la integridad de los datos de los ataques externos. El primer elemento de la red es un baneador de IP para interceder ante los intentos de ataques por fuerza bruta en el acceso al servidor. El segundo elemento es un cortafuegos que trabaja con sistema de prevención y detención de intrusos. Dentro del componente de almacenamiento se le aplican reglas de listas de control de acceso a las carpetas compartidas.

Figura 5: Despliegue del componente de seguridad.. Fuente: elaboración propia



En este componente maneja la interoperabilidad a través de las reglas de acceso de los clientes a los recursos compartidos a través de análisis de la conexión. En este análisis se verifica si el cliente tiene acceso al recurso compartido que solicita.

3.1.5 - Componente para la gestión de respaldos (CGR)

Las copias de seguridad es una copia de los datos originales que se realiza con el fin de disponer de un medio para recuperarlos en caso de su pérdida. Las copias de seguridad son útiles ante distintos eventos y usos: recuperar los sistemas informáticos y los datos de una catástrofe informática, natural o ataque; restaurar una pequeña cantidad de archivos que pueden haberse eliminado accidentalmente, corrompido, infectado por un virus informático u otras causas.

Tecnologías, estándares y protocolos aplicados

- **Rsync** es una aplicación que ofrece transmisión eficiente de datos incrementales, que opera también con datos comprimidos y cifrados. Permite sincronizar archivos y directorios entre dos máquinas de una red o entre dos ubicaciones en una misma máquina, minimizando el volumen de datos transferidos.
- **Duplicity** es una herramienta de copia de seguridad cifrada de ancho de banda utilizando el algoritmo Rsync.
- **Cron** es un administrador regular de procesos en segundo plano que ejecuta procesos o comandos en intervalos regulares.

Funcionalidades

- Soporte para copia de seguridad en nube: Compatibilidad para establecer comunicación con La Nube Privada Institucional Nova360 mediante el protocolo FTP.
- Soporte para copia de seguridad en una unidad USB externa con la opción de ejecutar una tarea de copia de seguridad automáticamente al conectar un dispositivo USB en cualquier ranura del servidor mediante la herramienta de sincronización de copias Rsync;
- Soporte para la automatización las etapas de la copia de seguridad, desde la selección de datos de diferentes tipos hasta la programación y encadenamiento de tareas de copia de seguridad planificadas previamente mediante la herramienta de programación de tareas cron y la herramienta de sincronización de copias Rsync.
- Soporte para la encriptación en la copias de seguridad mediante la herramienta Duplicity.

Despliegue

Figura 6: Despliegue del componente de gestión de respaldos.. Fuente: elaboración propia



En la Figura 6 se muestra como se despliega el componente de gestión de respaldos. Los directorios compartidos mediante los diferentes protocolos de compartición de archivos tienen asociado una tarea de copia de seguridad. Mediante la copia de seguridad encriptada la información se puede guardar en dispositivos USB externos, en la Nube Institucional o en otro servidor de almacenamiento presente en la topología de red.

En este componente la interoperabilidad se despliega a través de la compatibilidad en el intercambio de formatos de datos durante las operaciones de copias de seguridad hacia o desde dispositivos externos. También se despliega en el soporte del protocolo SFTP para establecer la comunicación en las conexiones locales y remotas.

3.1.6 - Componente para la gestión de integridad (CGI)

La integridad es la capacidad de garantizar que los datos no han sido modificados desde su creación sin autorización. La información que disponemos es válida y consistente.

Tecnologías, estándares y protocolos aplicados

- **Tecnología de detección de programas maliciosos.** Se refiere a la aplicación de antivirus que tengan como objetivo el escaneo activo de los recursos guardados en el sistema de almacenamiento para la detección y eliminación de programas que puedan atentar con la integridad de los datos.
- **Mecanismos de sumas de verificación** a partir de una función hash¹⁰ que tiene como propósito principal detectar cambios accidentales en una secuencia de datos para proteger la integridad.
- **Encriptación de sistemas** de archivos para contribuir a la confidencialidad de los datos a través de mecanismos de ocultación de la información a clientes no autorizados.

Funcionalidades

- Soporte para la gestión de tareas programas de eliminación y desinfección de programas malignos.
- Soporte para cifrado de carpetas compartidas y volúmenes físicos y virtuales personalizadas para clientes específicos
- Soporte para la programación tareas de chequeo de sumas de verificación a recursos persistentes en el sistema de almacenamiento antes y después de realizar tareas de escritura a los datos.

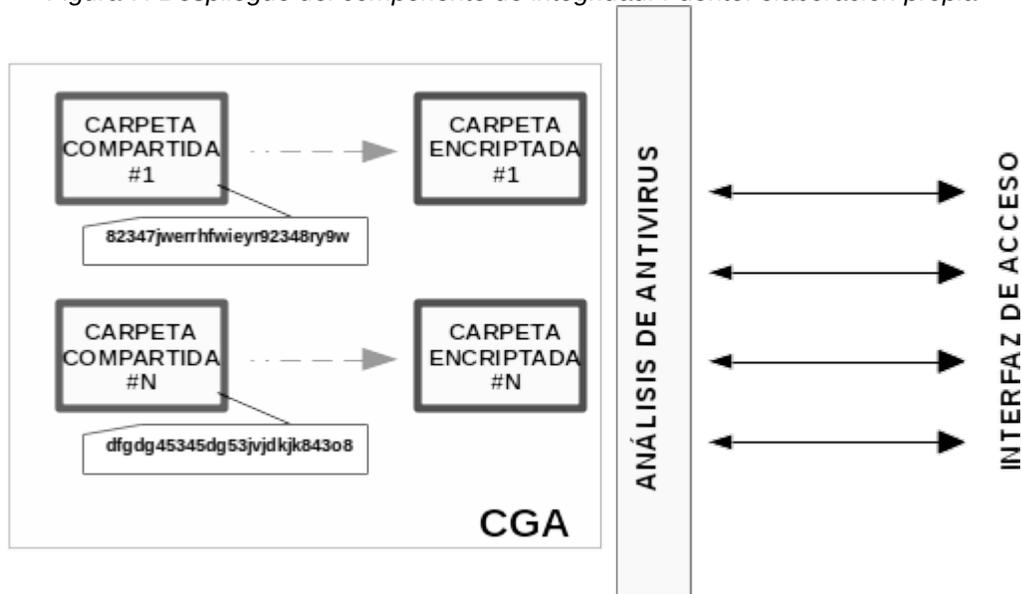
Despliegue

En la Figura 7 se muestra como se despliega el componente de gestión de la integridad. A cada carpeta compartida se le actualiza el *hash* de verificación tras una operación de escritura de sus datos por agentes autorizados. La información que se encuentra en estas carpetas son cifradas y solo pueden ser accedidas una vez que se descifren al montarlas como sistemas de archivos virtuales mediante llaves públicas que sólo las poseen los clientes autorizados. El antivirus mediante tareas programas escanea la información alojada en el sistema de almacenamiento tras operaciones de escritura.

Este componente contribuye a la interoperabilidad a través del resguardo de la integridad de los formatos de intercambio de los propios datos almacenados en el CGA.

¹⁰ Algoritmo matemático que transforma cualquier bloque arbitrario de datos en una nueva serie de caracteres con una longitud fija.

Figura 7: Despliegue del componente de integridad. Fuente: elaboración propia



3.1.7 - Componente para la gestión de monitoreo (CGM)

El monitoreo es el proceso de mantener la vigilancia sobre la existencia y la magnitud del cambio de estado y el flujo de datos en un sistema. Además tiene como objetivo identificar las fallas y ayudar a su eliminación posterior. Las técnicas utilizadas en la supervisión de sistemas de información se cruzan con los campos de procesamiento en tiempo real, estadísticas y análisis de datos. Un conjunto de componentes de software utilizados para la recopilación de datos, su procesamiento y presentación se denomina sistema de supervisión.

La detección rápida de problemas amenazantes es el objetivo más importante del monitoreo y es la función de las alertas en una solución de supervisión. La dificultad consiste en perseguir dos objetivos conflictivos: velocidad y precisión. Conocer si algo no está funcionando adecuadamente y además conocerlo lo más rápido posible. La precisión permite identificar la diferencia si una afectación provoca problemas temporales y transitorios de impacto insignificante. Detrás de cada valor de umbral razonable acecha el riesgo de que problemas potencialmente desastrosos pasen desapercibidos (Ligus 2012).

Tecnologías, estándares y protocolos aplicados

- **Tecnologías para la emisión de alarma.** La alarma es una pieza de configuración que describe una condición indeseable y alertas emitidas en respuesta a ella (Ligus 2013) .
- **Tecnologías para la generación de alertas.** Una alerta es un mensaje de notificación que informa

sobre un cambio de estado, que generalmente indica un posible problema (Ligus 2013).

- **Tecnologías para la medición y monitoreo de los atributos de los dispositivos físicos** de almacenamiento.
- **Estado de los servicios.** Estado de los servicios que implementan los protocolos del componentes de gestión de la disponibilidad.
- **Servicios para el control y seguimiento de los recursos de hardware del servidor.** Entre los recursos críticos para un servidor de almacenamiento se encuentran:
 - Tráfico de la red
 - Consumo de la memoria RAM
 - Consumo de los dispositivos físicos

Funcionalidades

- Control y seguimiento de los recursos de hardware y los servicios
- Gestión de alertas y alarmas.

3.2 - Conclusiones parciales

- La definición de los componentes que conforman la solución de almacenamiento de datos informáticos estructuran de forma organizacional las principales funcionalidades para la gestión del almacenamiento, la seguridad, la integridad, la disponibilidad y el monitoreo de datos informáticos
- Las responsabilidades de cada componente regulan el control sobre los datos almacenados en el sistema de almacenamiento y contribuyen a la interoperabilidad de los clientes con la solución de almacenamiento

Validación de la solución de almacenamiento de datos informáticos

En este capítulo se valida la solución de almacenamiento de datos informáticos a través de opiniones de los expertos en servicios telemáticos, mediante la satisfacción de los administradores de redes y a través de la aplicación de la solución de almacenamiento en la empresa ECOIND para medir los valores de intercambio de formatos de datos y el soporte de los protocolos implementados para validar la interoperabilidad.

4.1 - Valoración de los expertos sobre la solución de almacenamiento

El primer método aplicado fue el criterio de expertos en su variante Delphi (Okoli, D.Pawlowski 2004), con el objetivo de determinar la aplicabilidad de la solución de almacenamiento en entornos de red y de realizar su evaluación. Se define la aplicabilidad como la capacidad de que sea práctico, posible de utilizar y de adaptarse a las características de las empresas. Este es un método que tiene su base en la subjetividad del individuo escrutador, particularmente en su experiencia sobre un tema determinado (Okoli, D.Pawlowski 2004).

Para la selección de los expertos se confeccionó un listado de 12 administradores de redes con varios años de experiencia en la administración de servicios telemáticos y que, las características de los expertos demuestran que poseen los conocimientos para valorar la propuesta. Se tomaron en consideración los siguientes aspectos: título universitario, categoría docente y científica, años de experiencia en la administración de servicios telemáticos, el nivel de dominio sobre el tema que se encuesta y las fuentes de argumentación.

Para la determinación del coeficiente de competencia de los expertos (**K**) se utilizó la fórmula $K = (Kc + Ka) * 0,5$. Donde **Kc** representa el coeficiente de conocimiento que tiene el experto acerca del tema, y se calcula a partir de su propia valoración dentro de una escala del 0 (mínimo conocimiento) al 10 (total conocimiento) multiplicada por 0,1. En el Anexo 3 se muestra la auto-valoración de cada experto y el coeficiente de conocimiento correspondiente. En la Tabla 3 se muestra un resumen de los valores de **Kc** obtenidos.

Tabla 3: Resumen de la ubicación de los expertos según Kc. Elaboración propia.

Coeficiente de conocimiento	1.00	0.9	0.8	0.7	0.6
-----------------------------	------	-----	-----	-----	-----

Cantidad de expertos	1	3	6	1	1
----------------------	---	---	---	---	---

Por su parte **Ka** representa el coeficiente de argumentación o fundamentación de los criterios del experto, resultado de la suma de los puntos dados por cada experto en las fuentes de argumentación definidas en la Tabla 4.

Tabla 4: Grado de influencia de las fuentes de argumentación. Fuente: Moráquez, 2006

No	Fuentes de Argumentación	Alto (A)	Medio (M)	Bajo (B)
1	Estudios teóricos realizados por usted.	0.30	0.20	0.10
2	Experiencia adquirida durante su vida profesional.	0.50	0.40	0.30
3	Conocimiento de investigaciones y/o publicaciones nacionales e internacionales.	0.05	0.04	0.03
4	Conocimiento propio sobre el estado del tema de investigación.	0.05	0.04	0.03
5	Actualización en cursos de posgrado, diplomados, maestrías, doctorado.	0.05	0.04	0.03
6	Intuición	0.05	0.04	0.03

Los valores de **Ka** correspondientes a cada experto se encuentran registrados en el Anexo 4. Finalmente se calcularon los coeficientes de competencias (**K**) según la fórmula antes descrita y se definieron los niveles de competencia según los intervalos que se muestran en la Tabla 5.

Tabla 5: Intervalos para definir la competencia de un experto. Fuente: Moráquez, 2006

Nivel de competencia		
Alto (A)	Medio (M)	Bajo (B)

$1 \geq K \geq 0.8$	$0.8 > K \geq 0.5$	$K < 0.5$
---------------------	--------------------	-----------

En el Anexo 5 se muestran los valores de **Kc**, **Ka**, y **K** para cada experto, donde se refleja que el 83,3% de los expertos obtuvo un alto nivel de competencia y el 16,6% un nivel de competencia medio. Se escogieron los 12 expertos teniendo en cuenta que sus niveles de competencia sobrepasaban el valor de 0,74. De los 12 expertos elegidos uno es extranjero, el 30% posee el título académico de Máster y el 70% son ingenieros informáticos. El promedio de años de experiencia en la administración de servicios telemáticos es de 6 años y el 50% son administradores de redes en instituciones cubanas.

4.1.1 - Resultados de la valoración de los expertos

Para la validación de la solución de almacenamiento por los expertos se diseñó un cuestionario que se muestra en el Anexo 10, el cual les fue enviado en conjunto al material donde se explicaba los componentes que forman parte de la solución de almacenamiento y las tecnologías, protocolos y estándares que se utilizaron para su diseño. Se presentaron los diez aspectos definidos para valorar la solución, cada uno de los cuales cada experto midió con una valoración de Muy adecuado (**MA**), Bastante adecuado (**BA**), Adecuado (**A**), Poco adecuado (**PA**) e Inadecuado (**I**), además se posibilitó dar algún criterio u opinión acerca de la infraestructura de almacenamiento.

Los resultados de la evaluación realizada por los expertos a cada uno de los aspectos propuestos aparecen en el Anexo 6. Sobre la base de la tabla anterior, se determina la frecuencia absoluta por aspectos (ver Anexo 7) y luego la distribución de frecuencia acumulada de cada aspecto (ver Anexo 8). A partir del Anexo 8 , se calcula la distribución de frecuencias relativas acumuladas de cada aspecto (ver Anexo 9).

En la Tabla 6 se realiza el análisis estadístico final donde se calcula: los percentiles de la distribución normal estándar correspondientes a cada una de las frecuencias relativas acumuladas (que se consideran una aproximación de la probabilidad acumulada), la suma algebraica de todos los percentiles anteriores, los puntos de corte (media de los percentiles de cada categoría evaluativa), la suma algebraica de los percentiles dividida por el producto de la cantidad de aspectos sometidos a consulta y la cantidad de categorías evaluativas empleadas(**N**), la media de los percentiles de cada aspecto sometido a consulta (**P**), la diferencia (**N-P**) para cada aspecto analizado (filas). Cada uno de los resultados que se obtienen en la columna (**N-P**) se comparan con los puntos de corte y se determina en qué categoría evaluativa se

encuentra cada aspecto sometido a consulta de los expertos.

Tabla 6: Cálculo de los puntos de corte y escala de los aspectos. Elaboración propia

Aspectos	MA	BA	A	PA	I	Promedio(P)	N-P	Clasificación
A1	0,75	3,50	3,50	3,50	3,50	2,95	-0,1	MA
A2	0,67	3,50	3,50	3,50	3,50	2,93	-0.88	MA
A3	0,67	0,83	3,50	3,50	3,50	2,40	0,45	BA
A4	0,83	0,92	3,50	3,50	3,50	2,45	0,40	BA
A5	0,83	3,50	3,50	3,50	3,50	2,97	-0,12	MA
A6	0,67	3,50	3,50	3,50	3,50	2,93	-0.88	MA
A7	0,92	3,50	0,83	3,50	3,50	2,45	0,40	BA
A8	3,50	3,50	3,50	3,50	3,50	3,50	-0,65	MA
A9	0,83	3,50	3,50	3,50	3,50	2,97	-0,12	MA
A10	0,75	3,50	3,50	3,50	3,50	2,95	-0,1	MA
Puntos de corte (PC)	1,04	2,975	3,23	3,50	3,50	N = 2,85		

Como se puede observar en la Tabla 7 se ubican en una recta numérica todos los puntos de corte para cada una de las categorías y los resultados (N-P) para cada uno de los aspectos sometidos a consulta.

Tabla 7: Rangos y puntos de corte

A1	A2	A5	A6	A8	A9	A10	A3	A4	A7	-	-	-
(-0,1)	(-0.88)	(-0.12)	(-0.88)	(-0.65)	(-0.12)	(-0.1)	(0,45)	(0,40)	(0,40)			

MA (1.04)	MB (2.975)	A (3,23)	PA (3,50)	NA (3,50)
------------------	-------------------	---------------------	----------------------	----------------------

Luego se procedió a determinar el grado de consenso de los expertos (Astigarraga 2006) con los resultados mostrados anteriormente, para lo que se llevaron los resultados a una escala del 1 al 5, donde 5 es Muy adecuado, 4 es Bastante adecuado, 3 es Adecuado, 2 es Poco Adecuado y 1 es Inadecuado. Se confeccionó entonces la matriz correspondiente y se determinó el coeficiente de concordancia (**C**) para cada uno de los aspectos, a partir de la expresión $C=100*(1-Ds/Xm)$ donde **Ds** es la desviación

estándar que se calcula mediante la fórmula $Ds=\sqrt{1/(n-1)\sum_{i=0}^{10}(Xi-Xm)^2}$ y **Xm** equivale a la media

del criterio de los expertos por indicador y se calcula mediante la fórmula $Xm=(\sum_{i=0}^{10}CEi)/10$. Cada coeficiente de concordancia debe tener un valor superior a 75, los valores para cada indicador se muestran en la Tabla 8.

Tabla 8: Coeficiente de concordancia por aspectos. Elaboración propia

Número de experto	Aspectos									
	1	2	3	4	5	6	7	8	9	10
1	5	5	4	5	5	3	5	5	5	5
2	4	5	5	3	5	5	4	5	5	5
3	5	5	5	5	5	5	5	5	4	5
4	4	5	5	5	5	5	5	5	5	4
5	5	3	3	5	5	4	5	5	5	5
6	5	5	5	4	5	5	5	5	5	5
7	5	4	5	5	5	5	5	5	4	5

8	5	5	5	5	5	4	5	5	5	5
9	5	5	4	5	3	5	5	5	5	5
10	4	5	5	5	5	5	5	5	5	4
11	4	5	3	5	4	5	5	5	5	4
12	4	5	5	5	4	5	5	5	5	5
Xm	4,58	4,75	4,5	4,75	4,75	4,75	4,91	5,0	4,83	4,75
Ds	0,52	0,62	0,78	0,62	0,62	0,62	0,28	0,0	0,38	0,62
C = 100 * (1-Ds/Xm)	88,64	86,9	82,7	86,9	85,53	86,9	94,3	100,0	92,13	86,9

A partir de los resultados del análisis de concordancia (ver Tabla 8), se estableció que en la población de los 12 expertos seleccionados, los 10 aspectos evaluados de acuerdo con las puntuaciones de las categorías arriba descritas, alcanzan la condición de Muy adecuado y de Bastante adecuado y se obtuvo un grado de concordancia superior a 82 en todos los aspectos, por lo que se considera que los resultados obtenidos son válidos y fundamentan los criterios dados por los expertos.

Esto corrobora que el coeficiente de concordancia total (**Ct**) de la solución, que se calcula como se muestra en la Tabla 9, obtiene un 100% de concordancia, debido a que no se registraron votos negativos. Se consideran votos negativos aquellos que se refieren a un aspecto con la categoría de Poco adecuado o Inadecuado.

Tabla 9: Coeficiente de concordancia total de la solución de almacenamiento.

Votos negativos	Votos totales	Coeficiente de concordancia total
Vn	Vt	$Ct = (1 - Vn / Vt) * 100$
0	100	100

El consenso entre los expertos respecto a la evaluación de la solución de almacenamiento permite confirmar su aplicabilidad despliegue en un entorno de red. Este proceso de validación permitió perfeccionar y enriquecer la propuesta a partir de los criterios establecidos por los expertos seleccionados. De los resultados obtenidos se puede interpretar que:

- La solución de almacenamiento de datos informáticos a través de los componentes que la estructuran permiten la persistencia de la información en un entorno de red de forma centralizada.
- Los componente que forman parte de la solución de almacenamiento contribuyen a aumentar la seguridad y integridad de los datos.
- La solución de almacenamiento de datos informáticos permite control de acceso a los recursos en la red donde se despliega lo que contribuye a mejorar la interoperabilidad entre el sistema de almacenamiento y los clientes.

4.2 - Satisfacción de los administradores de redes con la solución de almacenamiento

La técnica de satisfacción del cliente se basa en la aplicación de un cuestionario que tiene una estructura interna determinada, que sigue una relación entre tres preguntas cerradas y un análisis posterior de otro conjunto de preguntas abiertas. La relación entre las preguntas cerradas se establece a través del denominado Cuadro Lógico de ladov; el cual posibilita determinar posteriormente el nivel de satisfacción del usuario y del grupo (Pérez 2017). Para aplicar el procedimiento se debe establecer una escala de satisfacción que responde a la siguiente estructura:

(1) Clara satisfacción, (2) Más satisfecho que insatisfecho, (3) No definida, (4) Más insatisfecho que satisfecho, (5) Clara insatisfacción y (6) Contradictoria.

Luego de aplicado el cuestionario a los administradores de redes y haber triangulado las preguntas cerradas, el número resultante de la interrelación de las tres preguntas cerradas indica la posición de cada cual en dicha escala de satisfacción. El resultado final de esta técnica es el índice de satisfacción grupal (ISG) , que refleja el grado de satisfacción de los encuestados. Para ponderar el ISG se establece una escala numérica entre +1 y -1 .

(+1)Máximo de satisfacción, (+0.5) Más satisfecho que insatisfecho, (0) No definido y contradictorio, (-0.5) Más insatisfecho que satisfecho y (-1) Máxima insatisfacción. El cálculo del ISG se realiza mediante la siguiente fórmula:

$$ISG = \frac{A(+1) + B(+0.5) + C(0) + D(-0.5) + E(-1)}{N}$$

Donde:

N: cantidad de usuarios encuestados

A: cantidad de usuarios con Clara satisfacción

B: cantidad de usuarios más satisfechos que insatisfechos

C: cantidad de usuarios no definidos

D: cantidad de usuarios más insatisfechos que satisfechos

E: cantidad de usuarios con Clara insatisfacción

La solución de almacenamiento de datos informáticos tiene como principales beneficiarios los usuarios que se desempeñan como administradores de redes en las instituciones cubanas por lo que se decide aplicar la prueba a este grupo de clientes utilizando el cuadro lógico de ladov del Anexo 11.

Se seleccionaron un total de 20 usuarios para la aplicación del cuestionario entre ellos se encuentran especialistas de migración en servicios telemáticas del centro CESOL de la Universidad de las Ciencias Informáticas y administradores de redes de Instituciones que adoptaron la solución en sus empresas. La preguntas realizadas fueron las siguientes:

1. Considera usted que se deba continuar desplegando el almacenamiento en los servidores de forma independiente sin una solución de almacenamiento de datos informáticos de manera centralizada.
2. ¿Utilizaría la solución de almacenamiento de datos informáticos ?
3. ¿Le satisface los componentes y tecnologías que forman parte de la solución de almacenamiento de datos informáticos?

Luego de calcular el índice de satisfacción grupal (ISG) se obtuvo como resultado de 0,863, pudiéndose constatar que este valor se encuentra en el intervalo de satisfacción, por lo que se puede concluir que la satisfacción de los usuarios que se benefician la solución de almacenamiento es alta.

4.3 - Aplicación de la solución de almacenamiento en la empresa ECOIND

Para medir los indicadores de medición de la interoperabilidad y el correcto funcionamiento de la solución de almacenamiento de datos informáticos se desplegó la solución en un entorno de red donde existen un conjunto de servicios que realizarán operaciones de lectura y escritura de varios tipos de datos a través de diferentes protocolos de comunicación en la red interna de la empresa ECOAIND.

Objetivo:

Medir a través de los indicadores de medición de interoperabilidad descritos en la (ISO/IEC 25023:2017 2016) el nivel de integración de la solución de almacenamiento con los servidores desplegados en la empresa.

En el entorno de despliegue se tuvieron en cuenta los siguientes factores:

1. Tecnologías utilizadas para el despliegue de la solución de almacenamiento.
2. Servicios de la empresa que realizarán operaciones de lectura y escritura en el sistema de almacenamiento.
3. Cantidad de usuarios que utilizarán los servicios desplegados en la empresa, tipo de información y protocolos a utilizar.

Tecnologías utilizadas para el despliegue de la solución de almacenamiento:

Para desplegar la solución de almacenamiento se utilizaron las tecnologías y herramientas descritas en la Metodología Cubana de Migración a Código Abierto las cuales están autorizadas para ejecutar los procesos de migración tanto en estaciones de trabajo como en servidores. Las tecnologías y la distribución por los diferentes componentes están descritos en el Anexo 12.

Servicios de la empresa que realizarán operaciones de lectura y escritura en el sistema de almacenamiento:

Tabla 10: Servicios utilizados en el caso de estudio para medir la interoperabilidad con la solución de almacenamiento.

Tipo de servidor	Descripción del servicio	Descripción de la operaciones de lectura y escritura
Correo	Servicio telématico que	En el servidor de almacenamiento se escribe y se

	implementa los protocolo SMTP y POP3 para la transferencia y descarga de mensajes	lee información cuando cada usuario realiza operaciones en el servidor de correo.
Transferencia de archivos (FTP)	Servicio telématico que implementa los protocolo FTP para la transferencia de archivos de cualquier formato	Operaciones de copiado y eliminaciones de archivos
Servidor Web Apache2	Servicio para el alojamiento de aplicaciones web.	Operaciones de lectura cada vez que un usuario instala un paquete desde el repositorio.
Rsyslog	Servicio para la persistencia de registros logs de las aplicaciones	Operaciones de escritura de archivos en formato logs
Servidor multimedia	Servicio para exportar en la red archivos de tipo audio.	Operaciones de lectura de archivos de audio

Cantidad de usuarios que utilizarán los servicios desplegados en la empresa, tipo de información y protocolos a utilizar:

El despliegue de la solución de almacenamiento tuvo una duración de 48 horas integrada a los servicios. En este tiempo accedieron 64 usuarios a los diferentes servicios desplegados en la empresa y directamente al servidor de almacenamiento. A continuación se listan los protocolos usados para el intercambio de información, cantidad y formato de información almacenada.

Tabla 11: Descripción de la información almacenada durante el caso de estudio.

Tipo de información	Cantidad	Formatos
---------------------	----------	----------

Archivos multimedia	456 GB	AVI, MKV, MP3, MP4, MPG, 3GP, DAT
Archivos de registros	36 MG	LOG
Buzones de correo	120 MG	MBOX
Archivos de texto	142 MB	PDF, ODF, TXT, JSON, XML, PHP, DOC

Protocolos utilizados:

- NFS: Protocolo para la exportación de sistemas de archivos en la red
- FTP: Protocolo para la transferencia de datos en una red
- SMB: Protocolo para la compartición de archivos en la red.
- SSH: Protocolo para la conexión remota segura a terminales.
- DAAP: Protocolo para la compartición de archivos de audio en la red.

Análisis de los resultados:

Para analizar los resultados arrojados por el despliegue del sistema de almacenamiento de datos informáticos se realiza la recolección de datos que se van a utilizar para establecer el calculo de los indicadores de medición de la interoperabilidad.

Medición del indicador Intercambio de formatos de datos

Según la (ISO/IEC 25023:2017 2016) este indicador mide la proporción de los formatos de datos intercambiables que existen en el sistema de almacenamiento y los formatos de datos que los clientes requieran intercambiar con el sistema de almacenamiento. En la recopilación de datos como parte del análisis de los resultados se detectaron 16 formatos de datos que los clientes intercambiaron con el sistema de almacenamiento donde todos fueron soportados durante las operaciones de escritura y lectura. Al aplicar la fórmula para el cálculo de la proporción se define como resultado un valor igual a 1 demostrando que existe compatibilidad en los formatos de datos intercambiables entre el sistema de almacenamiento y los clientes.

Medición del indicador suficiencia del protocolo de intercambio de datos

Según la (ISO/IEC 25023:2017 2016) este indicador mide la proporción en la compatibilidad en el uso protocolos de intercambio de datos que existe entre el sistema de almacenamiento y los cliente como parte del flujo de comunicación entre ambos. En la recopilación de datos como parte del análisis de los resultados se detectaron el uso de 5 protocolos de comunicación donde todos fueron soportados en el flujo de comunicación durante las operaciones de escritura y lectura. Al aplicar la fórmula para el cálculo de la proporción se define como resultado un valor igual a 1 demostrando que el sistema de almacenamiento es compatible con todos lo protocolos de comunicación empleados.

4.4 - Triangulación metodológica de los métodos aplicados

La triangulación metodológica es una técnica para evaluar distintos puntos de referencia y definir una posición. Reduce el sesgo producido en la comparación de resultados obtenidos en la cuantificación de variables mediante un método cuantitativo, las tendencias y dimensiones que surgen de la aplicación de métodos cualitativos (Valencia 2012). El resultado de la aplicación de esta técnica se muestra en la Tabla 12.

Tabla 12: Resultados de la Triangulación metodológica. Fuente: elaboración propia

Objetivo	Métodos cuantitativos	Métodos cualitativos	Conclusiones
Evaluar la capacidad de la solución de almacenamiento de datos informáticos para mejorar la interoperabilidad entre un servidor de almacenamientos y los clientes	<p>Caso de estudio:</p> <p>Se mide la interoperabilidad a través de los indicadores propuestas por la (ISO/IEC 25023:2017 2016)</p> <p>La proporción para el intercambio de formatos de datos es igual a 1.</p>	<p>Indicador: Alto grado de satisfacción:</p> <p>ISG:0.863</p> <p>Criterio de expertos:</p> <p>Se evaluaron 7 aspectos de muy adecuado y 3 de bastante adecuado.</p>	<p>Los resultados obtenidos en los métodos aplicados están en concordancia.</p> <p>Se valida positivamente la capacidad de la solución de almacenamiento para lograr la interoperabilidad entre el servidor de almacenamiento y los</p>

	La proporción en el uso de protocolos para el intercambio es igual a 1.		cliente.
--	---	--	----------

Conclusiones parciales

- La aplicación del criterio de expertos, a través del método Delphi, a expertos en el despliegue servicios telemáticos permitió corroborar la aplicabilidad de la solución de almacenamiento en un entorno de red para el intercambio y almacenamiento de la información
- La aplicación del cuestionario para conocer el grado de satisfacción de los administradores de redes respecto a los componentes y tecnologías que forman parte de la solución de almacenamiento de datos informáticos permitió conocer que los usuarios están satisfechos con la solución
- El despliegue de la solución de almacenamiento en la empresa ECOIND permitió a través del estudio de caso medir la correcta compatibilidad en el intercambio de formatos de datos y el uso de protocolos de comunicación entre el sistema de almacenamiento y los clientes.

Conclusiones

El desarrollo de la presente investigación permitió arribar a las siguientes conclusiones:

- Se define a NAS como arquitectura para el despliegue de la solución de almacenamiento de datos en las instituciones cubanas.
- El análisis de las herramientas de gestión de almacenamiento OpenmediaVault, FreeNAS y Free4NAS y de la Especificación de la Iniciativa de Gestión de Almacenamiento permitieron definir que los elementos a tener en cuenta para el despliegue de la solución del almacenamiento son la gestión de la capacidad de almacenamiento, la seguridad, el acceso a los datos y el monitoreo de los recursos.
- La definición de los componentes que conforman la solución de almacenamiento de datos informáticos estructuran de forma organizacional las principales funcionalidades para la gestión del almacenamiento, la seguridad, la integridad, la disponibilidad y el monitoreo de datos informáticos que regulan el control sobre los datos almacenados en el sistema de almacenamiento.
- La aplicación del criterio de expertos en su variante Delphi permitió validar a la solución de almacenamiento de muy adecuada para ser desplegada en las instituciones cubanas.
- El despliegue de la solución de almacenamiento en la empresa ECOAIND permitió calcular los indicadores de medición de la interoperabilidad entre la solución de almacenamiento donde los resultados arrojaron que existe una proporción adecuada entre los formatos de intercambio y entre los protocolos de comunicación

Recomendaciones

Para dar continuidad a la presente investigación se recomienda:

- Desarrollar una herramienta informática que automatice el procesos de despliegue de la solución de almacenamiento de datos informáticos a través de las tecnologías que complementan los componentes propuestos en la solución

Referencias bibliográficas

- AGRAWAL, Ankit and CHOUDHARY, Alok, 2016. Perspective: Materials informatics and big data: Realization of the “fourth paradigm” of science in materials science. . 2016.
- ASTIGARRAGA, E, 2006. El método delphi. *Techniques*. . 2006. P. 1–14.
- BARRETT, Daniel and SILVERMAN, Richard E., 2001. *The Secure Shell: The Definitive Guide*. ISBN 0-596-00011-1.
- BARTLETT, EJ and FAIRHURST, MJ, 2017. Redundant array of independent disk (RAID) storage recovery. . 2017.
- BARTOK, PD, TURNER, PA and SKEHAN, JD, 2018. System for managing cryptographic keys and trust relationships in a secure shell (SSH) environment. . 2018.
- BARZ, Hans W. and BASSETT, Gregory A., 2015. *Multimedia Networks: Protocols, Design and Applications*.
- CHAN, Boris Kai-Tik and BARIL, Allan, 2017. System and method of secure file sharing using p2p. [online]. 10 August 2017. Available from: <https://patents.google.com/patent/US20170339215A1/en>
- CRETELLA, Giuseppina and ESPOSITO, Antonio, 2015. Cloud Portability and Interoperability. . 19 March 2015.
- DAEHEE KIM, SEJUN SONG and BAEK-YOUNG CHOI, 2015. *Data Deduplication for Data Optimization for Storage and Network Systems*. USA: Springer publishing. ISBN 978-3-319-42280-0.
- DURÁN, Adrián Andrés Jara, 2017. ANÁLISIS, DISEÑO Y PROPUESTA DE UN SERVIDOR NAS CREADO A BASE DE RASPBERRY PI EN LA MAYORISTA DE TURISMO ALLTRAVEL A TRAVÉS DE UN SISTEMA OPERATIVO OPEN SOURCE.
- FREENAS, 2018a. FreeNAS Storage Operating System | Open Source. *Open Source Storage Operating System* [online]. 2018. [Accessed 6 February 2018]. Available from: <http://www.freenas.org/>
- FREENAS, 2018b. ZFS. [online]. 2018. Available from: <https://www.freenas.org/zfs/>
- FUENTES, Allan Pierra, 2011. *Nova, distribución cubana de GNU/LINUX. Reestructuración estratégica de su proceso de desarrollo*. La Habana: Universidad de las Ciencias Informáticas.
- IEEE, 1990. *A Compilation of IEEE Standard Computer Glossaries*. New York.
- ISO/IEC 25023:2017, 2016. *Ingeniería de software y sistemas – requisitos de la calidad y evaluación de software y sistemas (SQuaRE) – medición de la calidad del producto de software y DEL sistema*. 2016. Cuban National Bureau of Standards.
25023

- IXSYSTEMS, 2018. Free NAS Software | NAS4Free. *The Free Network Attached Storage Project* [online]. 2018.
- JOHN WILEY, 2010. *Information Storage and Management: Storing, Managing, and Protecting Digital Information*. EMC Education Services. ISBN 978-0-470-61833-2.
- KO, Y, 2014. System and method for load balancing multiple file transfer protocol (FTP) servers to service FTP connections for a cloud-based service. . 2014.
- LIGUS, Slawek, 2012. *Effective Monitoring and Alerting*. O'Reilly Media, Inc.
- LIGUS, Slawek, 2013. *Effective Monitoring and Alerting*. ISBN 9781449333522.
- LIPINSKI, GJ, 2017. Direct-attached/network-attached storage device. . 2017.
- LONVICK, Ed. C., 2006. The Secure Shell (SSH) Protocol Architecture. [online]. January 2006.
- M. ROUSE, 2017a. ¿Qué es Almacenamiento de archivos? *SearchDataCenter en Español* [online]. 2017. [Accessed 5 February 2018].
- M. ROUSE, 2017b. Almacenamiento de bloques. *SearchDataCenter en Español* [online]. 2017. [Accessed 5 February 2018].
- M. ROUSE, 2017c. ¿Qué es Almacenamiento conectado a la red? *SearchDataCenter en Español* [online]. 2017. [Accessed 5 February 2018].
- MATOTEK, Dennis, TURNBULL, James and LIEVERDINK, Peter, 2017. *Pro Linux System Administration: Learn to Build Systems for Your Business Using Free and Open Source Software*. Second Edition. New York: Springer Science+Business Media New York. ISBN 978-1-4842-2008-5.
- NISHIO, M, 2010. *Network service system, service proxy processing method, computer-readable storage medium storing program, and program therefor*. ISBN 7.805.493.
- OKOLI, Chitu and D.PAWLOWSKI, Suzanne, 2004. The Delphi method as a research tool: an example, design considerations and applications. . December 2004. P. 15–29.
- PÉREZ, José Felipe Ramírez, 2017. Modelo para la gestión y análisis de conocimiento para la selección de equipos de trabajo quirúrgico en sistemas de información en salud mediante técnicas de inteligencia organizacional. [online]. 2017.
- PERURENA, Raydel Montesino, 2012. *MODELO PARA LA GESTIÓN AUTOMATIZADA E INTEGRADA DE CONTROLES DE SEGURIDAD INFORMÁTICA*. La Habana.
- PRADHAN, Gyanendra and CAI, Bingxue, 2015. Centralized storage of storage system resource data using a directory server. [online]. 19 May 2015.
- REINA, David Miguel Poyatos, 2016. *S ERVICIO DE MENSAJERÍA INSTANTÁNEA CON MODERACIÓN*

- Y CONTROL DE TEMÁTICA CON CLIENTE PARA TERMINALES A NDROID. March 2016.
- REYNOLDS, J., 1985. FILE TRANSFER PROTOCOL (FTP). [online]. 1985.
- RUEDA NIETO, MARÍA LUCIA and RODRÍGUEZ REYES, ROSSANA, 2017. Administración y control de GNU/Linux Zentyal Server 5.0 como sistema operativo base para disponer de los servicios y plataformas de infraestructura IT. . 2017.
- SNIA, 2018a. Storage Networking Industry Association. *Concepts* [online]. 2018. [Accessed 2 February 2018].
- SNIA, 2018b. The Storage Management Initiative (SMI) | SNIA. *Advancing Storage and Information Technology* [online]. 2018.
- SUN MICROSYSTEMS, 1989. NFS: Network File System Protocol Specification. [online]. 1989. [Accessed 3 March 2018].
- TECHNET, Microsoft, 2013. *Common Internet File System*.
- TIEN VAN DO and UDO R. KRIEGER, 2008. Performance modeling of an Apache Web server with a dynamic pool of service processes. . 20 June 2008. Vol. 39, p. 1.
- UCI, 2018. Variantes de Nova. *Nova, Distribución Cubana GNU/Linux* [online]. 2018. [Accessed 10 February 2018]. Available from: <http://www.nova.cu>
- V. THEILE, 2018. OpenMediaVault. *The open network attached storage solution* [online]. 6 February 2018. [Accessed 6 February 2018]. Available from: <https://www.openmediavault.org/>
- VADALA, Derek, 2002. *Managing RAID on Linux* [online]. ISBN 978-1-56592-730-8.
- VALENCIA, Maria Mercedes Arias, 2012. La triangulación metodológica. Principios, alcances y limitaciones. [online]. 2012.
- VILLAZÓN, Yasiel Pérez and VILLAZÓN, Yadiel Pérez, 2013. *Módulo para administrar el servicio de correo*. La Habana: UCI.
- VILLAZÓN, Yoandy Perez and VITIER, Abel García, 2015a. Introducción. In: *BUENAS PRÁCTICAS PARA LA MIGRACIÓN A CÓDIGO ABIERTO*. La habana: Ediciones Futuro. p. 2. ISBN 978-959-286-041-4.
- VILLAZÓN, Yoandy Perez and VITIER, Abel García, 2015b. Aplicaciones autorizadas en el proceso de migración. In: *BUENAS PRÁCTICAS PARA LA MIGRACIÓN A CÓDIGO ABIERTO*. La habana: Ediciones Futuro. p. 91–107. ISBN 978-959-286-041-4.

WANG, Xianfu and LU, Xihong, 2014. Flexible Energy□Storage Devices: Design Consideration and Recent Progress. . 10 June 2014.

WANG, Yichuan and KUNG, Leeann, 2018. *Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations.*

XU, Rongbin, ZHAO, Kangkang and ZHANG, Pengfei, 2017. *A Novel Data Set Importance Based Cost-Effective and Computation-Efficient Storage Strategy in the Cloud.* ISBN 978-1-5386-0752-7.

ZADOK, EREZ, 2017. On Incremental File System Delopmen. . 28 August 2017.

Anexos

Anexo 1: Operacionalización de la variable dependiente. Fuente: ISO/IEC 25023:2017

ID	Variable	Descripción	Función de medición
CIn-1-G	Intercambio de formatos de datos	¿Qué proporción de los formatos de datos especificados es intercambiable con otros software o sistemas?	$X = A/B$ <p>A = Número de formatos de datos intercambiables con otros programas o sistemas</p> <p>B = Número de formatos de datos especificados como intercambiables</p>
CIn-2-G	Suficiencia del protocolo de intercambio de datos	¿Qué proporción de los protocolos de intercambio de datos especificados es compatible?	$X = A/B$ <p>A = Número de protocolos de intercambio de datos soportados</p> <p>B = Número de protocolos de intercambio de datos especificados para ser soportados</p>

Anexo 2: Encuesta para determinar el coeficiente de competencias de los expertos. Fuente: elaboración propia.

Encuesta para determinar nivel de competencia de los expertos

Estimado(a) compañero(a).

La presente encuesta forma parte de las acciones para validar la selección de las tecnologías existentes para el despliegue de un sistema de almacenamiento. Su análisis y colaboración en cuanto a los

aspectos que sometemos a su consideración serán de invaluable ayuda para el desarrollo de la investigación. Le solicitamos la mayor responsabilidad y sinceridad en la realización de la encuesta. Se necesita primeramente que evalúe su conocimiento acerca de las herramientas que brindan servicios de compartición de archivos y la gestión de los dispositivos físicos de almacenamiento disponibles en el repositorio de la distribución cubana GNU/Linux Nova, según las indicaciones que se dan a continuación. Le agradecemos de antemano por su valiosa contribución.

Datos generales del encuestado

Título universitario: _____

Categoría científica: _____ Categoría docente: _____

Años de experiencia en el desarrollo de la distribución cubana GNU/Linux Nova: _____

Instrucciones

- Según su criterio, marque con una X en la casilla que caracteriza su nivel de conocimiento sobre los temas referidos. "0" significa total desconocimiento del tema y "10" que tiene pleno conocimiento del mismo.

0	1	2	3	4	5	6	7	8	9	10

- Entre las fuentes que le han posibilitado enriquecer su conocimiento sobre los temas, se someten a consideración algunas de ellas, para que las evalúe en las categorías de: Alto (A), Medio (M) y Bajo (B), colocando una X.

FUENTES ARGUMENTACIÓN	DE	Grado de influencia de cada una de las fuentes		
		Alto (A)	Medio (M)	Bajo (B)
Estudios teóricos realizados por usted.				
Experiencia adquirida durante su vida profesional.				
Conocimiento de investigaciones y/o publicaciones nacionales				

e internacionales.			
Conocimiento propio sobre el estado del tema de investigación.			
Actualización en cursos de posgrado, diplomados, maestrías, doctorado.			
Intuición.			

Anexo 3: Cálculo del coeficiente de conocimiento para los expertos. Elaboración propia

Número de experto	Escala										Kc
	1	2	3	4	5	6	7	8	9	10	
1						x					0.6
2								x			0.8
3								x			0.8
4									x		0.9
5									x		0.9
6							x				0.7
7								x			0.8
8									x		0.9
9										x	1

10								x			0.8
11								x			0.8
12								x			0.8

Anexo 4: Matriz de coeficientes de argumentación por experto. Elaboración propia

Número de experto	Fuente de argumentación						Ka
	1	2	3	4	5	6	
1	0,30	0,40	0,05	0,03	0,05	0,05	0,88
2	0,20	0,50	0,05	0,04	0,04	0,05	0,88
3	0,20	0,50	0,03	0,05	0,04	0,05	0,87
4	0,30	0,50	0,05	0,04	0,05	0,05	0,99
5	0,30	0,50	0,05	0,04	0,05	0,05	0,99
6	0,30	0,50	0,05	0,04	0,05	0,05	0,99
7	0,20	0,40	0,04	0,04	0,05	0,05	0,78
8	0,20	0,50	0,04	0,04	0,05	0,05	0,88
9	0,30	0,40	0,05	0,05	0,05	0,05	0,9
10	0,20	0,50	0,04	0,05	0,04	0,05	0,88
11	0,20	0,50	0,04	0,04	0,04	0,05	0,88
12	0,30	0,40	0,04	0,04	0,04	0,05	0,87

Anexo 5: Niveles de competencia de los expertos. Elaboración propia

Número de experto	Kc	Ka	K	Nivel de competencia
1	0.6	0,88	0,74	Medio
2	0.8	0,88	0,84	Alto
3	0.8	0,87	0,84	Alto
4	0.9	0,99	0,95	Alto
5	0.9	0,99	0,95	Alto
6	0.7	0,99	0,85	Alto
7	0.8	0,78	0,79	Medio
8	0.9	0,88	0,85	Alto
9	1	0.9	0,95	Alto
10	0.8	0,88	0,84	Alto
11	0.8	0,88	0,84	Alto
12	0.8	0,87	0,84	Alto

Anexo 6: Resultados de la evaluación realizada por los expertos a los aspectos propuestos. Elaboración propia

Número de experto	Fuente de argumentación									
	1	2	3	4	5	6	7	8	9	10
1	MA	MA	BA	MA	MA	A	MA	MA	MA	MA

2	BA	MA	MA	A	MA	MA	BA	MA	MA	MA
3	MA	BA	MA							
4	BA	MA	BA							
5	MA	A	A	MA	MA	BA	MA	MA	MA	MA
6	MA	MA	MA	BA	MA	MA	MA	MA	MA	MA
7	MA	MB	MA	MA	MA	MA	MA	MA	BA	MA
8	MA	MA	MA	MA	MA	BA	MA	MA	MA	MA
9	MA	MA	BA	MA	A	MA	MA	MA	MA	MA
10	BA	MA	BA							
11	MA	MB	A	MA	BA	MA	MA	MA	MA	BA
12	MA	MB	MA	MA	MA	BA	MA	MA	MA	MA

Anexo 7: Frecuencia absoluta por aspectos. Elaboración propia

Número de aspecto	MA	BA	A	PA	I	Total
1	9	3	0	0	0	12
2	8	3	1	0	0	12
3	8	2	2	0	0	12
4	10	1	1	0	0	12
5	10	1	1	0	0	12
6	8	3	1	0	0	12

7	11	1	0	0	0	12
8	12	0	0	0	0	12
9	10	2	0	0	0	12
10	9	3	0	0	0	12

Anexo 8: Frecuencia acumulada por aspecto. Elaboración propia

Número de aspecto	MA	BA	A	PA	I
1	9	12	12	12	12
2	8	12	12	12	12
3	8	10	12	12	12
4	10	11	12	12	12
5	10	11	12	12	12
6	8	11	12	12	12
7	11	12	10	12	12
8	12	12	12	12	12
9	10	12	12	12	12
10	9	12	12	12	12

Anexo 9: Frecuencias relativas de las evaluaciones por aspecto. Elaboración propia

Número de aspecto	MA	BA	A	PA	I
1	0,7500	1,0000	1,0000	1,0000	1,0000

2	0,6667	1,0000	1,0000	1,0000	1,0000
3	0,6667	0,8333	1,0000	1,0000	1,0000
4	0,8333	0,9167	1,0000	1,0000	1,0000
5	0,8333	0,9167	1,0000	1,0000	1,0000
6	0,6667	0,9167	1,0000	1,0000	1,0000
7	0,9167	1,0000	0,8333	1,0000	1,0000
8	1,0000	1,0000	1,0000	1,0000	1,0000
9	0,8333	1,0000	1,0000	1,0000	1,0000
10	0,7500	1,0000	1,0000	1,0000	1,0000

Anexo 10: Encuesta a los expertos

Estimado(a) compañero(a).

Con la finalidad de someter a su consideración como experto(a) la solución de almacenamiento de datos informáticos, solicitamos su valoración sobre diferentes aspectos que a continuación se presentan. De antemano, le agradecemos su valiosa contribución.

Para expresar su evaluación, por favor, luego de analizar cuidadosamente el material que se adjunta, evalúe cada uno de los aspectos que se le presentan en la tabla, marcando con una cruz en la casilla correspondiente y teniendo en cuenta para ello el siguiente código de categorías de clasificación.

- **MA:** Muy adecuado.
- **BA:** Bastante adecuado.
- **A:** Adecuado.
- **PA:** Poco adecuado.
- **I:** Inadecuado.

Nº	Aspectos	MA	BA	A	PA	I
1	La solución de almacenamiento de datos informáticos					

	como infraestructura de comunicación entre los clientes y el sistema de almacenamiento, lo valoro como....					
2	Las funcionalidades y tecnologías propuestas en el Componente de Gestión del Almacenamiento (CGA), lo valoro como....					
3	Las funcionalidades y tecnologías propuestas en el Componente de Gestión de la Disponibilidad (CGD), lo valoro como....					
4	Las funcionalidades y tecnologías propuestas en el Componente de Gestión de la Seguridad (CGS), lo valoro como....					
5	Las funcionalidades y tecnologías propuestas en el Componente de Gestión de la Integridad (CGI), lo valoro como....					
6	Las funcionalidades y tecnologías propuestas en el Componente de Gestión de Respaldos (CGR), lo valoro como....					
7	Las funcionalidades y tecnologías propuestas en el Componente de Gestión del Monitoreo (CGM), lo valoro como....					
8	La aplicabilidad de la solución de almacenamiento de datos informáticos en un entorno de red con la interacción con servicios telemáticos la valoro como...					
9	La solución de almacenamiento de datos informáticos como mecanismo para disminuir la duplicidad de la información la valoro como...					

10	La solución de almacenamiento de datos informáticos como mecanismo para centralizar la datos informáticos en una red la valoro como...								
----	--	--	--	--	--	--	--	--	--

Anexo 11: Cuadro Lógico de ladov y preguntas abiertas para los administradores de redes. Fuente: elaboración propia.

¿Le satisface los componentes y tecnologías que forman parte de la solución de almacenamiento de datos informáticos?	Considera usted que se deba continuar desplegando el almacenamiento en los servidores de forma independiente sin una solución de almacenamiento de datos informáticos de manera centralizada.								
	Si			No sé			No		
	¿Utilizaría la solución de almacenamiento de datos informáticos ?								
	Si	No sé	No	Si	No sé	No	Si	No sé	No
Me gusta mucho	1	2	6	2	2	6	6	6	6
Me gusta más de lo que me disgusta	2	2	3	2	3	3	6	3	6
Me dalo mismo	3	3	3	3	3	3	3	3	3
Me disgusta más de lo que me gusta	6	3	6	3	4	4	3	4	4
No me gusta nada	6	6	6	4	4	6	4	5	6
No sé qué decir	2	3	6	3	3	3	6	3	4
Preguntas abiertas:									
1. ¿Qué elementos consideras positivos en la solución de almacenamiento de datos informáticos?									

2. ¿Qué elementos consideras negativos en la solución de almacenamiento de datos informáticos?
3. ¿Propondrías añadir alguna tecnología o componente a la solución de almacenamiento de datos informáticos?

Anexo 12: Aplicaciones utilizadas para el despliegue de la solución de almacenamiento en la empresa Ecoaind.

Aplicación	Descripción	Componente
mdadm	Herramienta para la gestión de RAID mediante software en GNU/Linux Nova.	CGA
lvm2	Herramienta para la gestión de volúmenes lógicos mediante software en GNU/Linux Nova.	CGA
quota	Herramienta para la gestión de cuotas de sistema de archivos en GNU/Linux Nova.	CGA
zfs	Sistema de archivos combinado y un administrador de volumen lógico	CGA
Samba	Herramienta que proporciona servicios de archivo e impresión seguros, estables y rápidos para todos los clientes que utilizan el protocolo SMB/ CIFS	CGD
OpenSSH	Es un conjunto de utilidades de nivel de red relacionadas con la seguridad basadas en el protocolo Secure Shell (SSH)	CGD
nfs-utils	Herramientas de cliente necesarias para usar las capacidades NFS del kernel	CGD
prodFTP	Implementa un servidor FTP.	CGD

iptables	Programa de utilidad de espacio de usuario que permite la administrador del <i>firewall</i> del kernel de Linux.	CGS
Snort	Sistema de detección de intrusión de red.	CGS
fail2ban	Software de prevención de intrusos que protege los servidores de la computadora de los ataques de fuerza bruta.	CGS
savUnix	Programa de detección y eliminación de programas malignos	CGI
eCryptfs	Paquete de software de cifrado de disco para Linux	CGI
Cksum	Comando de GNU/Linux Nova para la implementación de sumas de verificación criptográficamente seguras.	CGI
monit	Es una herramienta gratuita de supervisión de procesos.	CGM
free, top,lscpu,npr oc,dmidecod e	Utilidades de Linux para el seguimiento de los recursos de hardware	CGM
SMART	Tecnología de autosupervisión, análisis e informes utilizada en dispositivos de almacenamiento informático	CGM
rsync	Herramienta para la sincronización de copias de respaldo incrementales.	CGR
duplity	Herramienta para la encriptación de copias de respaldo.	CGR

Glosario de términos

- IDE: Entorno de desarrollo integrado
- SCSI: Interfaz de sistema de computadora pequeña
- TCP: Protocolo de control de transmisión
- IP: Protocolo de Internet
- AFP: Adobe Flash Player
- SMB: Bloque de mensajes del servidor
- FTP: Protocolo de transferencia de archivos
- ZFS: Sistema de archivos Zettabyte
- CPU: Unidad central de procesamiento
- Samba AD : Directorio activo con Samba
- NFS: Sistema de Archivos de Red
- TFTP: Protocolo de transferencia de archivos trivial
- SCP: Protocolo de copia segura
- SSH: Shell segura
- CGS: Componente de gestión de seguridad
- CGI: Componente de gestión de integridad
- CGA: Componente de gestión de almacenamiento
- CGM: Componente de gestión de monitoreo
- CGD: Componente de gestión de disponibilidad
- DAAP: Protocolo de Acceso a Audio Digital
- TI: Tecnologías de la información
- SNIA: Asociación de la Industria de Redes de Almacenamiento
- SMTP: Protocolo simple de transferencia de correo