

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE INGENIERIA Y ARQUITECTURA
ESCUELA DE INGENIERÍA DE SISTEMAS INFORMÁTICOS



**DESARROLLO DE PROTOTIPO DE INFRAESTRUCTURA EN
LA NUBE PARA LA COORPORACIÓN CRISTIANA DE RADIO
Y TELEVISIÓN ELIM**

PRESENTADO POR:
CATIVO ARGUETA, CARLOS CATIVO

OSEGUEDA AGUILAR, CRISTIAN ALEJANDRO

SOSA HERNÁNDEZ, JOSÉ RICARDO

ZELAYA TEJADA, CHRISTIAN ARIEL

PARA OPTAR AL TÍTULO DE:
INGENIERO DE SISTEMAS INFORMÁTICO
CIUDAD UNIVERSITARIA, OCTUBRE DE 2021

UNIVERSIDAD DE EL SALVADOR

RECTOR:

MSC. ROGER ARMANDO ARIAS ALVARADO

SECRETARIO GENERAL:

ING. FRANCISCO ANTONIO ALARCON SANDOVAL

FACULTAD DE INGENIERIA Y ARQUITECTURA

DECANO:

DOCTOR EDGAR ARMANDO PEÑA FIGUEROA

SECRETARIO:

ING. JULIO ALBERTO PORTILLO

ESCUELA DE INGENIERÍA DE SISTEMAS INFORMÁTICOS

DIRECTOR:

ING. ING. RUDY WILFREDO CHICAS VILLEGAS

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE INGENIERIA Y ARQUITECTURA
ESCUELA DE SISTEMAS INFORMÁTICOS

Trabajo de Graduación previo a la opción al Grado de:

INGENIERO DE SISTEMAS INFORMÁTICOS

Título:

**DESARROLLO DE PROTOTIPO DE INFRAESTRUCTURA
EN LA NUBE PARA UNA COORPORACIÓN DE RADIO Y
TELEVISIÓN**

Presentado por:

**CATIVO ARGUETA, CARLOS CATIVO
OSEGUEDA AGUILAR, CRISTIAN ALEJANDRO
SOSA HERNÁNDEZ, JOSÉ RICARDO
ZELAYA TEJADA, CHRISTIAN ARIEL**

Trabajo de Graduación Aprobado por:

Docente Asesor:

MSc. DAMIÁN MORALES

SAN SALVADOR, MES Y AÑO

Trabajo de Graduación Aprobado por:

Docente Asesor:

MSc. DAMIÁN MORALES

Contenido

Introducción.....	i
1. Objetivos.....	2
1.1. Objetivo General.....	2
1.2. Objetivo Específicos	2
2. Justificación	3
3. Definición del Problema.....	4
3.1. Pregunta de Investigación.....	4
3.2. Objetivo de la Investigación	4
3.3. Hipótesis	4
4. Importancia y Resultados Esperados	4
5. Alcances.....	5
6. Limitaciones	5
7. Marco Teórico	6
7.1. Cloud Storage	6
7.2. Escalabilidad.....	6
7.2.1. Escalado Vertical.....	9
7.2.2. Escalado Horizontal.....	10
7.2.3. Autoescalado	11
7.3. Seguridad de Datos en la Nube.....	12
7.3.1. Autenticación de Datos.....	12
7.3.2. Confidencialidad de los Datos.....	14
7.3.3. Criptografía con Umbral.....	15
7.3.4. Blockchain	16
7.4. Matriz Redundante de Disco Independientes (RAID).....	16

7.5.	Cloud NAS.....	20
7.5.1.	Beneficios del Uso de una Cloud NAS	22
7.6.	Respaldo en la Nube	24
7.6.1.	Proveedores de Respaldo en la Nube Pública.	25
7.6.2.	Respaldo en la Nube Privada.....	25
7.7.	Herramientas de Alta Disponibilidad	27
7.7.1.	Herramientas y Técnicas de Alta Disponibilidad.....	28
7.8.	OpenStack	31
7.8.1.	Servicios	32
7.8.2.	Comparación de distintas formas de instalación de OpenStack.....	34
7.9.	Kubernetes	36
7.9.1.	Velocidad.....	37
7.9.2.	Escalabilidad	37
7.9.3.	Abstracción de la infraestructura.....	38
7.9.4.	Eficiencia.....	38
7.10.	Sistemas de Almacenamiento Persistentes.....	39
7.10.1.	CEPH	40
8.	Metodología de Investigación	41
8.1.	Enfoque de la Investigación.....	41
8.1.1.	Enfoque Cualitativo.....	41
8.1.2.	Enfoque Cuantitativo.....	42
8.1.3.	Población y muestra.	42
8.2.	Instrumentos.....	44
8.2.1.	Entrevista semiestructurada.....	44
8.2.2.	Encuesta.....	45

8.3.	Análisis de Resultados	46
8.3.1.	Análisis de datos cualitativos	46
8.3.2.	Análisis Cuantitativo	50
9.	Desarrollo de Prototipo.....	68
9.1.	Análisis y Diseño	68
9.1.1.	Análisis	68
9.2.	Diseño	72
9.2.1.	Requerimientos.....	72
9.2.2.	Topología de Red	74
9.2.3.	Consideraciones previas	76
9.3.	Pruebas.....	77
9.3.1.	Funcionamiento de clúster Ceph	77
9.3.2.	Funcionamiento de Openstack.....	80
10.	Conclusiones.....	83
11.	Recomendaciones	84
12.	Bibliografía.....	85
13.	Anexos	87
	Anexo 1. Encuesta a empleados del área de IT	87
	Anexo 2. Encuesta a empleados del área de Cámara	88
	Anexo 3. Encuesta a empleados del área de Edición	89
	Anexo 4. Entrevista a jefe del área de IT	90
	Anexo 5. Instalación de Ceph.....	93
	Configuraciones De Cada Nodo.....	93
	Depliegue de Cluster ceph.....	94
	Verificar Funcionamiento.....	100

Anexo 6. Comparativo de Tecnologías Compatibles con Ceph.....	102
Ceph.....	102
Gluster	103
LVM	103
iSCSI.....	104
NFS.....	104
Sheepdog	104
ZFS	105
Anexo 7. Comparativo de Soluciones de Infraestructura como Servicio.....	106
Anexo 8. Comparativo de nube pública vs Privada.....	112

Índice de Figuras

Figura 1. Tiempo de respuestas de varios sitios web	8
Figura 2. Diagrama de ejemplo de RAID por Hardware.....	19
Figura 3. Diagrama de ejemplo de RAID por Software	19
Figura 4. Diferencias entre NAS, DAS y SAN	21
Figura 5. Esquema de Arquitectura de Servicio Cloud NAS	22
Figura 6. Diagrama de Diferentes Dispositivos que Acceden a la Nube	26
Figura 7. Esquema de Agregación Automática de Puertos	30
Figura 8. Introducción a OpenStack (2021).	33
Figura 9. Requerimientos Mínimos de Openstack	72
Figura 10. Diagrama de Topología Física	75
Figura 11. Diagrama de Topología Lógica.....	76
Figura 12. Dashboard de Ceph	77
Figura 13. OSDs de Ceph Funcionando	77
Figura 14. RGW funcionando correctamente.....	77
Figura 16. Bucket creado en Ceph usando RGW.....	78
Figura 15. Usuario de Object Gateway con sus credenciales Creadas	78
Figura 17. Listado de usuarios de Object Gateway	78
Figura 18. Prueba de Funcionamiento de Ceph usando AWS CLI.....	79
Figura 19. Descripción Bucket Ceph.....	79
Figura 20. Login de Openstack.....	80
Figura 21. Overview del Servicio de Compute	80
Figura 22. Comprobación de los servicios de Compute Funcionando Correctamente	81
Figura 23. Listado de Instancias en Openstack	81
Figura 24. Listado de Servicios de Neutron Funcionando Correctamente.....	82
Figura 25. Listado de Servicios de Volume Funcionando Correctamente	82
Figura 26. Listado de Servicios de Openstack	82

Índice de Tablas

Tabla 1. Tipos de Redudancia de Datos	18
Tabla 2. Niveles del Sistema y como Lograr Alta Disponibilidad.....	31
Tabla 3. Descripción de los Diferentes Servicios de Openstack	33
Tabla 4. Número de Personas por Área.....	43
Tabla 5. Análisis de Datos Cualitativos.....	47
Tabla . Configuraciones de Hardware de los Nodos del Sistema.....	74
Tabla . Asignaciones de IP de los Nodos	74

Índice de Gráficos

Gráfico 1. Mensajes de Error al Acceder al Sistema (Área de IT).....	51
Gráfico 2. Fallo en el área de trabajo (Área de IT).....	51
Gráfico 3. Tiempo de Respuesta Lento del Sistema (Área de IT).....	52
Gráfico 4. Corrupción de Archivos (Área de IT)	52
Gráfico 5. Navegabilidad del Sistema (Área de IT)	53
Gráfico 6. Consulta a Técnico (Área de IT)	53
Gráfico 7. Accesibilidad del Sistema (Área de IT)	54
Gráfico 8. Rapidez de Recuperación de Credenciales (Área de IT).....	54
Gráfico 9. Administrar y Mantenimiento del Sistema Actual	55
Gráfico 10. Facilidad de Agregar Almacenamiento Extra	55
Gráfico 11. Errores al Acceder al Sistema (Área de Cámara).....	56
Gráfico 12. Fallos en Área de Trabajo (Área de Cámara).....	57
Gráfico 13. Tiempos de Respuesta Lentos del Sistema (Área de Cámara).....	57
Gráfico 14. Corrupción de Archivos (Área de Cámara).....	58
Gráfico 15. Navegabilidad del Sistema (Área de Cámara)	58
Gráfico 16. Consulta a Técnico (Área de Cámara)	59
Gráfico 17. Accesibilidad del Sistema (Área de Cámara).....	59
Gráfico 18. Rapidez de Recuperación de Credenciales (Área de Cámara)	60
Gráfico 19. Tiempo de Subida es Lento	60
Gráfico 20. Forma de Guardar los Archivos	61
Gráfico 21. Error al Acceder al Sistema (Área de Edición).....	62
Gráfico 22. Falla en Área de Trabajo del Sistema (Área de Edición)	62
Gráfico 23. Tiempos de Respuesta Lentos del Sistema (Área de Edición).....	63
Gráfico 24. Corrupción de Archivos en el Sistema (Área de Edición)	63
Gráfico 25. Facilidad de Uso y Navegabilidad del Sistema (Área de Edición)	64
Gráfico 26. Facilidad de Consulta a Técnico (Área de Edición).....	64
Gráfico 27. Accesibilidad de Archivos del Sistema (Área de Edición)	65
Gráfico 28. Recuperación de Credenciales (Área de Edición).....	65
Gráfico 29. Descarga de Archivos del Sistema	66

Gráfico 30. Tiempos de Subida de Archivos..... 66

Introducción

La computación en la nube es una tecnología que se ha popularizado en los últimos años presentando ventajas como la reducción de los costos operativos, la reducción de la complejidad, el acceso inmediato, la facilidad de ampliación y reducción de recursos, sin embargo, el efecto final de la computación en la nube es que los usuarios dejan de ser los dueños reales de sus datos ya que los datos de la organización salen de sus instalaciones parcial o totalmente y se descargan en la nube, siendo el proveedor de la nube quien tenga todo el control de esos datos.

La nube ofrece diferentes servicios tales como “Infraestructura como servicio” (IaaS), “Plataforma como servicio” (PaaS) y “Software como servicio” (SaaS), la nube ofrece un modelo de negocio diferente al tradicional, en el cual se paga por lo que se usa, también ofrece características como escalado automático dependiendo del tráfico de red, la replicación de datos permitiendo una alta disponibilidad, respaldos en caso de fallos y poca latencia para el acceso a los datos, la computación en la nube ofrece la posibilidad de la implementación de una nube pública en donde la empresa es dueña de sus datos y los costes relacionados a la infraestructura, nube privada en donde los servicios y bondades de la nube pertenecen a un proveedor que los gestiona y también aloja a otros clientes, nube híbrida en donde se refiere a cualquier combinación de soluciones de nube privada y pública que funcionen en conjunto en entornos locales y externos, en la actualidad las empresas que ofrecen servicios en la nube pública destacan Amazon Web Services, Google Cloud Platform, Microsoft Azure, Digital Ocean entre otras, las cuales tienen la capacidad de trabajar en conjunto con nubes privadas

1. Objetivos

1.1. Objetivo General

- Desarrollar un prototipo de infraestructura en la nube que cumpla con las necesidades de la Corporación Cristiana de Radio y Televisión ELIM (CCR-TV) para el almacenamiento y administración de copias de seguridad de archivos multimedia (fotos, videos, audio, etc.).

1.2. Objetivo Específicos

- Aplicar una metodología de investigación e instrumentos para recolectar información de carácter cuantitativos y cualitativos, para obtener información previa del modelo de negocio de la organización.
- Analizar la problemática de la CCR-TV para desarrollar una solución que cumpla los requisitos.
- Identificar el tipo de infraestructura con la cual se implementará la solución.

2. Justificación

En la actualidad, la CCR-TV ha asumido el reto de digitalizar todos sus servicios. Este paso a la digitalización ha resultado en varios beneficios, uno de estos es la exposición a más personas al estar conectados por la internet.

No obstante, al expandirse en distintos medios de comunicación ha generado un incremento en la cantidad de archivos multimedia. Estos archivos multimedia a su vez son transformados para los distintos medios de comunicación de la corporación por lo que se generan más archivos. Este crecimiento de la información que se maneja en la corporación ha causado problemas al no tener la capacidad de almacenar y asegurar la integridad de todos los datos.

La corporación a su vez ha presentado interés en una solución que incorpore tecnologías de la nube, como solución actual la corporación ha incorporado servicios de Cloud NAS los cuales han aliviado en cierto grado los problemas presentados y mantiene en operación a la corporación.

Sin embargo, esta solución está presentando huecos al resurgir problemas de falta de espacio para almacenar información, caídas parciales del sistema y problemas de seguridad. Estos problemas nos guían a la presente investigación la cual se enfocará en desarrollar un prototipo que sea basado en la nube que cubra las deficiencias que se presentan actualmente.

3. Definición del Problema

3.1. Pregunta de Investigación

¿En qué medida el desarrollo de un prototipo de almacenamiento de nube para la CCR-TV beneficiaría a la institución para mejorar la disponibilidad, integridad, velocidad de transferencia de información y respaldos de datos?

3.2. Objetivo de la Investigación

- Satisfacer las necesidades de la CCR-TV en el área de almacenamiento de datos por medio de la construcción de un sistema basado en la nube.

3.3. Hipótesis

El uso de dispositivos Cloud NAS como solución de almacenamiento de datos en la organización no sule las necesidades actuales. El desarrollo de un sistema basado en la nube para el almacenamiento de datos mejorará la productividad en el área de trabajos de los empleados que hacen uso de los datos almacenados en el sistema. Dichas mejoras se presentarán en la disponibilidad, integridad de archivos, velocidad de transferencia de información y en el respaldo de datos.

4. Importancia y Resultados Esperados

La realización de este prototipo para la CCR-TV, servirá para resolver en gran manera, los diferentes problemas a los que incurre la institución al almacenar grandes cantidades de archivos multimedia. Y se espera que dichos problemas sean resueltos de forma exitosa.

5. Alcances

- Conocer el modelo de negocios de la organización.
- Desarrollar un prototipo de infraestructura en la nube que cubra las necesidades de la organización.
- Mediante la investigación, tener un mejor panorama de la problemática de la organización.
- Definir un plan estratégico para que el proyecto se realice en tiempo y forma.

6. Limitaciones

- La infraestructura en la nube no permite una solución tan personalizada.
- La migración de información local a la nube puede tomar mucho tiempo debido a la gran cantidad de datos.
- La CCR-TV se encuentra ubicada cerca de un sector de bloqueo de señal telefónica que en cierta manera afecta su conectividad.

7. Marco Teórico

7.1. Cloud Storage

El cómico George Carlin tiene un acto en el que habla de cómo los seres humanos parecen pasarse la vida acumulando "cosas". “Eso es todo lo que necesitas en la vida, un pequeño lugar para tus cosas. Eso es todo lo que es tu casa: un lugar para guardar tus cosas. Si no tuvieras tantas cosas, no necesitarías una casa. Podrías simplemente pasearte todo el tiempo.” (Carlin et al., *Comic Relief (I)* 1986). Si Carlin realizara ese acto hoy, podría hacer la misma observación sobre la información del internet. Ya que todos los que tienen un dispositivo electrónico pasan mucho tiempo adquiriendo y generando datos y luego tratando de encontrar una forma y lugares donde almacenarlos.

Para algunas personas encontrar suficiente espacio para guardar toda su información puede ser realmente difícil. Algunos invierten en grandes discos duros, otros prefieren los USB o la velocidad de los *SSDs*. Pero algunos prefieren el Almacenamiento en la Nube (Cloud Storage en inglés).

Entonces podemos definir almacenamiento en la nube como aquella forma de guardar cualquier tipo de información digital, de forma externa a la organización de forma remota y usualmente manejada por un tercero.

7.2. Escalabilidad

La escalabilidad es la capacidad que tiene un sistema de adaptarse a las necesidades o demanda de sus usuarios con el fin de no afectar a estos o el rendimiento del mismo sistema, por ejemplo,

Consideremos un servidor web que normalmente recibe 500 peticiones por minuto y responde con un tiempo de acceso de 15 ms. Ahora el servidor web recibe 1000 peticiones por minuto y responde con un tiempo de acceso de 500 ms (Fox y Hao, 2018, p.469).

El sistema debería poder mantener el tiempo de acceso o respuesta para cualquier cantidad de peticiones que reciba, en situaciones como esta es donde se ve aprovechada de la escalabilidad, Fox y Hao (2018) afirman que “para que un sistema sea escalable, esperamos que su rendimiento se vea afectado de forma proporcional al cambio en el tamaño o las peticiones” (p.510). Un sistema puede tener las siguientes formas de escalado: escalado constante, escalado lineal y escalado exponencial.

El escalado constante se refiere a un rendimiento del sistema que no cambia a medida que aumenta la carga de trabajo del sistema. En el ejemplo anterior, esperaríamos que el sistema siguiera respondiendo a una velocidad de 15 ms por petición si tuviera un escalado constante.

El escalado lineal significa que el rendimiento del sistema cambia proporcionalmente a los cambios en la carga de trabajo del sistema. Por ejemplo, como la carga de trabajo se ha duplicado, podríamos esperar que el tiempo de respuesta se duplicara hasta los 30 ms. El escalado exponencial se refiere a un rendimiento del sistema que cambia de forma desproporcionada a los cambios en la carga de trabajo del sistema. (Fox et al., 2018, p.496).

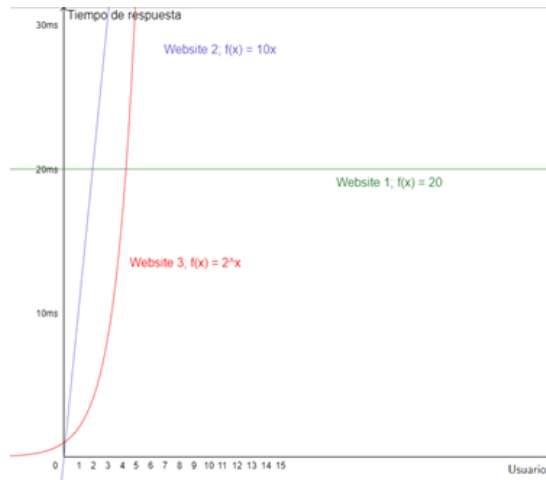


Figura 1. Tiempo de respuestas de varios sitios web

En la **figura 1** se hace una comparativa del tiempo de respuesta de tres sitios web con las etiquetas Website 1, Website 2 y Website 3. Cada sitio web tiene un tipo de escalado diferente, el Website 1 tiene un escalado constante en donde el rendimiento del sistema no se ve afectado por la demanda de usuarios, la respuesta sigue siendo de 20 ms para cualquier número usuarios recurrentes, la función que modela este escalado para este caso es , el Website 2 tiene un escalado lineal, la función que describe este tipo de escalado es , en donde por cada usuario el rendimiento se ve afectado por un factor de 10, es decir, un usuario obtendrá una respuesta de 10 ms, dos usuarios obtendrán una respuesta de 20 ms y así sucesivamente, en este caso cada usuario hace que el tiempo de respuesta aumente 10 ms, Y el Website 3 tiene un escalado exponencial, para este caso es la función que modela este comportamiento, notaremos que los tiempos se ven duplicados cada vez que un nuevo usuario es añadido, es decir que para 1, 2, 3 usuarios los tiempos de respuesta serán 2, 4, 8 ms respectivamente.

Podemos definir En el entorno web, la escalabilidad se refiere a la capacidad del sistema web para adaptarse al aumento de la demanda. La escalabilidad no tiene que ver con la rapidez de un sistema, sino que se centra en la cuestión de cómo funciona el sistema a medida que añadimos más recursos de computación/red/almacenamiento para aumentar su capacidad cuando surgen demandas crecientes (Fox et al., 2018, p.511).

7.2.1. Escalado Vertical

Este tipo de escalado se basa esencialmente en cambiar el sistema actual por un de más potencia y eficacia que el sistema actual, es decir, un hardware más grande y de mayor capacidad que reemplaza a otro de menor prestaciones, por ejemplo, el servidor actual de una empresa podría llegar a no soportar toda la demanda de sus usuarios a causa del crecimiento de la empresa, la empresa, decide realizar un escalado vertical para solucionar su problemática, para eso, compra un servidor con más recursos en hardware que el anterior servidor, si el negocio sigue creciendo, la empresa en un futuro tendrá que comprar un nuevo servidor de más capacidad que el anterior, este tipo de escalado tiene ventajas y desventajas. Fox y Hao (2018) afirman:

El escalado vertical es la forma más fácil de añadir capacidad a un sistema, pero es muy caro porque la organización está literalmente resolviendo el problema invirtiendo más dinero en servidores de gama alta mientras pierde dinero en los servidores comprados anteriormente. También hay un límite máximo en lo que un servidor de gama alta puede proporcionar (pp.470-471).

Esto plantea que el escalado vertical es una manera fácil de implementar, pero llegará un momento en que deba hacer un nuevo escalado vertical, donde se está limitado a los avances tecnológicos de los componentes del sistema, por ejemplo, de hardware, hay componentes hardware que tardan mucho tiempo en ampliar su potencia de forma significativa, además, este tipo de escalado resulta caro.

7.2.2. Escalado Horizontal

Tipo de escalado está basado en conformar un grupo de equipos que dan soporte a una funcionalidad compleja, en la escalabilidad horizontal se añaden equipos para dar más potencia al sistema que se escala, la flexibilidad del escalado es mucho mayor que la escalabilidad vertical, ya que si se necesita más potencia se añade más hardware, si la demanda disminuye el hardware que ya no se necesita puede ser retirado, también puede ser utilizado para añadir fiabilidad, por ejemplo es más fiable un clúster de dos servidores a tener un servidor, en caso de fallas en uno de los servidores, el que está en buen estado puede atender la demanda de los usuarios, o en caso de mantenimiento resulta muy útil, las peticiones se seguirán atendiendo las peticiones, esto claramente con un rendimiento menor hasta que el sistema sea restablecido. El escalado horizontal presenta ventajas y desventajas.

Fox y Hao (2018) afirman:

El escalado horizontal es mucho más viable económicamente. El coste de pasar de un servidor de gama baja a un servidor de gama media puede ser mayor que el de comprar otro servidor de gama baja (y lo mismo para el paso de gama media a gama alta). La investigación ha demostrado que el coste del escalado vertical aumenta exponencialmente y el coste del escalado horizontal aumenta linealmente. Además, como se ha señalado anteriormente, mediante el escalado vertical, estamos

desechando hardware que antes era útil. En el escalado horizontal, todo el hardware adquirido permanece en uso (p.472).

Y como desventaja, Fox y Hao (2018) afirman:

Es mucho más complejo de configurar y ejecutar de forma eficaz que el escalado vertical. Esto se debe a varias razones. En primer lugar, tenemos que asegurarnos de que los servidores de nuestro clúster horizontal son compatibles. Imaginemos que queremos comprar un segundo servidor de gama baja, pero el servidor que estamos utilizando actualmente ha sido descatalogado. Adquirir un servidor de gama baja diferente tendrá sus propios problemas, como tener que utilizar un software de servidor configurado de forma diferente y tener que reescribir los scripts del lado del servidor. Sin embargo, lo que más preocupa es la necesidad de contar con alguna forma de equilibrio de carga para que las peticiones entrantes se envíen a uno de los servidores disponibles (pp.472).

7.2.3. Autoescalado

Esta forma de escalado es donde se realiza la asignación de recursos automáticamente dependiendo de la carga total del sistema, es decir, a mayor carga habrá más recursos a disposición del sistema, si la carga disminuye, esos recursos serán reasignados hasta tener los recursos necesarios para la carga actual. El autoescalado tiene las ventajas de ahorro de electricidad, ahorro en costos, mayor tiempo de actividad y alta disponibilidad, otros tipos de características son las que Fox y Hao (2018) afirman:

La asignación de recursos no es instantánea, de modo que la asignación de recursos se produce una vez que la demanda pide más recursos, y la reasignación se produce sólo

cuando hay recursos ociosos. El autoescalado es, en esencia, un escalado horizontal realizado automáticamente. Puede lograr la mejor utilización de los recursos manteniendo un buen rendimiento del sistema (pp.472-473).

7.3. Seguridad de Datos en la Nube

La seguridad en la nube es uno de los aspectos de mayor preocupación de los usuarios y potenciales usuarios de plataforma en la nube, recordando que al hacer uso de este tipo de solución el usuario deberá externalizar toda o una parte de sus datos, lo que nos lleva a concluir que en este tipo de servicio los usuarios no son físicamente propietarios de sus datos y que los mismos estarán totalmente controlados por el proveedor de servicio de la nube, lo que implica nuevos desafíos y amenazas para la seguridad de los datos en la nube comparándolo con la manera tradicional de almacenar los datos de forma local. Zhang, Xu y Shen (2020) afirman que, desde la perspectiva de los usuarios, la seguridad de los datos en el almacenamiento en la nube incluye principalmente tres partes: fiabilidad de los datos, confidencialidad de los datos e investigación segura de los datos (pp. 5-6). A continuación, se presentan algunas técnicas básicas para lograr seguridad de los datos:

7.3.1. Autenticación de Datos

7.3.1.1. Código de autenticación de mensajes (MAC)

Es una técnica criptográfica utilizada para autenticar un mensaje, el objetivo de MAC es evitar que un tercero, ajeno a la comunicación entre emisor y receptor logre manipular el mensaje enviado, impidiendo que este tercero genere una etiqueta valida e inyecte o altere el mensaje original.

Lo que quiere decir que, la seguridad de MAC requiere que un atacante no pueda generar una etiqueta válida en ningún mensaje "nuevo" cuya etiqueta MAC no se haya generado previamente. Esta seguridad es capturada por una noción de seguridad de "existencialmente infalsificable bajo un ataque adaptativo de mensaje elegido". Aquí, la "infalsificabilidad existencial" se refiere al hecho de que el adversario no debe ser capaz de falsificar una etiqueta válida en ningún mensaje (Zhang, Xu y Shen, 2020, p.12).

Los valores MAC son calculados a partir de una función hash criptográfica con clave secreta K , que solo conocen el remitente y destinatario, pero no los atacantes o terceros ajenos a la conversación emisor receptor.

7.3.1.2. Función Hash

Es una función que tiene como entrada un conjunto de elementos, normalmente suelen ser cadenas de caracteres generando como salida otra cadena de caracteres de tamaño finito. Una función hash se conforma de dos algoritmos.

Lo cual KeyGen y h son:

- KeyGen es un algoritmo de generación de claves, que toma como entrada un parámetro de seguridad y produce una clave s .
- h toma como entrada s y una cadena $x \in \{0,1\}^*$ y da como salida una cadena $hs(x) \in \{0,1\}^f$, donde f es una función que mapea bits a algunos bits dependiendo de la configuración, y $hs(x)$ se refiere al hecho de que s está inherentemente incrustado en la función (Zhang et al., 2020, p.14).

La idea básica de un valor hash es que sirva como una representación compacta de la cadena de la entrada cumpliendo un principal requisito siendo este la resistencia a las colisiones, una colisión es una situación donde una función hash para dos o más entradas distintas producen la misma salida.

Informalmente, una función hash h es resistente a las colisiones si es computacionalmente inviable para cualquier adversario encontrar una colisión en h (Zhang et al., 2020, p.12).

7.3.2. Confidencialidad de los Datos

7.3.2.1. Cifrado de Clave Simétrica

Es una técnica criptográfica que tiene como característica principal usar la misma clave para cifrar y descifrar mensajes entre el emisor y receptor, una vez que estos tiene acceso a dicha clave en común el emisor cifra el mensaje con dicha clave, lo envía al destinatario, y este lo descifra con la misma clave.

El siguiente ejemplo confirma que, dos entidades (por ejemplo, Alice y Bob) comparten una clave y la aprovechan cuando quieren comunicarse en secreto. En concreto, Alice y Bob comparten una clave k para las comunicaciones posteriores. Alice genera primero un mensaje M y cifra M utilizando k para obtener un texto cifrado C . A continuación, Alice envía C a Bob. Al recibir C , Bob lo descifra utilizando k y recupera el mensaje original M (Zhang et al., 2020, p.17).

7.3.2.2. Cifrado de clave pública

Es el método criptográfico en donde el emisor y receptor poseen un par de llaves o claves cada uno, llamadas clave pública y clave privada, a la clave pública puede tener acceso

cualquier persona mientras que la clave privada solo la puede tener su propietario. Cuando el emisor desea enviar un mensaje este utiliza la clave pública del receptor para cifrarla, mensaje que ahora solo podrá ser descifrado por el poseedor de la clave privada que empareje con la clave publica utilizada para cifrar el mensaje.

Podemos corroborar con ejemplo, un emisor (por ejemplo, Bob) y un receptor (por ejemplo, Alice). Alice genera un par de claves (sk, pk) , donde sk se llama clave privada (o clave secreta) y pk se llama clave pública. Al igual que en el caso de la firma digital, dado sk , es fácil calcular pk , pero dado pk , es inviable calcular sk . Alice publica pk y permite que todos los que quieran enviar mensajes conozcan pk . Cuando Bob quiere enviar un mensaje M a Alice, cifra M utilizando pk para obtener el texto cifrado C , y envía C a Alice. Al recibir C , Alice es capaz de descifrarlo utilizando sk para obtener M . (Zhang et al., 2020, p.18).

7.3.3. Criptografía con Umbral

Esta criptografía busca distribuir un secreto entre un cierto número de participantes, de manera tal que solo teniendo al menos el número de umbral se podrá descifrar el secreto, es decir que, ningún conjunto de los participantes menor al umbral puede descifrar el secreto.

Considera un problema de este tipo: un distribuidor tiene un secreto s y quiere compartirlo entre un conjunto de n usuarios U_1, U_2 y U_n proporcionando a cada uno de ellos una parte secreta; también desea que cualquier t de ellos pueda poner en común sus partes secretas y reconstruir s , pero la agrupación de menos de t usuarios no obtendría nada sobre s a partir de sus partes colectivas. Un esquema de compartición de secretos que satisface el

requisito anterior se denomina esquema de compartición de secretos de umbral (t, n) (Zhang et al., 2020, p.18).

7.3.4. Blockchain

Es un tipo de estructura de datos que colecciona elementos llamados bloques, los cuales almacenan datos, uno de los datos que almacena es la información del bloque anterior al actual, consiguiendo así una “cadena de bloques”, donde solo el primer bloque llamado “bloque génesis” no posee la información del bloque anterior ya que este es el origen de la cadena. La información que contiene un bloque solo puede ser editada o repudiada modificando los bloques posteriores. Otro tipo de información que contienen un bloque es, “un puntero de hash como enlace a un bloque anterior, una marca de tiempo y datos de la transacción” (Zhang et al., 2020, p.20).

En la práctica ha permitido, gracias a la criptografía asimétrica y las funciones hash, la implementación de un registro contable (en inglés ledger) distribuido que permite soportar y garantizar la seguridad de dinero digital. Normalmente en las cadenas de bloque se almacenan transacciones, pero podría almacenarse otro tipo de información, en este registro contable se añaden las transacciones y tienen que ser avaladas por la mayoría de los participantes.

7.4. Matriz Redundante de Disco Independientes (RAID)

Del inglés Redundant Array of Independent Disks (RAID) es un mecanismo para asegurar la disponibilidad del almacenamiento en donde se hace uso de múltiples unidades de disco duros o unidades de estado sólido en los cuales los datos se replican. Existen diversas formas de aplicar RAID, estas configuraciones suelen llamarse niveles que poseen diferentes formas

de redundancia y diferentes grados de división de datos entre superficies, los niveles RAID se denominan por número, pero este número no tiene relación alguna con el anterior o el siguiente, más adelante se explican los siete niveles estándares, hay muchos otros niveles, en los que están clasificados como anidados, híbridos y propietarios.

Entre los beneficios que brinda RAID están mayor integridad, tolerancia frente a fallos, tasa de transferencia y capacidad. RAID proporciona una mayor tasa de transferencia de datos, de dos formas que Fox y Hao (2018) afirman:

1. Si un bloque de disco se reparte entre varias superficies, se puede acceder a cada una de ellas simultáneamente, de modo que el acceso al bloque tarda menos tiempo. Por ejemplo, si un bloque se divide en cuatro partes y se coloca en cuatro superficies, el bloque se puede leer o escribir aproximadamente cuatro veces más rápido.
2. Si se dispone de varias unidades independientes, es posible que se produzcan dos o más accesos diferentes simultáneamente porque un bloque se encuentre en una unidad y otro en otra. Como las unidades de disco pueden trabajar de forma independiente, los dos accesos a los bloques pueden producirse simultáneamente en lugar de secuencialmente (pp.462).
3. El rendimiento se ve mejorado ya que, al tener acceso más rápido a un solo bloque, la unidad puede realizar una petición y pasar a la siguiente mucho más rápido, lo mismo ocurre al atender peticiones en simultaneo, donde las peticiones pasan menos tiempo de espera.

La disponibilidad mejora con la redundancia de datos, por ejemplo, al presentarse pequeños fallos en algún sector de un disco duro, la información aún podría ser accesible si este dispositivo posee redundancia, una forma de proporciona redundancia es duplicar completamente una unidad de disco en otra, es decir, una unidad vendría siendo un espejo exacto de la otra, este tipo de redundancia proporcionaría una mayor disponibilidad para

cualquier sector, superficie o unidad de disco o conjunto al que se le presente algún tipo de falla, ya que la unidad duplicada seguiría siendo accesible, el único inconveniente es que se debe utilizar el 50% del almacenamiento, aunque hoy en día las unidades de almacenamiento tiene un costo relativamente bajo.

Tabla 1. Tipos de Redundancia de Datos

Nivel	Tipo de redundancia	Nivel de banda	Descripción
0	Ninguno	Algún número de bytes	No hay redundancia, sólo se utiliza para mejorar la velocidad
1	Espejo completo	Algún número de bytes	Redundancia total, más cara
2	Códigos de Hamming	Nivel de bits	Los códigos Hamming consumen demasiado tiempo para calcular, no se utilizan
3	Bits de paridad	Nivel de bits	Proporciona el acceso más rápido a un solo bloque
4	Bytes de paridad en un disco	Algún número de bytes	Un solo disco almacena toda la paridad y por lo tanto es un cuello de botella de promover los accesos paralelos al disco
5	Bytes de paridad distribuidos en los discos	Algún número de bytes	Dado que los bytes de paridad se distribuyen entre los discos hay un mayor potencial para los accesos paralelos
6	Los bytes de paridad se duplican y se distribuyen entre los discos	Algún número de bytes	Con la redundancia duplicada, esto es más caro que el RAID 5, pero ofrece una mayor redundancia y un mayor potencial de accesos paralelos.

Existen dos tipos de formar de implementar RAID: RAID por hardware y RAID por software.

RAID por hardware es la que utilizar un controlador especial que realizar la gestión de los dispositivos de almacenamiento y presenta el acceso al host, este controlador tiene un procesador como una memoria, de esta forma puede gestionar las operaciones RAID independientemente del CPU de host.

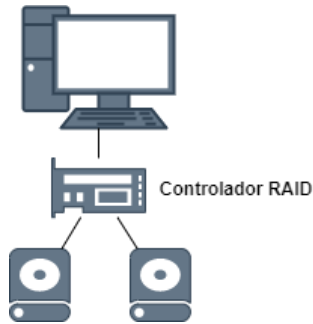


Figura 2. Diagrama de ejemplo de RAID por Hardware

RAID por software: las funciones RAID son proporcionadas por un software que se implementa a nivel de sistema operativo del host, el cual hace uso de la CPU del host para funcionar sin ningún hardware dedicado para esta función.

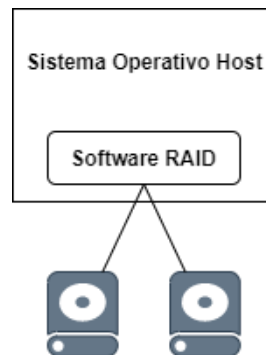


Figura 3. Diagrama de ejemplo de RAID por Software

Fox y Hao (2018) concluyen que:

La implementación del RAID por hardware tiene un mejor rendimiento que el RAID por software. Sin embargo, el coste del RAID por software es más barato que el RAID por hardware porque está integrado en el sistema operativo. Linux, por ejemplo, viene

con el programa mdadm (administración de múltiples dispositivos). A través de mdadm, puedes crear, eliminar y supervisar el RAID por software (pp.464).

7.5. Cloud NAS

Uno de los crecientes retos del internet es la seguridad e integridad de los datos almacenados. Garth y Rodney (2000) refieren que “La mayoría de organizaciones confía en sus entornos distribuidos, los cuales consisten de estaciones de trabajo y sistemas de archivos compartidos”. Generalmente estos sistemas son ejecutados dentro de un servidor de almacenamiento el cual es administrado por un usuario con suficientes privilegios.

Almacenamiento conectado en red, Network Attached Storage (NAS) se refiere a la tecnología que permite compartir la capacidad de almacenar datos de una maquina servidor con otras computadoras clientes, esto a través de una red utilizando un sistema operativo optimizado para su uso (Garth y Rodney, 2000). Los sistemas NAS cuentan con más dispositivos de almacenamiento para poder aumentar su capacidad y por lo general se encuentran dispuestos en arreglos de discos para poder generar redundancia de datos y evitar fallas o perdidas de información.

Aparte de NAS se tienen otros tipos de tecnologías que ofrecen una solución parecida, pero con rasgos diferentes hablando de costos y tipos de conexión, Direct Attached Storage (DAS) es la cual se conecta mediante interfaces (SCSI) o la conexión Storage Área Network (SAN) mediante fibra óptica, en todos los casos estas conexiones utilizan tarjetas de conexión dedicadas. Garth y Rodney (2000) afirman que en el modelo NAS, los clientes realizan peticiones de datos a los sistemas de archivos de manera remota mediante protocolos, sin

embargo, DAS y NAS realizan las peticiones directamente al sistema de archivos. Como se puede ver en la figura 4.

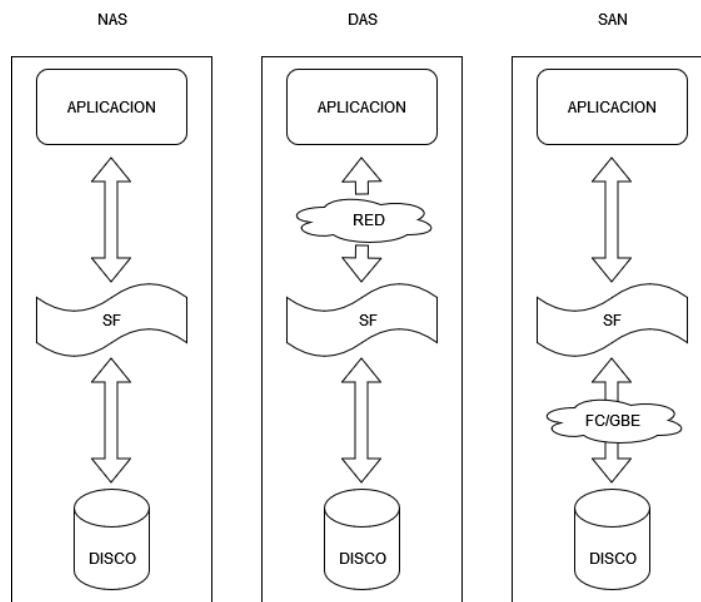


Figura 4. Diferencias entre NAS, DAS y SAN

Una de las principales características de una NAS es el acceso a los datos como si fueran locales, dejando detrás de escena los procedimientos complicados para la configuración que conlleva un acceso seguro a la información.

El servicio de Cloud NAS se desarrolla como cualquier otro servicio de almacenamiento local en red, la diferencia es que generalmente un tercero proporciona una dirección IP a los clientes, esta dirección contiene un conjunto de aplicaciones especializadas en manejar de manera privada el almacenamiento en la nube, evitando así implementaciones y mantenimientos complejos (Garth y Rodney, 2000).

Estos terceros pueden ser revendedores del servicio Cloud, lo cual implica agregar costos para utilizar sus servicios. Se presenta un esquema de la arquitectura en la siguiente figura.

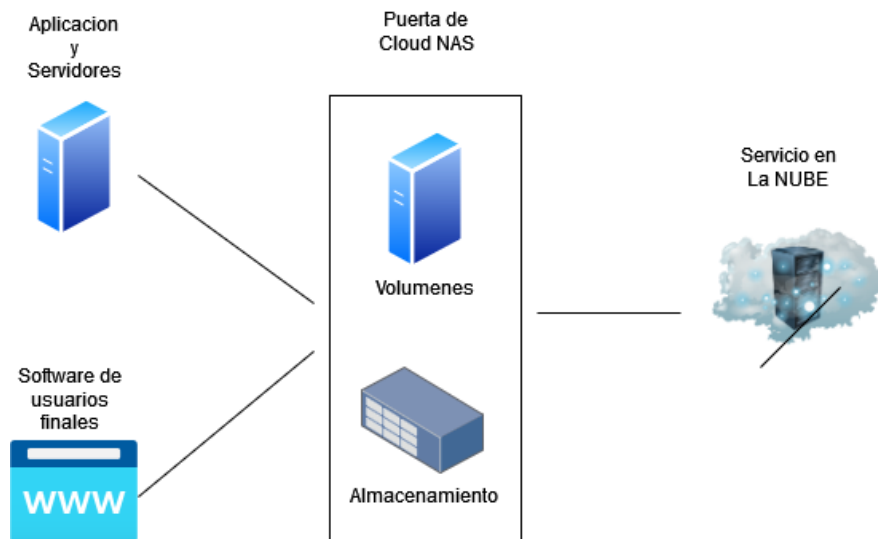


Figura 5. Esquema de Arquitectura de Servicio Cloud NAS

7.5.1. Beneficios del Uso de una Cloud NAS

7.5.1.1. Flexibilidad

Una gran cantidad de organizaciones decide adoptar el modelo cloud ya que incrementa la flexibilidad del negocio. “Le permite a cada usuario ser más flexible dentro y fuera de su espacio de trabajo. Empleados que se encuentran en viajes de negocios pueden acceder a la información siempre y cuando tenga una conexión a internet mediante cualquier dispositivo” (Xue, C. y Xin, F, 2016, p.5). Los dueños de negocios no necesitan estar presentes en una ubicación para poder tener acceso a la misma información que sus empleados.

7.5.1.2. Reducción de Costos

De acuerdo con Xue, C. y Xin (2016) la principal razón por la que las organizaciones adoptan el almacenamiento en la nube es porque les permite reducir costos. Estas organizaciones son atraídas por la posibilidad de reducir costos en sus operaciones ya que los proveedores de servicios cloud les brindan todas las soluciones necesarias. Aparte de esto las organizaciones pueden pagar solamente por lo que utilizan, es decir que solamente pagan por los servicios usados mientras acceden a la nube.

7.5.1.3. Agilidad

“La manera para hacer que una organización se diferencie de la otra es la habilidad y tiempo usado para responder a los cambios en las necesidades de sus clientes” (Xue, C. y Xin, 2016, p.7), Debido a la disponibilidad del internet, la computación en la nube se mantiene a la vanguardia, esto permite que las organizaciones puedan hacer uso de la plataforma en un corto tiempo, esto se convierte en una herramienta competitiva.

Además, incrementa la habilidad de poder adaptarse rápido a los cambios en el ambiente de negocios. Xue, C. y Xin (2016) “Un servicio en la nube tomaría poco proceder y provisionar estos cambios, mientras que un servicio normal podría tomar hasta días.”

7.5.1.4. Escalabilidad

Otro de los beneficios de la computación en la nube es que permiten ajustar los recursos basados en los cambios en su negocio. Xue, C. y Xin (2016) aseguran que “Esto se logra expandiendo la infraestructura de cómputo mediante interfaces de usuario amigables a diferencia de las soluciones tradicionales de IT que no permiten la escalabilidad y presentan

inconvenientes a las organizaciones”. Los servicios en la nube brindan grandes ventajas para las organizaciones especialmente para aquellas pequeñas porque pueden expandir sus recursos cuando sea necesario.

7.6. Respaldo en la Nube

Últimamente el tema de la computación en la nube se ha vuelto muy popular y seguirá en aumento. Esto supone un enorme reto para los administradores informáticos considerando la nube para poder realizar respaldos (*Cloud Backups*).

Cloud Backup

El significado de la computación en la nube puede tener muchas definiciones y esto depende totalmente de quien esté usándolo.

Para usuarios individuales o administradores de organizaciones pequeñas, la computación en la nube, incluyendo el respaldo de datos en la nube, toma lugar sobre una nube pública, no obstante, para organizaciones de tamaños considerables se hace uso de nubes privadas, de acuerdo a los expertos una de las desventajas de usar nubes públicas son los tiempos largos que puede tomar la recuperación de los datos (Harold, Tipton y Krause, 2012, p.380).

Las diferencias entre las nubes públicas y privadas se desvanecen rápidamente por lo tanto es difícil hacer declaraciones precisas y rápidas concernientes a las ventajas y desventajas que presentan los respaldos en la nube.

7.6.1. Proveedores de Respaldo en la Nube Pública.

Segun (Harold, Tipton y Krause, 2012, p.380) los usuarios de nubes públicas en su mayoría usuarios individuales y organizaciones pequeñas que utilizan servicios en la nube mediante terceros, actualmente mediante servidores de nube públicos son capaces de recolectar, comprimir y encriptar datos que es transferida usualmente cada 24 horas o menos mediante internet. Los servicios públicos de respaldo en la nube evolucionan de manera rápida, incluyendo un conjunto de servicios como el hacinamiento de almacenamiento, gastos en el espacio de almacenamiento y ancho de banda en tiempos de recuperación y restauración. Es necesario recalcar que este almacenamiento es visto de diferentes puntos dependiendo del rubro de la industria.

7.6.2. Respaldo en la Nube Privada.

Como lo dice su nombre este tipo de respaldo es privado o en términos simples es de propiedad de la organización que lo utiliza o un consorcio de estas. (Harold, Tipton y Krause, 2012, p.380) afirman que algunas organizaciones de tamaño considerable han estado usando los respaldos de datos en la nube con fines de resguardo, principalmente para evitar tiempos prolongados en la recuperación de datos. Algunas de estas organizaciones han construido y siguen construyendo nubes privadas para su uso personal.

En cualquiera de los escenarios ya sea una nube pública o privada, la administración corporativa de estrategias requiere de una optimización en el tamaño y tipo de datos que serán almacenados y sincronizados. Esto implica tener un conjunto de protocolos

y procesos almacenados por el usuario para mantener la información organizada y accesible para cuando se necesite (Harold, Tipton y Krause, 2012, p.380).

Las técnicas de computación en la nube incluyen la transmisión real o virtual de copias de datos mediante redes públicas o privadas para ser almacenadas y accesibles como respaldos.

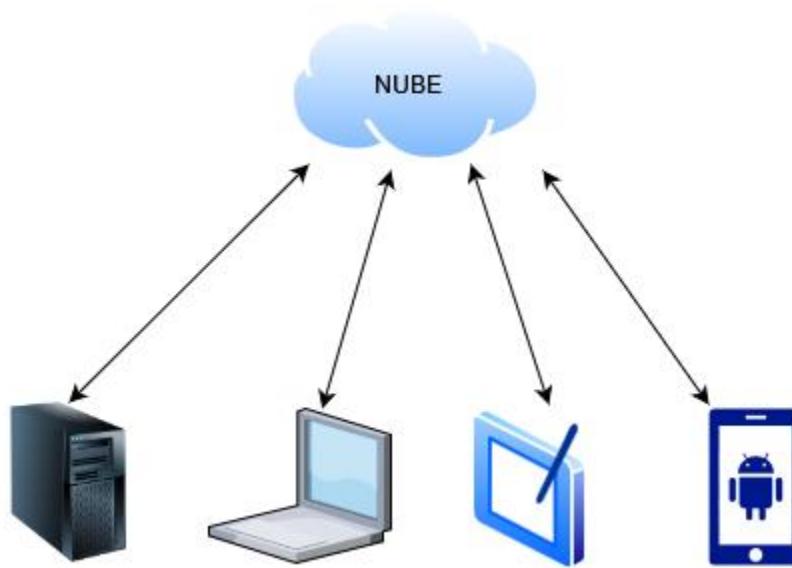


Figura 6. Diagrama de Diferentes Dispositivos que Acceden a la Nube

Este proceso puede ser realizado por la organización misma o puede realizarse mediante la contratación de un tercero. De cualquier manera, se debe tener cuidado al asegurar que la solución cumpla con los requisitos de la integridad de datos, privacidad, disponibilidad, y puntos de vista de la seguridad. Se encuentra muchísima información concerniente a como los administradores deben tomar decisiones sobre cómo utilizar de mejor manera los servicios que la computación en la nube ofrece. (Harold, Tipton y Krause, 2012, p.380)

Dentro de la nube se puede almacenar cualquier tipo de datos, uno de los más recientes métodos de resguardar la información de los usuarios es mediante el respaldo de máquinas

virtuales. Los administradores a menudo utilizan técnicas más avanzadas como los respaldos de máquinas virtuales con fines de recuperación de datos. Una vez completada la recuperación de una máquina virtual esta puede pasar desapercibida para los usuarios, aunque este proceso se aborda de una manera diferente que el almacenamiento de datos. Harold, Tipton y Krause (2012) afirman “En una virtualización el software observa uno o muchos elementos de hardware, no de una manera física sino de una manera virtual” (p.381). Por lo tanto, el contenido de grandes cantidades de dispositivos puede ser almacenado de manera virtual en una menor cantidad que la original.

Agregando, el costo de los respaldos en la nube puede escalar cuando se utilizan respaldos de tipo virtual, los costos pueden incluir la adquisición de múltiples copias de respaldos de software que faciliten el proceso de restauración (Harold, Tipton y Krause, 2012, p.380).

7.7. Herramientas de Alta Disponibilidad

Con el amplio esparcimiento de tecnologías de virtualización en la nube debido a su costo beneficio y eficiencia, el modelo de virtualización se ha adaptado rápidamente por organizaciones y proveedores de servicios como su principal infraestructura y entrega de servicios, ejecutando aplicaciones internas y externas críticas para el modelo de negocios de diferentes consumidores. Para minimizar el tiempo de baja de los servicios debido a errores dentro de las máquinas virtuales, se hace uso de un sistema de **alta disponibilidad**, usualmente incrustado dentro de la infraestructura Hoi y Trieu (2012).

En un ambiente computacional virtualizado, a diferencia del ambiente computacional tradicional en donde los servidores y hardware son estáticos el número de máquinas virtuales que se despliegan puede incrementar o decrecer en un periodo de tiempo

corto, la rápida naturaleza elástica del modelo de computación en la nube presenta un alto número de cambios en las soluciones de alta disponibilidad y también para los diseñadores de infraestructura como la escalabilidad, fiabilidad y el incremento en la complejidad del mantenimiento Hoi y Trieu (2012).

Es evidente que la habilidad de la infraestructura en la nube para brindar soluciones simples, flexibles, escalables y de alta disponibilidad agregan beneficio a proveer servicios en ambientes virtualizados Hoi y Trieu (2012) afirman “Las soluciones de alta disponibilidad son críticas para las aplicaciones en negocios como, compras de divisas y transacciones de bancos” (p.1). Muchas soluciones como la replicación, clúster de conmutación por errores y espejos están disponibles y lo suficientemente explorados por lo tanto muchos revendedores ofrecen la alta disponibilidad como un solo paquete.

7.7.1. Herramientas y Técnicas de Alta Disponibilidad

“La alta disponibilidad es un requerimiento importante de aplicaciones críticas para diversas organizaciones como bancos, esta mantiene activos y accesibles sistemas en casos de fallas de software o hardware que resulta en tiempos de baja perceptibles para los usuarios finales” (Hoi y Trieu,2012,p.2).

La alta disponibilidad de los servicios en la nube es esencial para que los usuarios puedan mantener la funcionalidad, confidencialidad y prevenir costos innecesarios en las operaciones. (Mesbahi *et al* 2018)

Algunas de las herramientas y técnicas más comunes para soluciones de alta disponibilidad son:

1. Espejos (Mirroring)

Un típico sistema de espejos consiste en la configuración de un servidor principal y un servidor espejo (Mirror) que continuamente se comunica con el servidor principal (Hoi y Trieu, 2012, p.2).

2. Replicas (Replication)

Se basa en el modelo de publicar y suscribirse, el servidor primario publica datos y detalles de transacciones que son distribuidas a uno o más servidores suscritos (Hoi y Trieu, 2012, p.2).

3. Clusters de Conmutacion por Error (Fail Over Cluster)

Un grupo independiente de servidores colabora para maximizar la disponibilidad de aplicaciones y servicios. Cada uno de estos servidores en el clúster es un nodo el cual está conectado a otros mediante cables físicos y el clúster en total es administrado mediante un software especial. Si alguno de los nodos en el clúster falla, otro nodo toma su lugar y continúa proveyendo servicios, la interrupción de servicios a los usuarios es mínima con esta herramienta (Hoi y Trieu, 2012, p.2).

4. Captura Instantánea (Snapshot)

Una snapshot congela y preserva el estado de los datos de una máquina virtual en un punto específico del tiempo. Los datos incluyen todos los archivos que constituyen la máquina virtual tales como discos, memoria hasta tarjetas de red. Cuando una snapshot se crea un disco hijo es iniciado con el disco base de su

padre. El disco hijo es de reemplazo o repuesto y utiliza el mecanismo *copy-on-write* (Hoi y Trieu, 2012, p.2).

5. Agregación Automática de Puertos (APA)

Otra herramienta para alcanzar la alta disponibilidad es el uso de la agregación automática de puertos (APA), un producto especial en el ámbito de redes que está disponible en muchos dispositivos de conmutación actuales “Esta tecnología agrega múltiples conexiones Fast Ethernet o Gigabit Ethernet mediante los puertos para asignar una sola IP” (Weygant *et al.*, 2001)). Esto permite la detección de fallas y la recuperación de las mismas, también como balanceador de carga entre enlaces físicos como se observa en la figura 7.

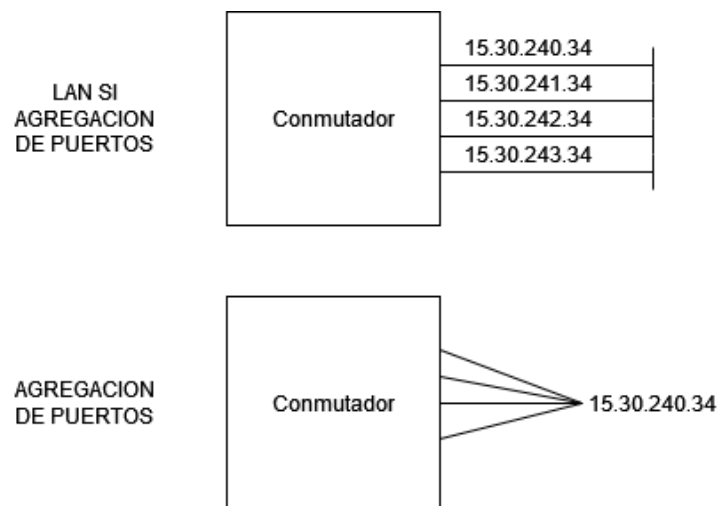


Figura 7. Esquema de Agregación Automática de Puertos

Para lograr una solución de alta disponibilidad es necesario mantener una jerarquía de diferentes niveles de sistemas y herramientas. Uno de los principales retos es que se debe detectar de manera rápida errores en cualquiera de los niveles y de igual manera una solución rápida. Al mismo tiempo el mantenimiento planeado en todos los niveles debe ser lo menos disruptivo posible para el servicio (Weygant *et al.*, 2001).

Tabla 2. Niveles del Sistema y como Lograr Alta Disponibilidad

Nivel del Sistema	Como se Logra la Alta Disponibilidad
Nivel Componentes de Hardware	Componentes redundantes y técnicas de enrutamiento deben ser utilizadas.
Nivel Firmware	La corrección de errores debe ser incorporada.
Nivel Servidores	Las entradas y salidas de la información deben ser redundantes.
Nivel Sistema Operativo	Espejos del software del sistema deben ser implementados.
Nivel Manejo de red	Administración del sistema distribuida y herramientas de monitoreo de red deben ser de alta disponibilidad
Nivel Clúster	La comunicación entre nodos debe ser altamente disponible.
Nivel de base de datos.	La base de datos de ser capaz de inicializarse desde un nodo diferente o debe correr desde varios nodos a la vez.
Nivel de Transacciones de procesos	Las transacciones o servicios de monitoreo deben ser altamente disponibles
Nivel de aplicación	Las aplicaciones o herramientas deben ser robustas y capaces de recuperarse de errores. Estas deben ser capaces de cambiarse automáticamente de procesador y reiniciarse de manera automática sin intervención

7.8. OpenStack

OpenStack es un software de código abierto que permite la implementación de Infraestructure as a Service (IaaS) a través de múltiples servicios que, de manera coordinada,

cumplen diferentes propósitos para lograr el correcto funcionamiento de dicha infraestructura (Encinas, D, 2015).

OpenStack presenta de varias características que la vuelven una atractiva alternativa para el desarrollo del proyecto de en cuestión, entre estas se encuentra:

- **Escalable:** Las soluciones en OpenStack ya se encuentra integradas en compañías reconocidas mundialmente que manejan cantidades exorbitantes de datos como Walmart, Target, American Airlines, Tencent entre otras. La escalabilidad que provee OpenStack en estas compañías en cuanto al manejo de los volúmenes de datos que manejan dan prueba suficiente del cumplimiento de esta característica.
- **Compatible y Flexible:** OpenStack soporta la mayoría de soluciones de virtualización que se encuentran en el mercado: ESX, Hyper-V, KVM, LXC, QEMU, UML, XEN y otras (Takako,P, 2010).
- **Código abierto:** Al ser tecnología de código abierto nos es posible modificar y adaptar los servicios que solamente son necesarios utilizar en cada proyecto.

7.8.1. Servicios

En la página oficial de OpenStack se listan de 6 servicios básicos estables que manejan computación, redes, almacenamiento, identidad e imágenes, mientras que más de una docena de servicios opcionales varían en la madurez del desarrollo (OpenStack, 2021). Esos 6 servicios centrales son la infraestructura que permite que el resto de los proyectos manejen el tablero, la orquestación, el aprovisionamiento completo, la mensajería, los contenedores y la gobernanza.

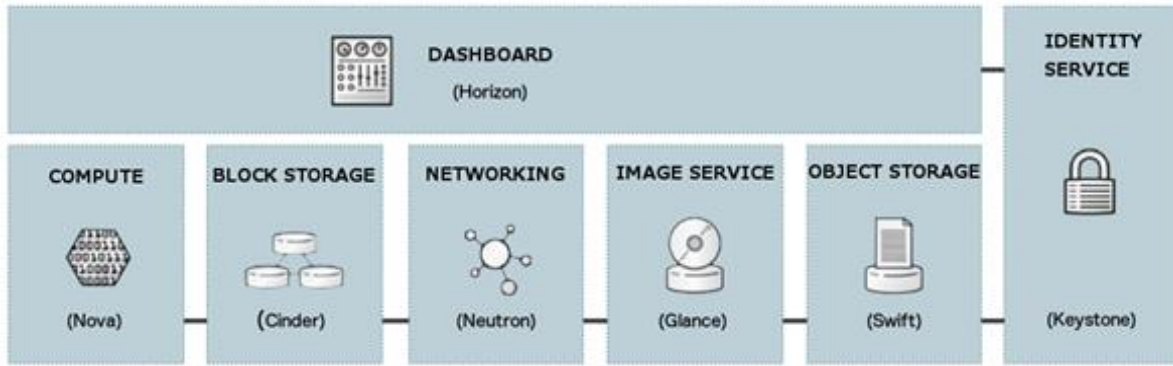


Figura 8. Introducción a OpenStack (2021).

Tabla 3. Descripción de los Diferentes Servicios de Openstack

Servicio	Descripción
Nova	Nova es una herramienta completa de gestión y acceso a los recursos informáticos de OpenStack, manejando la programación, la creación y la eliminación de estos.
Neutron	Neutron conecta las redes a través de otros servicios de OpenStack.
Swift	Swift es un servicio de almacenamiento de objetos altamente tolerante a fallos que almacena y recupera objetos de datos no estructurados mediante una API RESTful.
Cinder	Cinder proporciona un almacenamiento de bloques persistente accesible a través de una API de autoservicio.
Keystone	Keystone autentifica y autoriza todos los servicios de OpenStack. También es el catálogo de puntos finales para todos los servicios.
Glance	Glance almacena y recupera las imágenes de disco de las máquinas virtuales desde diversas ubicaciones.

7.8.2. Comparación de distintas formas de instalación de OpenStack

7.8.2.1. *Devstack*

Devstack es un script de shell de código abierto que puede utilizarse para desplegar OpenStack (Paramjot Singh, 2014). Sin embargo, Devstack soporta la instalación de OpenStack sólo en versiones de Ubuntu, Fedora y RHEL. Puede ser utilizado por los desarrolladores para obtener una configuración fácil de OpenStack.

Los servicios por defecto constituidos por Devstack son (OpenStack, septiembre 2021):

- Identidad (keystone)
- Almacenamiento de Objetos (swift)
- Servicio de imagen (glance)
- Almacenamiento en bloque (cinder)
- Computación (nova)
- Redes (neutron)
- Panel de control (horizon)
- Orquestación (heat)

Devstack sólo soporta el despliegue de un solo nodo para una formación práctica o una visión rápida de OpenStack.

7.8.2.2. *MAAS (Metal As A Service)*

MAAS (Metal as a Service) es un conjunto de software para aprovisionar los servidores bare-metal en componentes para otros servicios como OpenStack. MAAS trata la máquina del servidor físico de forma similar a las máquinas virtuales en la nube (Catherine Spence, 2014). MAAS ofrece la facilidad de desplegar un número de máquinas juntas y tratarlas en un grupo en lugar de individuales. MAAS proporciona una sencilla interfaz web y de línea de

comandos. Con el uso de la sencilla interfaz web, la máquina física puede ser añadida, desplegada, actualizada, reciclada y gestionada.

Cualquier nodo puede ser destruido y reimplantado cuando sea necesario. Los componentes clave de MAAS (Metal as a Service) incluyen: Controlador de Región, Controlador de Cluster y Nodos (servidores físicos).

- Los nodos son las máquinas físicas gestionadas por MAAS.
- El controlador de región son los nodos utilizados para separar diferentes clústeres de subredes.
- El controlador de clúster se utiliza para aprovisionar servidor DHCP y es responsable de la gestión de sus clusters servidores a través de IPMI.

7.8.2.3. *Microstack*

MicroStack proporciona un despliegue de OpenStack de uno o varios nodos que puede ejecutarse directamente en cualquier estación de trabajo (Ubuntu, 2021).

MicroStack es un OpenStack en un instante, lo que significa que todos los servicios de OpenStack y las bibliotecas de apoyo están empaquetados juntos en un solo paquete que puede ser fácilmente instalado, actualizado o eliminado.

MicroStack incluye todos los componentes clave de OpenStack: Keystone, Nova, Neutron, Glance y Cinder.

7.8.2.4. *OpenStack-Ansible (OSA)*

Ansible permite la automatización de configuraciones a través de líneas de código llamadas playbooks, en un playbook usualmente se agregan cada una de las tareas que se desean ejecutar en los equipos remotos, está basado en lenguaje Python y usa protocolo SSH para conectarse y ejecutar tareas.

OpenStack-Ansible proporciona playbooks y roles de Ansible para el despliegue y la configuración de un entorno OpenStack (OpenStack, 2021).

7.8.2.5. *Kolla-Ansible*

Kolla-Ansible proporciona contenedores listos para producción y herramientas de despliegue para operar infraestructuras OpenStack. En Kolla, cada servicio de OpenStack se encapsula con sus dependencias en un contenedor dedicado (Ronan-Alexandre Cherrueau, 2017).

Las imágenes de los contenedores pueden ser construidas bajo demanda, almacenadas y utilizar durante el despliegue. Kolla cuenta con muchos comportamientos por defecto, permitiendo la creación rápida de prototipos además de permitir la completa personalización de estos.

Las versiones modificadas o limpias de OpenStack pueden ser instaladas, topologías de despliegue pueden ser adaptadas a las necesidades de los usuarios y la configuración de todos los servicios pueden ser cambiados.

7.9. Kubernetes

Kubernetes es un orquestador de código abierto para desplegar aplicaciones en contenedores. Fue desarrollado originalmente por Google, inspirado en una década de experiencia en el despliegue de sistemas escalables y fiables en contenedores a través de APIs orientadas a aplicaciones (Burns, B., Beda, J., & Hightower, K. 2018).

Kubernetes es una arquitectura para sistemas distribuidos que son adecuados para el desarrollo de cualquier tipo de nube. Provee el software necesario para desarrollar de forma correcta el despliegue de sistemas distribuidos fiables y escalables

El Libro "Kubernetes: Up and Running" (2018) nos habla sobre 4 características por la cual una persona se inclinaría a usar Kubernetes en vez de las opciones disponibles en el mercado.

Estas características son:

7.9.1. Velocidad

La velocidad es un aspecto importante durante el desarrollo de software hoy en día.

El tiempo que se toma en realizar un software su trabajo o el tiempo de repuesta cuando se realiza una búsqueda o se realiza una consulta son aspectos de suma importancia para los desarrolladores y usuarios.

Sin embargo, la velocidad no es solo medido en que tan rápido puede correr tareas. Los usuarios no están solamente interesados en la velocidad, ellos están interesados también en un servicio de alta disponibilidad lo cual contenedores y Kubernetes proveen las herramientas necesarias para cubrir esta necesidad.

7.9.2. Escabilidad

Kubernetes facilita el cumplimiento de la escabilidad por el hecho de favorecer el desarrollo de arquitecturas desacopladas. Estas arquitecturas desacopladas separan cada componente por APIs definidas y balanceadores de carga de servicios.

Las APIs proporcionan un regulador entre implementador y consumidor mientras que los balanceadores de carga proporcionan un regulador entre las distintas instancias en ejecución en cada servicio.

El desacoplamiento de los componentes a través de los balanceadores de carga facilita el escalado de los programas que componen el servicio porque permiten el aumento del tamaño del programa sin necesidad de realizar otro tipo de configuración a otras capas del servicio al encontrarse separadas.

El desacoplamiento de los servidores a través de las APIs facilita el escalado de los equipos de desarrollo porque cada equipo puede encargarse de trabajar en un microservicio.

7.9.3. Abstracción de la infraestructura

Cuando desarrolladores construyen aplicaciones en términos de imágenes de contenedores y despliegan en términos de APIs portátiles de Kubernetes, transferir su aplicación entre entornos, o incluso ejecutando en entornos híbridos, es simple.

Construir sobre las abstracciones orientadas a la aplicación de Kubernetes asegura que el esfuerzo en la construcción, el despliegue y la gestión de aplicaciones es realmente portátil a través de una amplia variedad de entornos.

7.9.4. Eficiencia

La eficiencia puede medirse por la relación entre el trabajo útil realizado por una máquina o proceso con respecto a la cantidad total de energía gastada. Cuando se trata de desplegar y gestionar aplicaciones, muchas de las herramientas y procesos disponibles son algo

ineficientes. Cuando se habla de eficiencia, suele ser útil pensar tanto en el coste de funcionamiento de un servidor como en el coste humano necesario para gestionarlo.

El funcionamiento de un servidor conlleva un coste basado en el uso de energía, los requisitos de refrigeración, el espacio del centro de datos y la potencia de cálculo en bruto. Una vez que el servidor se ha montado en el bastidor y se ha encendido, el contador empieza a funcionar literalmente. Cualquier tiempo de CPU ocioso es dinero desperdiciado. Por lo tanto, se convierte en parte de las responsabilidades del administrador del sistema para mantener la utilización en niveles aceptables, lo que requiere una gestión continua. Aquí es donde entran en juego los contenedores y el flujo de trabajo de Kubernetes. Kubernetes proporciona herramientas que automatizan la distribución de aplicaciones en un clúster de máquinas, asegurando niveles de utilización más altos que los posibles con las herramientas tradicionales.

7.10. Sistemas de Almacenamiento Persistentes

Los sistemas de archivos distribuidos funcionan en un clúster de máquinas, cada una de las cuales tiene asignados uno o más roles, como el de monitoreo del estado del clúster, servidor de metadatos y servidor de almacenamiento.

Los servidores de almacenamiento, que constituyen la mayor parte de las máquinas del clúster, reciben solicitudes de E/S a través de la red y las sirven a dispositivos de forma local utilizando el software de backend de almacenamiento. Situado en la ruta de E/S, el backend de almacenamiento desempeña un papel clave en el rendimiento del sistema en general. (A. Aghayev, S. Weil, M. Kuchnik, M. Nelson, 2019).

7.10.1. CEPH

Ceph es un sistema de archivos paralelo basado en objetos cuyo diseño se basa en dos ideas clave (C. Maltzahn, E. Molina-Estolano, A. Khurana, 2010). La primera idea clave es el almacenamiento basado en objetos, que divide la arquitectura tradicional del sistema de archivos en un componente cliente y un componente de almacenamiento. El componente de almacenamiento gestiona localmente la programación y la disposición de los discos. La segunda idea clave en el diseño de Ceph es la separación de datos y metadatos. La gestión de los datos difiere fundamentalmente de la gestión de los metadatos.

Ceph desacopla las operaciones de datos y metadatos eliminando las tablas de asignación de archivos y sustituyéndolas por funciones de generación. Esto permite a Ceph aprovechar la inteligencia presente en los OSD para distribuir la complejidad que rodea el acceso a los datos, la serialización de las actualizaciones, la replicación y la fiabilidad, la detección de fallos y la recuperación. Ceph utiliza una arquitectura de clúster de metadatos distribuidos altamente adaptable que mejora drásticamente la escalabilidad del acceso a los metadatos y, con ello, la escalabilidad de todo el sistema (SA. Weil, SA. Brandt, 2006).

8. Metodología de Investigación

8.1. Enfoque de la Investigación.

Para el diseño metodológico de investigación se hará uso de instrumentos cuantitativos y cualitativos. Con el fin de obtener mayores perspectivas, profundización, innovación, análisis con más seguridad y rigor en el prototipo que se desarrolle.

Todos los fenómenos y problemas que enfrentan actualmente las ciencias son tan complejos y diversos que el uso de un enfoque único, tanto cuantitativo como cualitativo, es insuficiente para lidiar con esta complejidad. (Hernández Sampieri & Mendoza Torres, 2018). Por lo tanto, se plantea de estos dos métodos como el más acertado y como los que se ajustan a las expectativas del grupo interesado en la indagación y sistematización del problema a resolver con el prototipo a desarrollar.

8.1.1. Enfoque Cualitativo

La metodología cualitativa, como indica su propia denominación, tiene como objetivo la descripción de las cualidades de un fenómeno. Busca un concepto que pueda abarcar una parte de la realidad. No se trata de probar o de medir en qué grado una cierta cualidad se encuentra en un cierto acontecimiento dado, sino de descubrir tantas cualidades como sea posible (RM Palacios, 2006).

Dentro del enfoque cualitativo se tiene contemplado hacer el uso de una entrevista a las personas interesadas en la solución con el propósito de obtener información de primera mano que pueden darnos las personas involucradas en el área de TI que se encargan del manejo del almacenamiento de información.

8.1.2. Enfoque Cuantitativo

Para que exista metodología cuantitativa se requiere que entre los elementos del problema de investigación exista una relación cuya naturaleza sea lineal. Es decir, que haya claridad entre los elementos del problema de investigación que conforman el problema, que sea posible definirlo, limitarlos y saber exactamente donde se inicia el problema, en cual dirección va y que tipo de incidencia existe entre sus elementos (RM Palacios, 2006).

Edelmira G. La Rosa (1995) dice que para que exista metodología cuantitativa debe haber claridad entre los elementos de investigación desde donde se inicia hasta donde termina, el abordaje de los datos es estático, se le asigna significado numérico.

Para el método cuantitativo se hará uso de encuestas que permitan obtener datos que nos ayuden cotejar la información obtenida por el método cualitativo, con el propósito de poder identificar las mejores tecnologías a usar, también nos ayudará de mejor forma a diseñar el prototipo y así poder dar una mejor solución a los posibles a los problemas.

8.1.3. Población y muestra.

Es muy importante en un trabajo de investigación poder identificar las diferencias entre población y muestra, para poder así poder llevarla a cabo de la mejor manera y lograr sacar mejores resultados.

En primer lugar, una muestra es entendida como un subconjunto de la población conformado por unidades de análisis (Velasco Rodríguez et al. 2003).

La población es un conjunto de elementos que contienen ciertas características que se pretenden estudiar (Gallego 2004a). Por esa razón, entre la población y la muestra existe un

carácter inductivo (de lo particular a lo general), esperando que la parte observada (en este caso la muestra) sea representativa de la realidad (entiéndase aquí a la población); para de esa forma garantizar las conclusiones extraídas en el estudio (Silva Ayçaguer 2007).

Existen dos niveles de población, el primero, la población diana, que generalmente es muy grande y el investigador no logra tener acceso a ella y el segundo, la población accesible, en donde el número de elementos es menor y está delimitado por criterios de inclusión y exclusión (Gallego 2004b). Es en la última población en la que se realiza el muestreo y tamaño muestral.

En el presente caso de la corporación se hará uso de toda la población que tiene acceso a los servicios de nube, ya que no es demasiado grande para realizar un muestreo que nos devuelva datos de valor.

En la población se han logrado identificar 17 personas que tienen acceso a los servicios de nube de la corporación los cuales se encuentran divididos en 4 áreas. Los cuales son:

Tabla 4. Número de personas por área

Área	# Personas
Área de IT	4
Área de Edición	8
Área de Cámara	4
Gerencia	1

Las funciones de estos grupos son diferentes entre sí, lo cual es importante para poder identificar los diferentes problemas que cada área presenta, las actividades que cada grupo realiza son:

- **Área de IT:** Encargada de administrar y mantener los dispositivos de almacenamiento, así como de ayudar en el guardado y obtención de archivos hacia las otras áreas.
- **Área de edición:** Encargada de hacer uso de los archivos para la edición, depuración y publicación de estos.
- **Área de cámara:** Encargada de grabar los archivos y guardar la información en los dispositivos de almacenamiento.

8.2. Instrumentos.

La entrevista nos permite obtener información de los individuos que están directamente relacionados con la realidad de la problemática. En este caso se realizará una entrevista al jefe del departamento de IT del establecimiento donde se está realizando el proyecto de graduación con el objetivo de obtener de una forma más directa y personal la situación real de la problemática

8.2.1. Entrevista semiestructurada

La entrevista semiestructurada de investigación es un instrumento capaz de adaptarse a las diversas personalidades de cada sujeto, en la cual se trabaja con las palabras del entrevistado y con sus formas de sentir, no siendo una técnica que conduce simplemente a recabar datos acerca de una persona, sino que intenta hacer hablar a ese sujeto, para entenderlo desde dentro (Corbetta,2003).

- **Entrevista al jefe del área de informática**

Interesa conocer la opinión de un alto mando de la empresa sobre la problemática planteada de resguardo y acceso de archivos multimedia en la corporación, siendo el encargado de mantener en función el área de informática nos asegura que nos pueda brindar información relevante sobre la realidad de la corporación.

8.2.2. Encuesta

Se realizará una encuesta a los empleados de las tres áreas siguientes. La encuesta nos permitirá conocer similitudes o diferencias sobre aspectos que pueden influenciar la construcción del prototipo logrando identificar variables (Corbetta, 2006) que se pueden presentar según sea la frecuencia de uso del sistema, el grado de conocimiento de los aspectos técnicos del sistema, la complejidad de las tareas que se realizan y entre otros que pueden surgir durante el desarrollo de la investigación.

- **Encuesta a empleados en el área de IT**

Es el área que tienen un mayor grado de conocimiento del funcionamiento del sistema, al ser el área que se encarga del mantenimiento y velar por el funcionamiento correcto del sistema, y pueden aportar un punto de vista técnico sobre las dificultades que se presentan en la organización y los retos que se tienen que superar para mejorar la situación actual.

- **Encuesta a empleados en el área de Edición**

Es el área que tienen un mayor grado de involucramiento con el uso de los datos que se almacenan en el sistema por lo cual su punto de vista nos dará información importante

sobre individuos que utilizan frecuentemente los datos que se almacenan actualmente en el sistema.

En el desempeño de sus actividades laborales manejan frecuentemente los archivos multimedia para la creación y edición de los videos, audios e imágenes publicitarias de la organización de radio y televisión.

- **Encuesta a empleados en el área de Cámara**

Es la principal fuente de archivos multimedia que se almacenan en el sistema, el punto de vista que proveen nos dará información pertinente sobre la existencia de problemas durante el proceso de almacenamiento y respaldo de datos en el sistema.

8.3. Análisis de Resultados

Para alcanzar los objetivos planteados al inicio de la investigación, se procedió con el tratamiento, análisis e interpretación de la información obtenida a través de la encuesta y entrevista semiestructurada, además se realizaron gráficos para una mejor comprensión visual de los resultados cuantitativos. Para la perspectiva cualitativa se busca analizar la opinión del encargado de IT sobre la eficiencia del sistema que almacena los datos de la corporación. Con respecto a la perspectiva cuantitativa se busca obtener información sobre la experiencia de los usuarios que están directamente relacionada con el sistema de archivo actual.

8.3.1. Análisis de datos cualitativos

Para el desarrollo del proyecto se hizo uso de la entrevista como una de las formas de encarar la realidad. Una entrevista es una conversación estructurada y con un propósito definido. Como una de las herramientas más usadas en la investigación cualitativa, la entrevista busca entender el mundo desde la perspectiva del entrevistado, y desmenuzar los significados de sus experiencias” (Alvarez & Jurgenson, 2003). Para Sampieri, Fernández, & Baptista

(2010), Una entrevista es una reunión para conversar e intercambiar información entre una persona (el entrevistador) y otra (entrevistado) u otras (entrevistados).

Para poder realizar un análisis cualitativo correcto se harán uso de técnicas hermenéuticas y métodos como la teoría fundamentada descrita en (Hernández Samperi, Álvaro Obregón, & Baptista Lucio, 2010).

Primeramente, para la realización del análisis cualitativo se realizó la recolección de datos mediante la entrevista, donde posteriormente se realizó una transcripción de esta, esto para poder realizar una categorización de las respuestas brindadas, e identificar la información que da respuesta concreta a la pregunta realizada. Se identificaron las categorías siguientes:

- Problemas al acceder a contenido multimedia
- Quejas recibidas de parte de los usuarios del almacenamiento
- Datos más importantes que no se cubren de forma adecuada
- Frecuencia de acceso a los datos

Tabla 5. Análisis de datos cualitativos

Tema/ Participante	Encargado de IT
Problemas al acceder a contenido multimedia	Presentamos dos grandes problemas, los que son los medios físicos de comunicación de los servidores como fibra óptica, cables UTP, aún la misma red refiriéndose al wifi y lo otro son los ISP

<p>Quejas recibidas de parte de los usuarios del almacenamiento</p>	<p>Otras que pasan difícilmente es porque el servidor se ha caído o necesita una actualización, eso por parte de radio. Por parte del canal, se divide en editores, cámaras, productores los servidores de ellos manejan suficientes TB de almacenamiento, han existido ciertas quejas, una de las más frecuentes es que se han quedado sin almacenamiento, nosotros contamos con 80 TB de almacenamiento, pero el contenido que ellos suben es en HD, ellos piensan que el almacenamiento es ilimitado</p>
<p>Datos más importantes que no se cubren de forma adecuada</p>	<p>Te podría decir que por nuestro tipo de organización que somos podría hablar de videos y predicaciones del pastor general, que están almacenadas, pero me gustaría que tuvieran algún tipo de encriptación o una biblioteca especialmente como respaldo, que no se sabe lo que puede pasar y tal vez cierto material de documental, refiriéndonos más que todo a la cuestión de seguridad quiero tocar lo que es el acceso, si se pueden crear ciertas restricciones que cierto usuario no acceda a ciertas carpetas</p>
<p>Frecuencia de acceso a los datos</p>	<p>En mi caso como IT puede ser 2 o 3 veces a la semana tenemos una carpeta especial donde almacenamos las ISO de los sistemas operativos, librerías, paqueterías, en el caso de productores o editores es diario, su acceso a los</p>

	<p>servidores de almacenamiento, en la nube de ellos es a diario, aunque ellos editan sus archivos offline pero todos sus archivos están en la nube cuando ellos hacen la post producción o hacen edición final lo suben, igual en el canal las programadoras tienen que estar subiendo contenido para pasarlo en el canal señal al aire lo agarran directamente de los servidores y eso es a diario</p>
--	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Se obtuvo los siguientes resultados para cada categoría:

Problemas al acceder a contenido multimedia

Se determina que usuarios técnicos como el encargado de IT presentan problemas de acceso a su contenido multimedia siendo estos los relacionados a medio físico como cableado UTP, la conectividad a la red inalámbrica (WiFi) y su proveedor de internet (ISP), se descarta algún tipo de inconformidad considerable referente al acceso de los datos que provenga directamente de las nubes que la institución posee actualmente.

Quejas recibidas de parte de los usuarios del almacenamiento

Las principales quejas que recibe el departamento de TI son de usuarios externos a este departamento, entre las razones están la disponibilidad de los servidores y el agotar los recursos de almacenamiento de la nube de video, estos causado por caídas de servidores, falta de actualización de los mismos, esto con respecto a la disponibilidad, con respecto al almacenamiento, este recurso es limitado para la institución, y dado que son usuarios externos

al departamento, son usuarios no técnicos, como los editores de videos que desconocen de este tipo de limitantes, los cuales manifiestan quedarse sin almacenamiento frecuentemente.

Datos más importantes que no se cubren de forma adecuada

La institución produce contenido multimedia a diario, y del que podemos determinar que el sistema actual con el que se resguardan sus datos no posee ningún sistema de identidad que permita otorgar permisos o restringir accesos a los usuarios que entra al sistema, teniendo la posibilidad que cualquier tipo de usuario pueda realizar cambios sobre cualquier archivo de la institución, lo que representa un grave problema de seguridad del sistema actual que resguarda los datos.

Frecuencia de acceso a los datos

Los usuarios que hacen más uso del almacenamiento actual de la institución son productores, editores y personas programadores de contenido para el canal, a los que corresponden tareas relacionadas con archivos de videos, su relación con la nube es constante teniendo un acceso diario, esto es cinco veces por semana, por otra parte, los usuarios técnicos del departamento de IT representan un acceso levemente menor con un máximo de tres veces por semana, lo que representa una alta disponibilidad.

8.3.2. Análisis Cuantitativo

A continuación, se presentan los resultados del análisis e interpretación de los datos obtenidos mediante la encuesta realizada al área de IT durante la primera semana de septiembre

Pregunta 1. ¿Alguna vez se ha encontrado con algún mensaje de error cuando quiere acceder al sistema de almacenamiento?

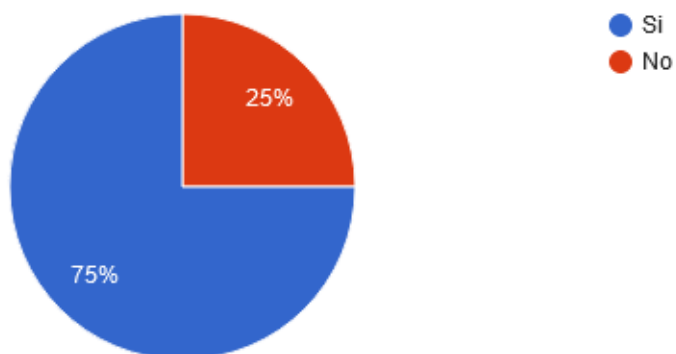


Gráfico 1. Mensajes de error al acceder al sistema (área de IT)

Mediante el análisis de la información podemos observar que 3 de los 4 encuestados aseguran que han encontrado un mensaje de error cuando tratan de acceder al sistema de almacenamiento.

Pregunta 2. ¿Alguna vez el sistema de almacenamiento falló en su área de trabajo y le impidió trabajar durante horas o días?

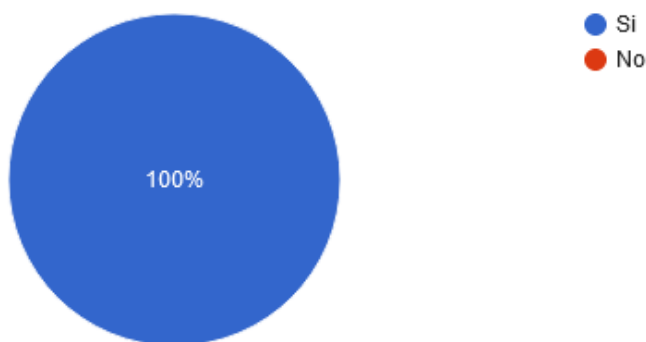


Gráfico 2. Fallo en el área de trabajo (área de IT)

Claramente se puede observar que los 4 trabajadores del área han experimentados fallos en el sistema lo cual les impidió poder trabajar durante un periodo de tiempo.

Pregunta 3. ¿Siente que los tiempos de respuesta del sistema es lento?

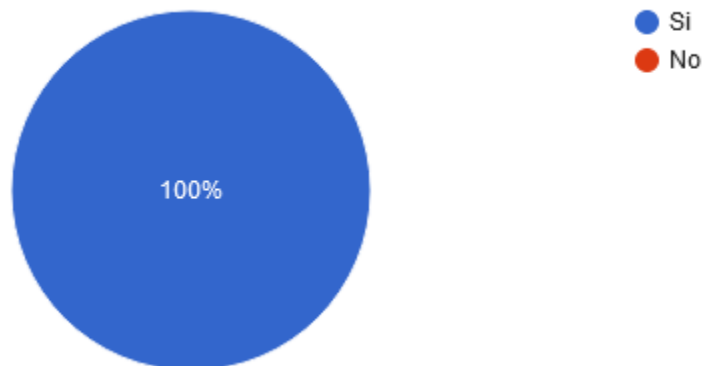


Gráfico 3. Tiempo de respuesta lento del sistema (área de IT)

Los 4 trabajadores del área mencionan que los tiempos de respuesta del sistema es lento, lo cual retrasa las operaciones que ellos realizan.

Pregunta 4. ¿Alguna vez algún archivo que usa se corrompe o no puede tener acceso de nuevo?

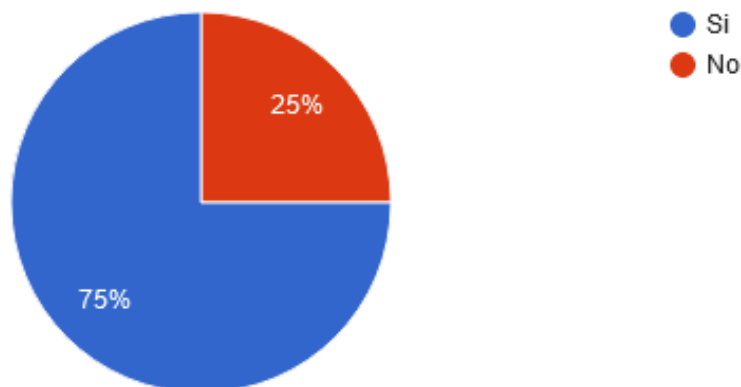


Gráfico 4. Corrupción de archivos (área de IT)

Podemos observar que 3 de los 4 trabajadores han experimentado en algún momento el malfuncionamiento de algún archivo o se les es imposible acceder a este.

Pregunta 5. ¿Siente que es fácil usar y navegar el sistema de archivos actual?

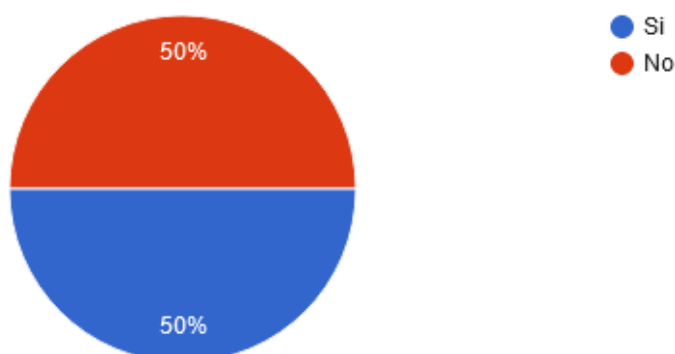


Gráfico 5. Navegabilidad del sistema (área de IT)

La mitad de los trabajadores del área consideran que el sistema de archivos actual es fácil de utilizar y de navegar a través de él, el resto está de acuerdo que es complicado su uso y navegación.

Pregunta 6. ¿Ha tenido que consultar con un técnico o superior por la razón que se presentó un tipo de mensaje de error o no respondía el sistema?

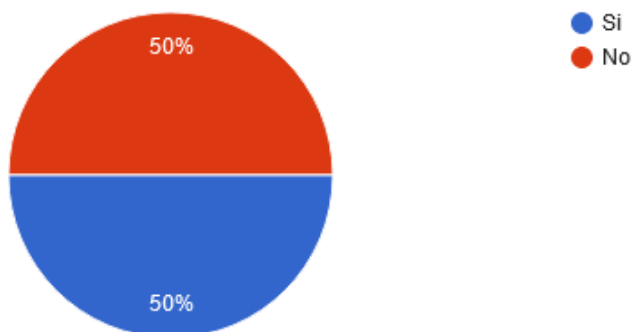


Gráfico 6. Consulta a técnico (área de IT)

El 50% de los trabajadores menciona que, si han tenido que recibir algún tipo de soporte o ayuda a un superior debido a algún fallo dentro del sistema de archivos, mientras que el resto aseguran que no han tenido que recurrir a alguna ayuda.

Pregunta 7. ¿Siente que la información es accesible cuando usted lo desea?

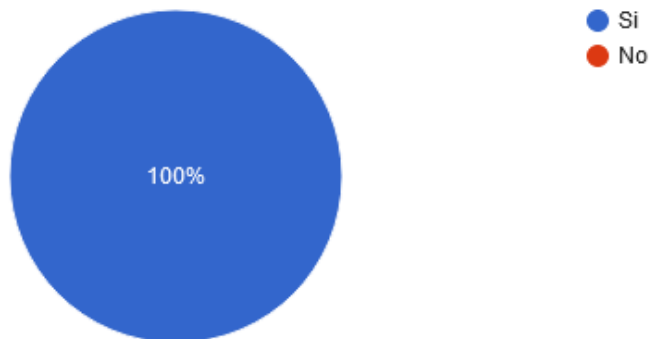


Gráfico 7. Accesibilidad del sistema (área de IT)

Los 4 trabajadores del área están de acuerdo que la información se encuentra disponible en el sistema de archivos cuando se desea.

Pregunta 8. En caso de pérdida de credenciales, ¿La experiencia ha sido rápida para su resolución?

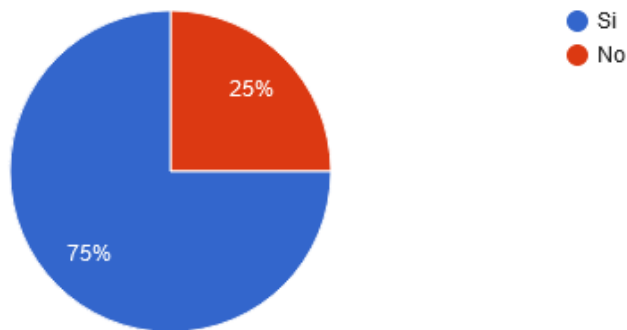


Gráfico 8. Rapidez de recuperación de credenciales (área de IT)

Podemos observar que 3 de los 4 trabajadores mencionan que la recuperación de credenciales se realiza de manera rápida, el resto califica este procedimiento como no tan rápido.

Pregunta 9. ¿Siente que es fácil administrar o darle mantenimiento al sistema actual?

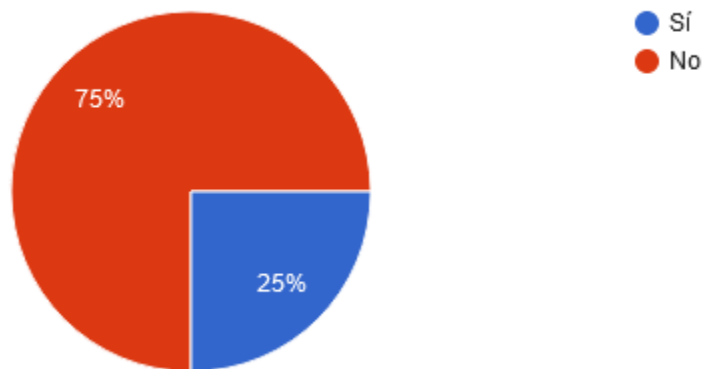


Gráfico 9. Administrar y mantenimiento del sistema actual

Según los datos obtenidos 3 de los 4 trabajadores concuerdan que administrar y darle mantenimiento al sistema de archivos actual no es fácil, en cambio 1 de los trabajadores asegura que es fácil

Pregunta 10. ¿Es fácil agregar almacenamiento extra cuando se necesita al sistema actual?

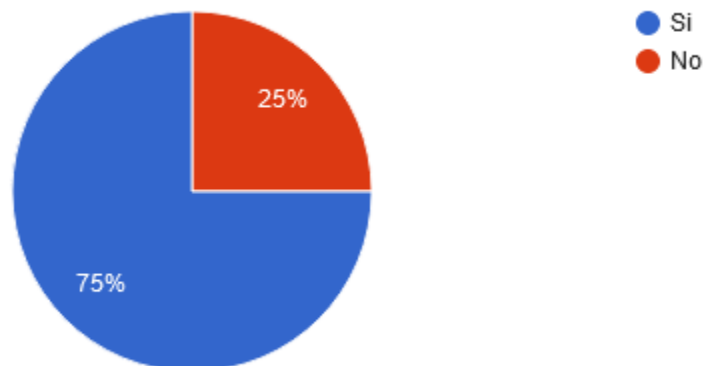


Gráfico 10. Facilidad de agregar almacenamiento extra

Como se puede observar el 75% es decir 3 de los 4 trabajadores dicen que es fácil agregar almacenamiento extra al sistema de archivos cuando se necesita.

Según los resultados obtenidos en el área de TI podemos destacar que el sistema de archivos actual presenta fallas aun para los trabajadores que le dan mantenimiento al mismo, también que para la mitad de los trabajadores es complicado su uso y su administración es complicada para 3 de los 4 encuestados, no obstante, la mayoría de los trabajadores concuerdan que es fácil poder agregar almacenamiento al sistema cuando se necesita y que la información se encuentra disponible en cualquier momento que se desee, 3 de los 4 trabajadores consideran fácil el proceso de recuperar credenciales de usuarios y la mitad de estos han tenido que consultar a algún supervisor debido a complicaciones con el sistema.

A continuación, se presentan los resultados obtenidos del análisis e interpretación de los datos obtenidos mediante la encuesta realizada al área de cámara durante la primera semana de septiembre

Pregunta 1. ¿Alguna vez se ha encontrado con algún mensaje de error cuando quiere acceder al sistema de almacenamiento?

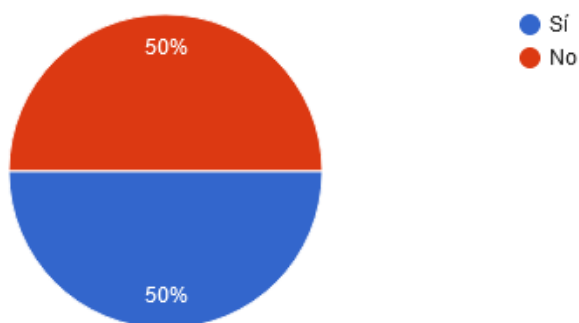


Gráfico 11. Errores al acceder al sistema (área de cámara)

Según la información obtenida y la cantidad de personas encuestadas podemos observar que 2 de los 4 trabajadores del área han encontrado un error al momento acceder al sistema de almacenamiento.

Pregunta 2. ¿Alguna vez el sistema de almacenamiento falló en su área de trabajo y le impidió trabajar durante horas o días?

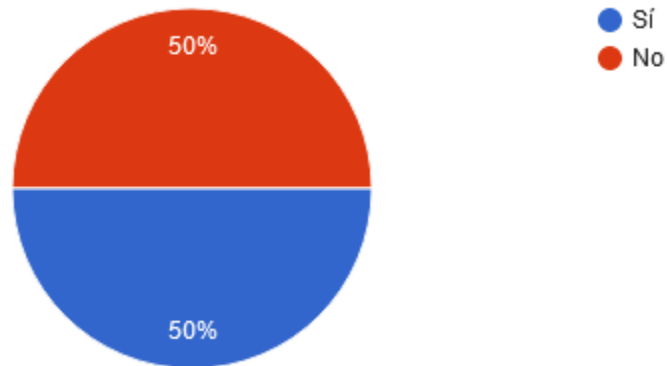


Gráfico 12. Fallos en área de trabajo (área de cámara)

Como podemos observar el 50% de los trabajadores del área ha experimentado una falla del sistema de almacenamiento en su área de trabajo lo cual le impidió trabajar por un periodo de tiempo.

Pregunta 3. ¿Siente que los tiempos de respuesta del sistema es lento?

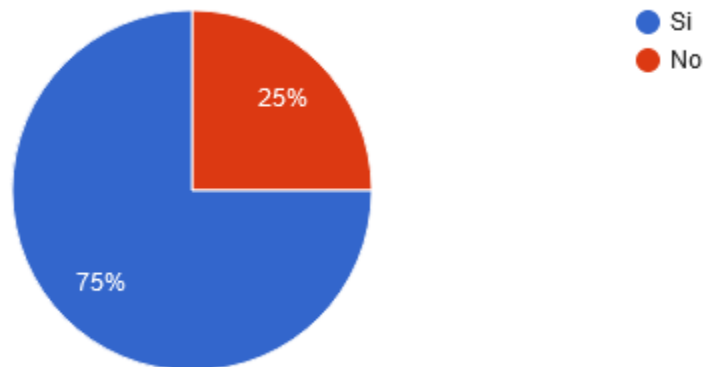


Gráfico 13. Tiempos de respuesta lentos del sistema (área de cámara)

Según la información obtenida se observa que el 75% de los trabajadores es decir 3 de ellos consideran que el tiempo de respuesta del sistema de archivos es lento y solamente 1 considera que el tiempo de respuesta es normal.

Pregunta 4. ¿Alguna vez algún archivo que usa se corrompe o no puede tener acceso de nuevo?

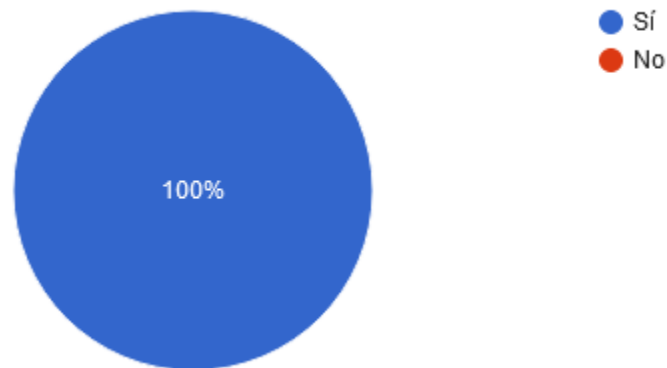


Gráfico 14. Corrupción de archivos (área de cámara)

Como se puede observar el 100% de los encuestados es decir 4 de los trabajadores han experimentado la pérdida de algún archivo o han tenido acceso limitado a estos.

Pregunta 5. ¿Siente que es fácil usar y navegar el sistema de archivos actual?

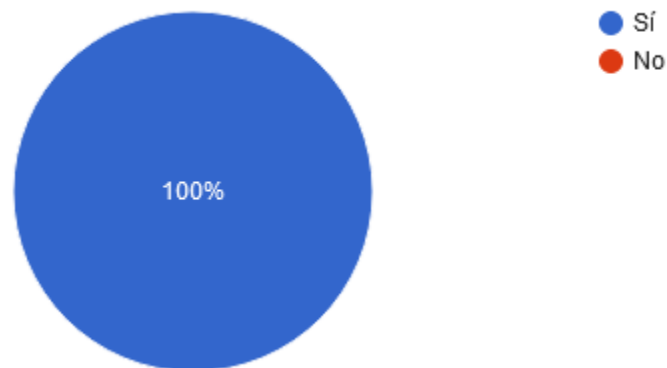


Gráfico 15. Navegabilidad del sistema (área de cámara)

Los 4 trabajadores encuestados concuerdan que el sistema de archivos actual es fácil de usar y navegar.

Pregunta 6. ¿Ha tenido que consultar con un técnico o superior por la razón que se presentó un tipo de mensaje de error o no respondía el sistema?

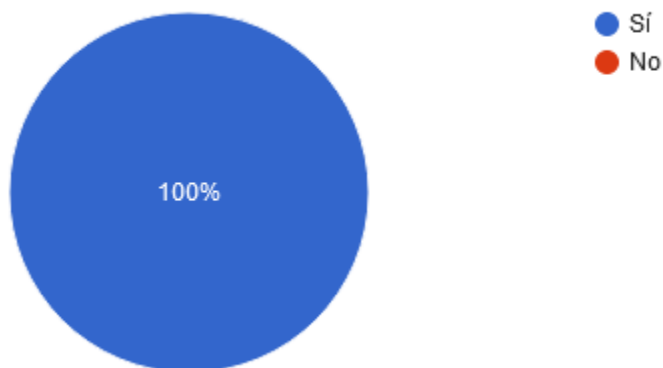


Gráfico 16. Consulta a técnico (área de cámara)

El 100% es decir los 4 trabajadores del área en algún momento han tenido que solicitar apoyo técnico a un superior para poder seguir operando el sistema de archivos

Pregunta 7. ¿Siente que la información es accesible cuando usted lo desea?

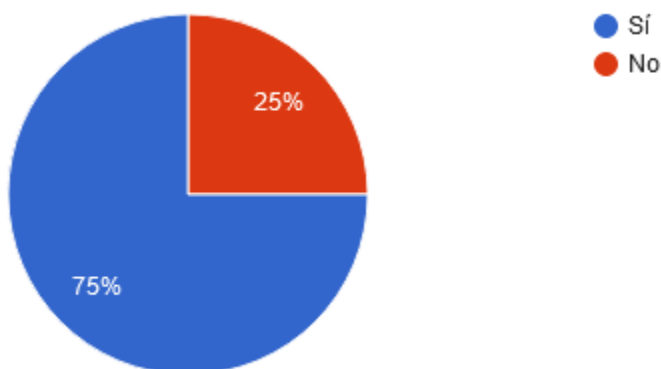


Gráfico 17. Accesibilidad del sistema (área de cámara)

El 75%, 3 de los 4 trabajadores del área concuerdan que la información se encuentra accesible cuando lo desean dentro del sistema de archivos.

Pregunta 8. En caso de pérdida de credenciales, ¿La experiencia ha sido rápida para su resolución?

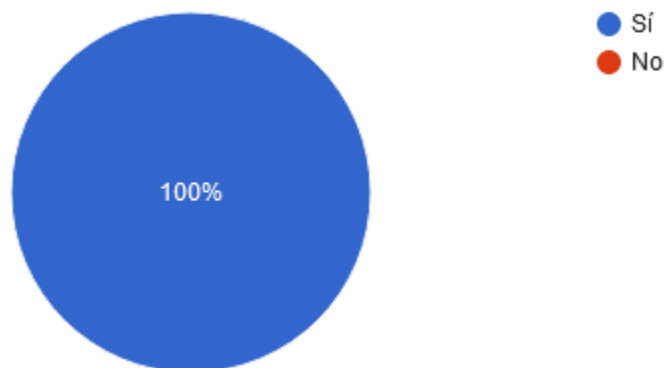


Gráfico 18. Rapidez de recuperación de credenciales (área de cámara)

Los 4 trabajadores del área concuerdan que en caso de pérdidas de credenciales se les ha dado una resolución de manera rápida.

Pregunta 9. ¿Considera que el tiempo de subida de archivos es lento?

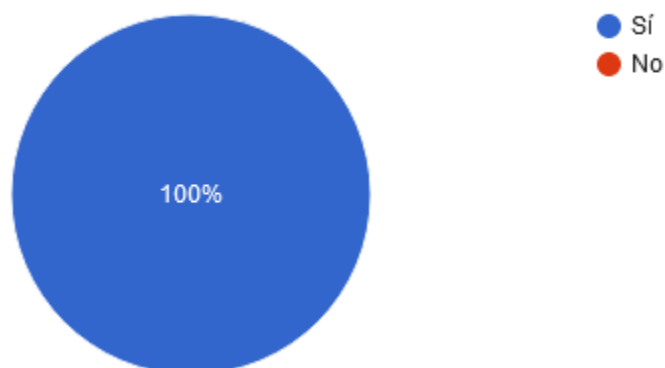


Gráfico 19. Tiempo de subida es lento

Como se puede observar según el gráfico los 4 trabajadores del área consideran que el sistema de archivos presenta tiempos de subida lentos.

Pregunta 10. ¿Considera que los archivos se guardan de forma correcta cuando usted lo solicita?

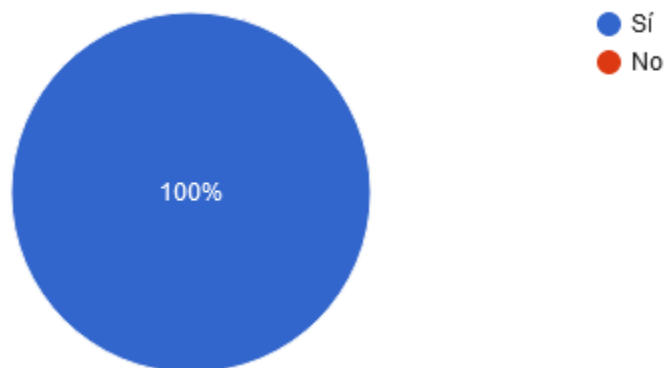


Gráfico 20. Forma de guardar los archivos

El 100% de los trabajadores aseguran que el sistema guarda de manera correcta los archivos cuando se le solicita.

Según la información recopilada dentro del área de cámaras podemos observar que la mitad de los trabajadores en algún momento ha recibido un mensaje de error al utilizar el sistema de archivos actual, esto concuerda con los datos obtenidos de los trabajadores que en algún momento no pudieron continuar desarrollando sus labores por un periodo de tiempo debido a un error en el sistema. Para 3 de los 4 trabajadores del área el sistema de archivos actual presenta tiempos de respuesta largos lo cual retrasa el desarrollo de sus funciones, de igual manera la subida de archivos es lenta lo cual puede deberse a que estos cargan los archivos crudos obtenidos de las cámaras de alta resolución. Todos los trabajadores concuerdan que el sistema ya ha dañado archivos en algún momento, que es fácil de utilizar y que los datos se guardan de manera correcta.

A continuación, se presentan los resultados del análisis e interpretación de los datos obtenidos mediante la encuesta realizada al área de edición durante la primera semana de septiembre

Pregunta 1. ¿Alguna vez se ha encontrado con algún mensaje de error cuando quiere acceder al sistema de almacenamiento?

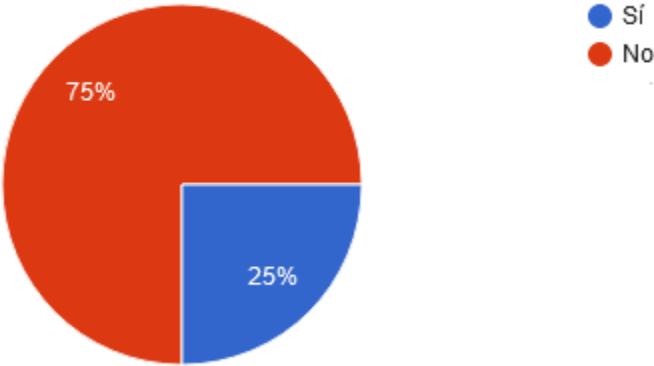


Gráfico 21. Error al acceder al sistema (área de edición)

Como se puede observar 6 de los 8 trabajadores del area no han encontrado algun mensaje de error cuando tratan de acceder al sistema de almacenamiento.

Pregunta 2. ¿Alguna vez el sistema de almacenamiento falló en su área de trabajo y le impidió trabajar durante horas o días?

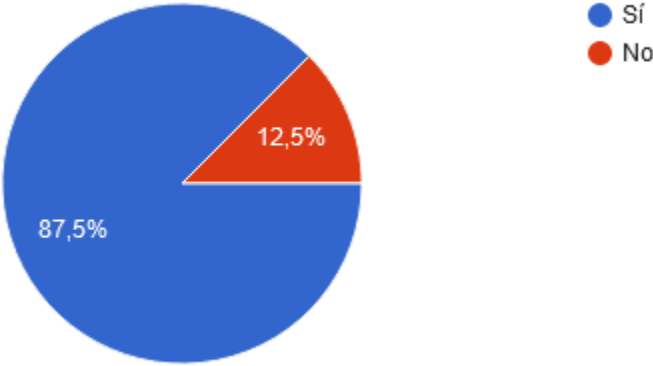


Gráfico 22. Falla en área de trabajo del sistema (área de edición)

Según la información obtenida el 87.5% es decir 7 de los 8 trabajadores ha experimentado la falla del sistema de almacenamiento en su área de trabajo lo cual les impidió trabajar por un periodo de tiempo.

Pregunta 3. ¿Siente que los tiempos de respuesta del sistema son lentos?

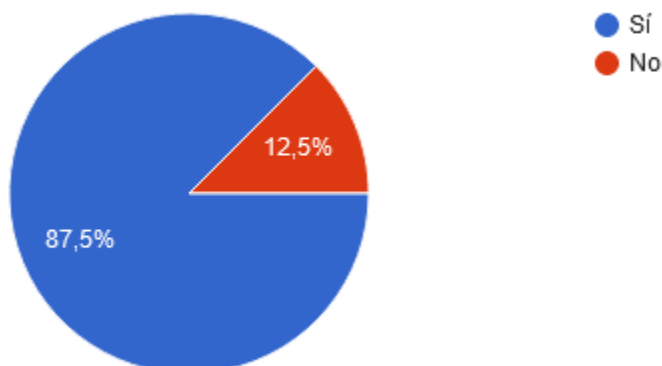


Gráfico 23. Tiempos de respuesta lentos del sistema (área de edición)

Como se puede observar 7 de los 8 trabajadores consideran que los tiempos de respuesta del sistema son lentos.

Pregunta 4. ¿Alguna vez algún archivo que usa se corrompe o no puede tener acceso de nuevo?

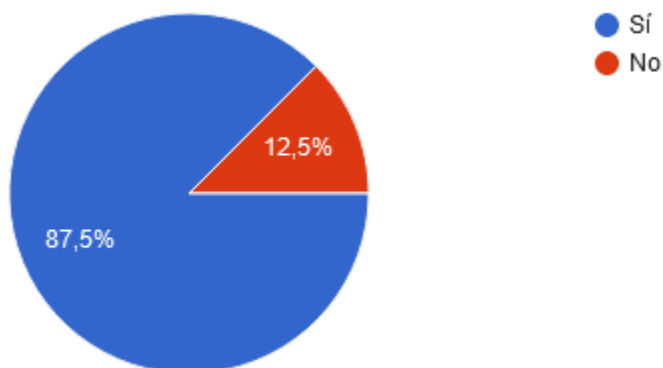


Gráfico 24. Corrupción de Archivos en el Sistema (Área de Edición)

Como se puede observar el 87.5% es decir 7 de los 8 trabajadores del área han experimentado la falta de acceso o pérdida de información en los archivos del sistema

Pregunta 5. ¿Siente que es fácil usar y navegar el sistema de archivos actual?

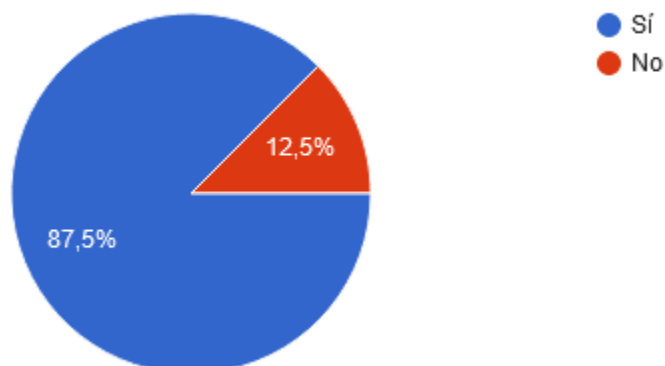


Gráfico 25. Facilidad de Uso y Navegabilidad del Sistema (Área de Edición)

Según la información obtenida, 7 de los 8 trabajadores consideran que el sistema de archivos actual es fácil de usar y navegar.

Pregunta 6. ¿Ha tenido que consultar con un técnico o superior por la razón que se presentó un tipo de mensaje de error o no respondía el sistema?

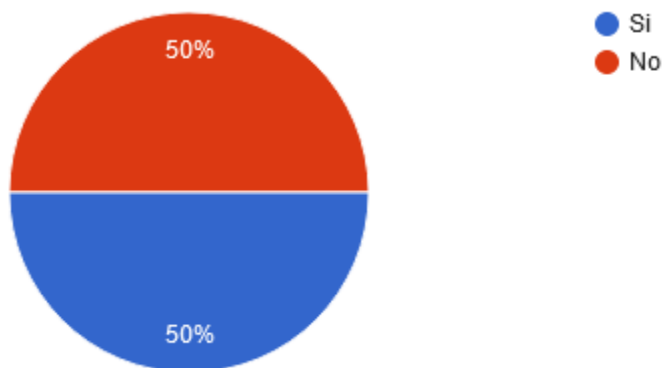


Gráfico 26. Facilidad de Consulta a Técnico (Área de Edición)

El 50% de los trabajadores es decir 4 de ellos han tenido que solicitar ayuda o asistencia técnica con algún superior debido a errores en el sistema.

Pregunta 7. ¿Siente que la información es accesible cuando usted lo desea?

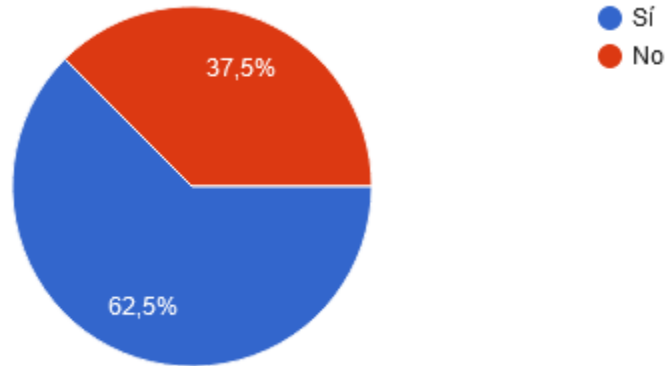


Gráfico 27. Accesibilidad de Archivos del Sistema (Área de Edición)

Como se puede observar el 62.5% de los encuestados es decir 5 de los 8 trabajadores del área aseguran que la información es accesible cuando se lo desea dentro del sistema de archivos.

Pregunta 8. En caso de pérdida de credenciales, ¿La experiencia ha sido rápida para su resolución?

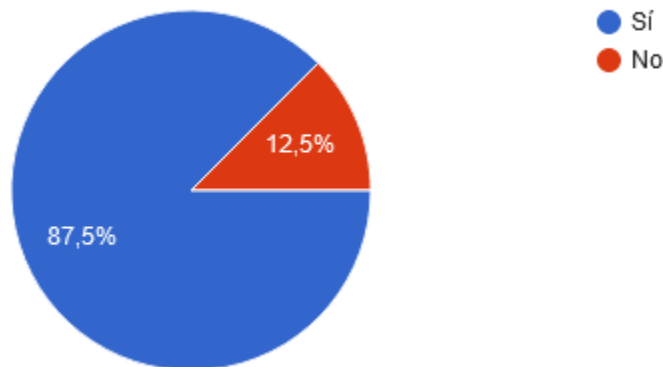


Gráfico 28. Recuperación de Credenciales (Área de Edición)

Según la información obtenida podemos observar que 7 de los 8 trabajadores han dicho que cuando se tienen pérdidas de credenciales se les ha dado una resolución de manera rápida.

Pregunta 9. ¿Considera que la descarga de archivos del sistema actual es lenta?

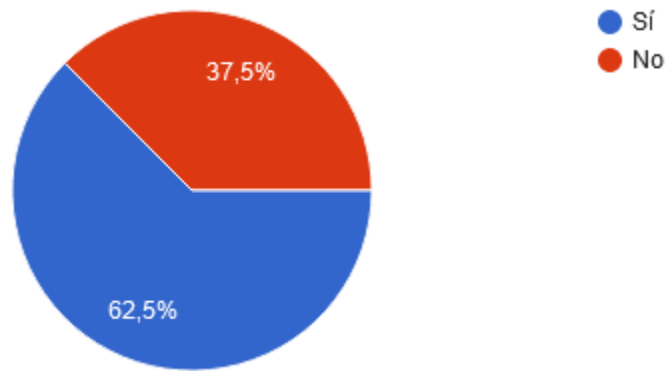


Gráfico 29. Descarga de Archivos del Sistema

Como se puede observar el 62.5% de los trabajadores consideran que el sistema de archivos brinda velocidades de descarga lentas.

Pregunta 10. ¿Considera que el tiempo de subida de archivos es lento?

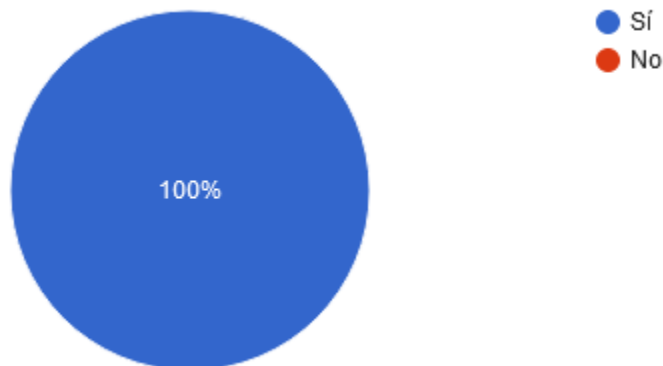


Gráfico 30. Tiempos de Subida de Archivos

Con la información obtenida se puede observar que los 8 trabajadores encuestados del área consideran que los tiempos de subida de los archivos son lentos.

Al realizar un análisis de la información recopilada podemos observar la situación del área de edición con respecto al sistema de archivos actual, a diferencia de las demás áreas encuestadas podemos notar que el área de edición no presenta la misma tendencia con respecto a los errores en el sistema en este caso solo 2 de los 8 trabajadores han recibido un mensaje de error y 6 de ellos han percibido fallos en el sistema lo cual les ha impedido continuar sus labores durante un periodo de tiempo, según el área encuestada solo 2 de los 8 trabajadores coinciden que el sistema presenta tiempos lentos, que los archivos se pueden corromper al manipularlos y que es fácil de usar, la gran mayoría concuerda que los aspectos de administración de usuarios y recuperación de credenciales es fácil y rápida, 5 de los 8 trabajadores concuerdan que la información es accesible cuando se le desea y que al subir y descargar los archivos se toma bastante tiempo.

9. Desarrollo de Prototipo

9.1. Análisis y Diseño

9.1.1. Análisis

9.1.1.1. *¿Qué se hará?*

Según el comparativo de Nube Privada y Nube Pública (Ver Anexo 8), se puede decir que la Nube Privada posee la mayor cantidad de cualidad que la corporación necesita para el soporte de sus servicios como lo son:

- **Costo:** Con una nube privada a largo plazo presenta un ahorro a comparación de una Nube Pública.
- **Seguridad:** Con una nube privada la CCR-TV es la encargada de la seguridad de esta, mientras que en la pública todo depende del proveedor.
- **Privacidad:** En la nube privada La CCR-TV tiene el control total de los archivos, a comparación de una nube pública los cuales se deja en manos del proveedor el manejo de la información.
- **Rendimiento:** En la nube privada la CCR-TV puede elegir el hardware de acuerdo a las necesidades y capacidades económicas.
- **Flexibilidad:** La nube privada se pueden instalar los elementos que se necesitan y ser escalable de acuerdo a lo que la CCR-TV necesite.

Entonces para darle una solución a los problemas obtenidos en la investigación, se hará despliegue de una nube privada. Esto es importante ya que la corporación necesita tener la escalabilidad si se quiere aumentar la nube en cuestiones de almacenamiento y aplicativo web para el manejo de archivos, también disminuir costos de mantenimiento del

almacenamiento ya que lo que se requiere es una inversión grande al inicio para la infraestructura inicial pero luego la inversión es mínima y solo es necesaria cuando se requiera aumentar recursos. Otro punto importante es la seguridad y control sobre la información, ya que una nube privada permite el manejo propio de toda la infraestructura, y así no depender de terceros para garantizar estos aspectos importantes.

También se tendrá que buscar la implementación de un sistema de almacenamiento distribuido capaz de replicación de datos, alta disponibilidad, lo suficiente robusta para almacenar grandes cantidades de archivos multimedia y flexible si se desea aumentar la capacidad de forma fácil.

No solo basta con tener la infraestructura y el sistema de almacenamiento funcionando, también es necesario poder tener un aplicativo en el cual los usuarios puedan administrar de forma gráfica y sencilla todos los archivos, para lo cual se desarrollará un aplicativo web que posea alta disponibilidad.

9.1.1.2. *¿Qué se usará?*

Para el despliegue de la nube privada hay varias soluciones, entre estos los proyectos más grandes que dominan el mercado son:

- **Apache Cloudstack**
- **OpenNebula**
- **Eucalyptus**
- **Nimbus**
- **Openstack**

Según el comparativo de las cinco soluciones (Ver Anexo 7) podemos decir que Openstack es la mejor opción por la mayor cantidad de soporte tiene de parte de la comunidad, la extensa documentación, además que está respaldada por organizaciones como la NASA, además poco a poco se ha vuelto un estándar de la industria para el despliegue de nubes privadas, por su modularidad lo hace flexible, es decir solo se instala lo que se necesita y si en un momento se quieren agregar otros servicios se puede hacer de forma sencilla. OpenStack nos permite desplegar máquinas virtuales y otras instancias que manejan las diferentes tareas de la nube. OpenStack satisface todas las necesidades de la corporación: flexible, escalable, seguridad y almacenamiento flexible, escalable, redundante y distribuido.

El prototipo estará enfocado al almacenamiento por lo cual se necesitará implementar alguna tecnología de almacenamiento compatible con OpenStack de acuerdo a las necesidades. De acuerdo a la documentación de Openstack están las siguientes:

- **Ceph**
- **Gluster**
- **LVM**
- **iSCSI**
- **NFS**
- **Sheepdog**
- **ZFS**

Con base a las características de cada uno (Ver Anexo 6), se ha elegido a Ceph a utilizar ya que es el único que tiene características que se adaptan a las necesidades de la corporación, entre las que están:

- Los administradores tienen mejor control sobre las estrategias de distribución y replicación de los datos.
- Provee consolidación del almacenamiento de objetos y del almacenamiento en bloque.
- Da un aprovisionamiento rápido de instancias de arranque desde volumen mediante **aprovisionamiento ligero**.
- Fácil compatibilidad e implementación con Openstack.

Solo el aprovisionamiento de la nube privada con el clúster Ceph no es necesario para el prototipo, si no también es necesario el despliegue de una aplicación web de alta disponibilidad, y se hará uso de Kubernetes (k8s), ya que Kubernetes nos permite tener de forma rápida y sencilla un despliegue de la aplicación escalable y con alta disponibilidad y capaz de tener aisladas cada aplicación.

9.1.1.3. ¿Cómo se va a desplegar?

Instalar OpenStack manualmente es una tarea tediosa, ya que conlleva la instalación de varios componentes y configuración de diferentes archivos necesarios para el correcto funcionamiento de la plataforma, etc. por lo tanto se han analizado varias formas de desplegar OpenStack. La mejor forma que se ha encontrado es **Kolla-Ansible**. Ya que la misión de Kolla es proporcionar contenedores listos para producción y herramientas de despliegue para operar nubes OpenStack. (OpenStack, 2021).

Para la instalación de Ceph se tendrá que buscar una solución aparte, ya que desde la versión Rocky de Openstack Kolla-Ansible no proporciona soporte para el aprovisionamiento y la configuración de un clúster Ceph directamente. En su lugar, los administradores deben utilizar una herramienta dedicada a este propósito (OpenStack, 2021b), por lo tanto, se usará

la herramienta **Ceph-ansible**, ya que igual que Kolla-ansible se despliegan contenedores usando ansible para implementar el clúster ceph, de acuerdo a la documentación este se tendrá que desplegar primero.

Para el despliegue de Kubernetes se hará uso de **Magnum** de OpenStack ya que provee de un servicio que tiene como objetivo hacer que los motores de orquestación de contenedores como Kubernetes sean accesibles como recursos de primera clase dentro de OpenStack.

9.2. Diseño

9.2.1. Requerimientos

El prototipo se desarrollará en base a los requerimientos mínimos de OpenStack¹, los cuales se detallan a continuación en la siguiente imagen:

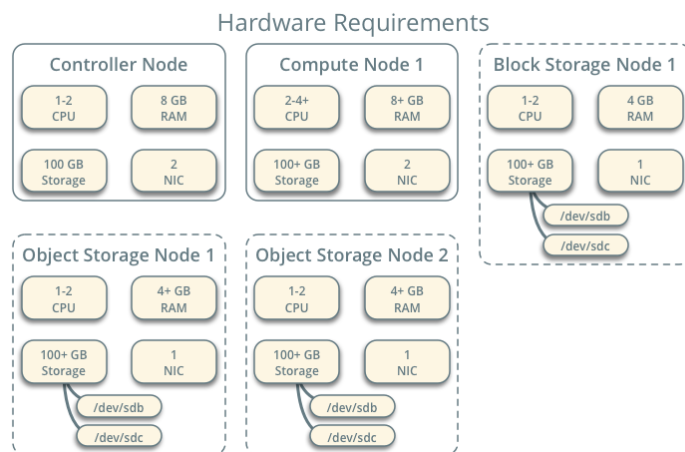


Figura 9. Requerimientos Mínimos de Openstack

1 <https://docs.openstack.org/install-guide/overview.html#example-architecture>

Se hará uso de 7 máquinas las cuales se detalla a continuación:

- **Deployer:** Máquina que contiene los archivos de configuración del clúster Ceph y Openstack usando Ansible, tiene que tener suficiente tamaño de disco para guardar estos programas y suficiente RAM para poder desplegar el sistema.
- **Monitor:** Servidor encargado de contener los servicios de monitoreo del clúster Ceph como lo son Grafana, Prometheus y el Ceph Monitor
- **Controller:** Servidor encargado de contener los servicios de control de Openstack (Glance, Neutron, Cinder, Keystone, Heat, Horizon, Magnum, Nova-Conductor y Nova-Scheduler) y también el Ceph manager del clúster de Ceph
- **Compute1, 2, 3 y 4:** Nodos Compute de Openstack y también contienen un disco agregado para un OSD de Ceph cada uno.

Las recomendaciones para la máquina que hará el deploy de OpenStack, son las siguiente:

- 2 Interfaces de Red
- 2GB RAM
- 40GB Espacio de Disco

En resumen, se tienen los siguientes requerimientos:

Tabla 6. Configuraciones de Hardware de los Nodos del Sistema

Requerimientos Para Nodos				
Nombre	vCPUs	RAM	Almacenamiento	Sistema Operativo
Deployer	1	4 GB	60 GB	Ubuntu Server Focal 20.04
Controller	2	8 GB	100 GB	Ubuntu Server Focal 20.04
compute1 (ceph-osd0)	4	8 GB	60GB +100 GB	Ubuntu Server Focal 20.04
compute2 (ceph-osd1)	4	8 GB	60GB +100 GB	Ubuntu Server Focal 20.04
compute3 (ceph-osd2)	4	8 GB	60GB +100 GB	Ubuntu Server Focal 20.04
Monitor	1	2 GB	40 GB + 100 GB	Ubuntu Server Focal 20.04
Total	16 vCPUs	38 GB	780 GB	Ubuntu Server Focal 20.04

9.2.2. Topología de Red

La topología de la red sería la siguiente:

Tabla 7. Asignaciones de IP de los Nodos

	Red Externa (Pública)	Red de administración (Interna)	Red de almacenamiento	Red Neutron
Hostname	enp1s0	enp2s0	enp3s0	enp4s0
Deployer	192.168.122.3	192.168.2.3	192.168.4.3	N/A
Controller	192.168.122.4	192.168.2.4	192.168.4.4	N/A

Compute1 (ceph-osd1)	192.168.122.5	192.168.2.5	192.168.4.5	N/A
Compute2 (ceph-osd2)	192.168.122.6	192.168.2.6	192.168.4.6	N/A
Compute3 (ceph-osd3)	192.168.122.7	192.168.2.7	192.168.4.7	N/A
Monitor	192.168.122.8	192.168.2.8	192.168.4.8	N/A

9.2.2.1. Diagrama Topología Física

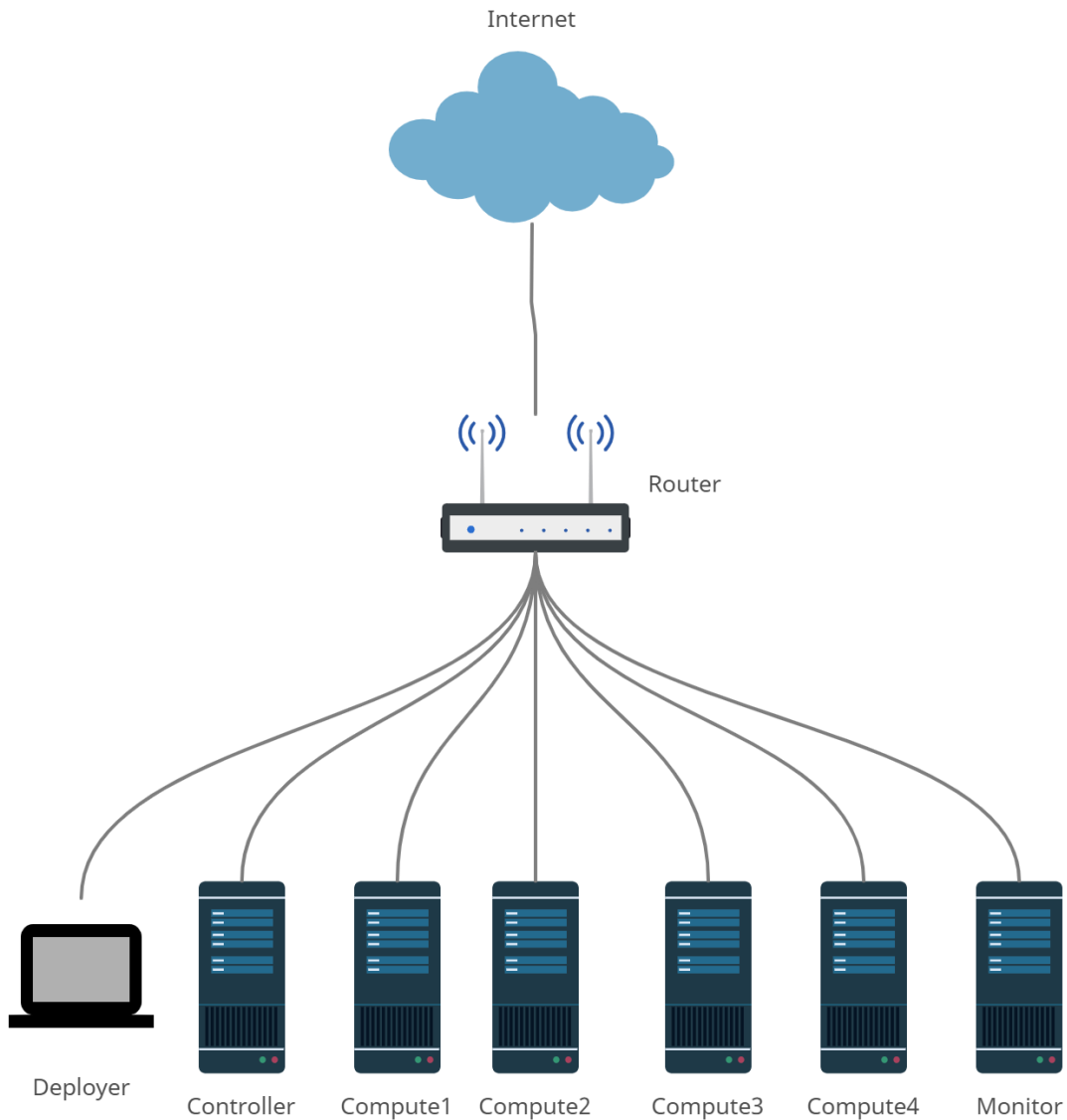


Figura 10. Diagrama de Topología Física

9.2.2.2. Diagrama Topología Lógica

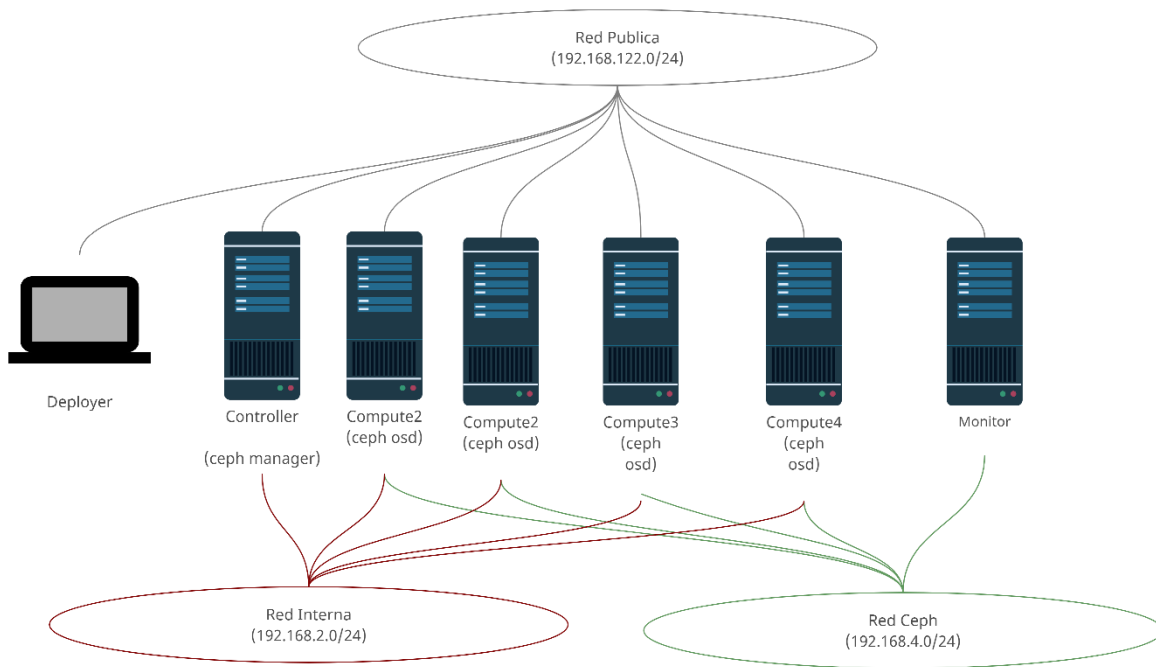


Figura 11. Diagrama de Topología Lógica

9.2.3. Consideraciones previas

- Numero de NIC tiene que ser 4:
 - enp1s0: Es la interfaz de red para el segmento de red externa, encargada de dar conexión a Internet y despliegue del sistema.
 - enp2s0: Es la interfaz de red para la gestión de interna de Openstack.
 - enp3s0: Es la interfaz de red para la red de almacenamiento del clúster.
 - enp4s0: Es la interfaz de red para la red de neutron, no debe ser asignada.
- Cada nodo tiene que tener un usuario sudo sin contraseña y que la máquina “Deployer” pueda conectarse con *ssh* a cada nodo sin pedir ningún tipo de contraseña.
- Primero se debe implementar **Ceph** en los nodos compute1, compute2 y monitor.

9.3. Pruebas

9.3.1. Funcionamiento de clúster Ceph

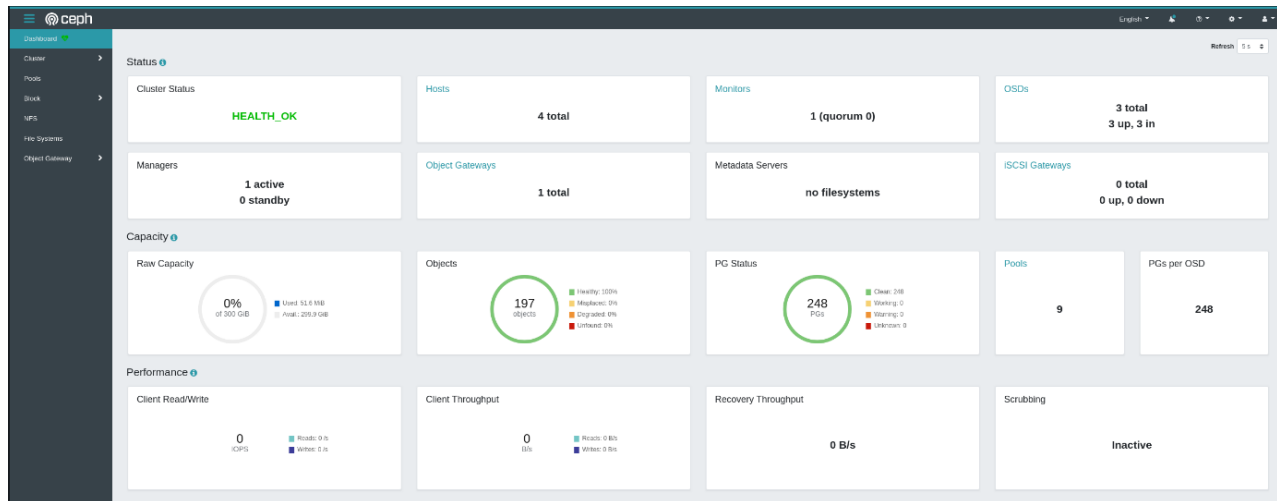


Figura 12. Dashboard de Ceph

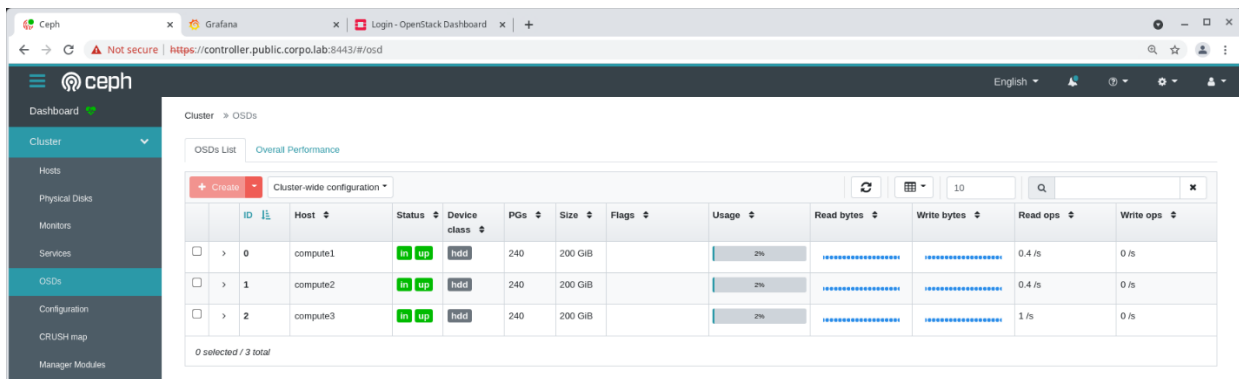


Figura 13. OSDs de Ceph Funcionando

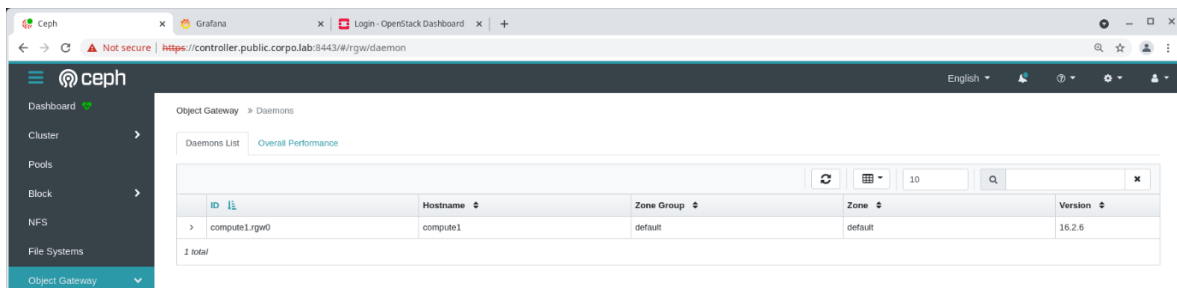


Figura 14. RGW funcionando correctamente

		Name	Owner	Used Capacity	Capacity Limit %	Objects	Object Limit %
<input type="checkbox"/>	>	bucket01	corporacion	0 B	No Limit	0	No Limit

0 selected / 1 total

Figura 16. Bucket creado en Ceph usando RGW

Object Gateway » Users

Username

corporacion

Details Keys

1 selected / 1 total

1 selected / 3 total

Figura 15. Usuario de Object Gateway con sus credenciales Creadas

		Username	Tenant	Full name	Email address	Suspended	Max. buckets	Capacity Limit %	Object Limit %
<input type="checkbox"/>	>	ceph-dashboard		Ceph dashboard			1000	No Limit	No Limit
<input type="checkbox"/>	>	corporacion		Corporación			1000	No Limit	No Limit
<input type="checkbox"/>	>	dashboard		Ceph Dashboard			1000	No Limit	No Limit

0 selected / 3 total

Figura 17. Listado de usuarios de Object Gateway

```

>>> ~ aws --profile=corpo --endpoint=http://compute1.public.corpo.lab:8080/ s3
ls s3://prueba
2021-10-03 22:38:00          30 prueba.txt
>>> ~ aws --profile=corpo --endpoint=http://compute1.public.corpo.lab:8080/ s3
cp prueba.txt s3://prueba
>>> ~ aws --profile=corpo --endpoint=http://compute1:8080 s3 mb s3://test
make_bucket failed: s3://test Could not connect to the endpoint URL: "http://compu
te1:8080/test"
>>> ~ aws --profile=corpo --endpoint=http://compute1.public.corpo.lab:8080/ s3
mb s3://test
make_bucket: test
>>> ~
>>> ~ aws --profile=corpo --endpoint=http://compute1.public.corpo.lab:8080/ s3
ls
2021-10-03 22:33:33 prueba
2021-10-07 18:45:17 test
>>> ~ aws --profile=corpo --endpoint=http://compute1.public.corpo.lab:8080/ s3
cp prueba.txt s3://test
upload: ./prueba.txt to s3://test/prueba.txt
>>> ~

```

Figura 18. Prueba de Funcionamiento de Ceph usando AWS CLI

Name	test
ID	a0f99d6d-95f9-4407-a751-c46b16827791.4204.2
Owner	corporacion
Index type	Normal
Placement rule	default-placement
Marker	a0f99d6d-95f9-4407-a751-c46b16827791.4204.2
Maximum marker	0#,1#,2#,3#,4#,5#,6#,7#,8#,9#,10#
Version	0#1,1#1,2#1,3#1,4#1,5#1,6#1,7#1,8#1,9#2,10#1
Master version	0#0,1#0,2#0,3#0,4#0,5#0,6#0,7#0,8#0,9#0,10#0
Modification time	9/27/21 7:29:56 PM
Zonegroup	b6ebafbe-1800-4c86-a1b9-f5dae2e39b54
Versioning	Suspended
MFA Delete	Disabled

Bucket quota

Figura 19. Descripción Bucket Ceph

9.3.2. Funcionamiento de Openstack

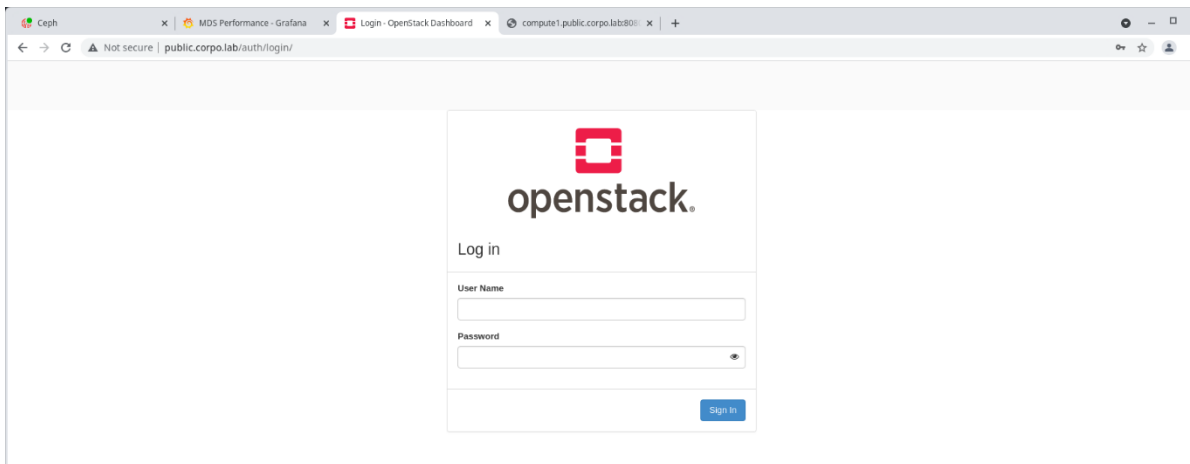


Figura 20. Login de Openstack

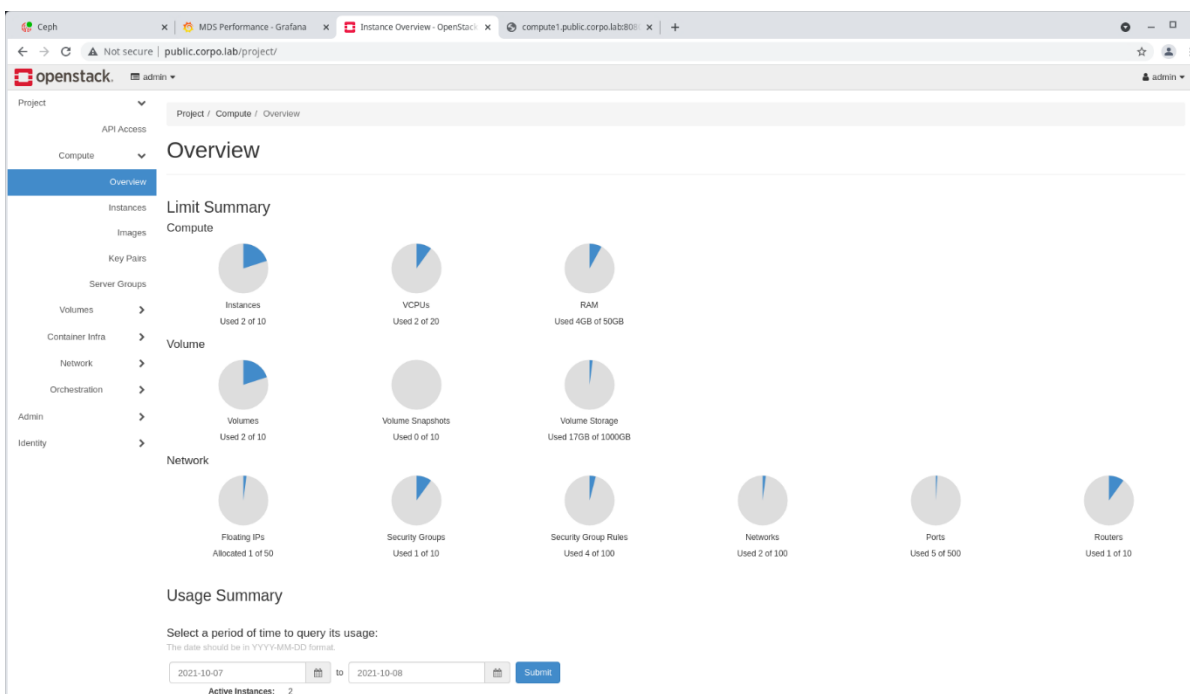


Figura 21. Overview del Servicio de Compute

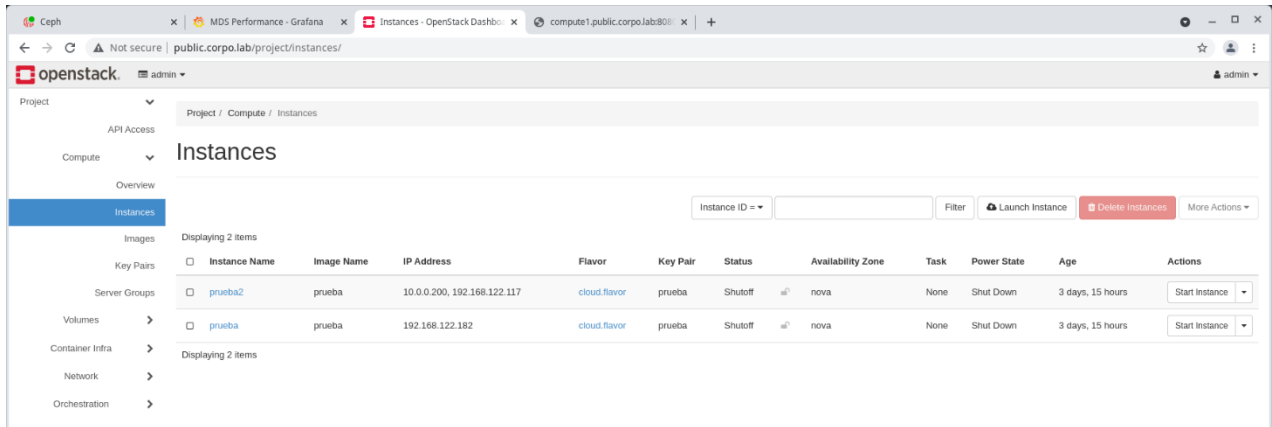


Figura 23. Listado de Instancias en Openstack

```
>>> ~ ssh corpo@deployer.public.corpo.lab
corpo@deployer.public.corpo.lab's password:
Welcome to Ubuntu 20.04.3 LTS (GNU/Linux 5.4.0-88-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

System information as of Fri 08 Oct 2021 12:51:31 AM UTC

System load:  0.0          Users logged in:      0
Usage of /:   28.3% of 28.91GB  IPv4 address for enp1s0: 192.168.122.3
Memory usage: 2%          IPv4 address for enp2s0: 192.168.2.3
Swap usage:   0%          IPv4 address for enp3s0: 192.168.4.3
Processes:   151

 * Super-optimized for small spaces - read how we shrank the memory
   footprint of MicroK8s to make it the smallest full K8s around.

   https://ubuntu.com/blog/microk8s-memory-optimisation

0 updates can be applied immediately.

Last login: Mon Oct  4 09:14:32 2021 from 192.168.122.1
corpo@deployer:~$ cd openstack/
corpo@deployer:~/openstack$ source admin-openrc.sh
corpo@deployer:~/openstack$ source openstack/bin/activate
(openstack) corpo@deployer:~/openstack$ openstack compute service list
+-----+-----+-----+-----+-----+-----+-----+
| ID | Binary          | Host       | Zone  | Status | State | Updated At          |
+-----+-----+-----+-----+-----+-----+-----+
| 1  | nova-scheduler | controller | internal | enabled | up    | 2021-10-08T00:52:07.000000 |
| 1  | nova-conductor | controller | internal | enabled | up    | 2021-10-08T00:52:15.000000 |
| 3  | nova-compute   | compute2  | nova  | enabled | up    | 2021-10-08T00:52:08.000000 |
| 4  | nova-compute   | compute1  | nova  | enabled | up    | 2021-10-08T00:52:07.000000 |
| 5  | nova-compute   | compute3  | nova  | enabled | up    | 2021-10-08T00:52:08.000000 |
+-----+-----+-----+-----+-----+-----+-----+
(openstack) corpo@deployer:~/openstack$
```

Figura 22. Comprobación de los servicios de Compute Funcionando Correctamente

```
(openstack) corpo@deployer:~/openstack$ openstack service list
```

ID	Name	Type
06458e5f5bc245958c12eb0217aeda3a	heat	orchestration
19cbfa00508048d9b6ea9b581b66aafe	heat-cfn	cloudformation
2701920561a7472c8e8d11d3336a9287	monasca-log-api	logging
34522e33127143f8b4188e051588a2c5	monasca-api	monitoring
3a714f19546741cb9182a5b5d4a8224e	nova_legacy	compute_legacy
6918666473d04685b28238e1844add49	nova	compute
7ee1033bb9394571901280eda8f1900c	keystone	identity
b9597d9431db473bb631d9e84b995b10	neutron	network
bf2318b04acc4e7ebd72c324db66c4f0	glance	image
ca8baf40bfd8419b94e87b0382b5f481	magnum	container-infra
cd7c6ffdf519408eb4b063ab9355bf81	cinderv3	volumev3
fa6ec6f748f446aa9f5288abe30b101f	placement	placement

```
(openstack) corpo@deployer:~/openstack$
```

Figura 26. Listado de Servicios de Openstack

```
(openstack) corpo@deployer:~/openstack$ openstack volume service list
```

Binary	Host	Zone	Status	State	Updated At
cinder-scheduler	controller	nova	enabled	up	2021-10-08T00:53:17.000000
cinder-volume	controller@rbd-1	nova	enabled	up	2021-10-08T00:53:17.000000
cinder-backup	controller	nova	enabled	down	2021-10-08T00:08:46.000000

```
(openstack) corpo@deployer:~/openstack$
```

Figura 25. Listado de Servicios de Volume Funcionando Correctamente

```
(openstack) corpo@deployer:~/openstack$ openstack network agent list
```

ID	Agent Type	Host	Availability Zone	Alive	State	Binary
controller	OVN Controller agent	controller		: -)	UP	ovn-controller
compute1	OVN Controller agent	compute1		: -)	UP	ovn-controller
compute2	OVN Controller agent	compute2		: -)	UP	ovn-controller
compute3	OVN Controller agent	compute3		: -)	UP	ovn-controller

```
(openstack) corpo@deployer:~/openstack$
```

Figura 24. Listado de Servicios de Neutron Funcionando Correctamente

10. Conclusiones

- Se aplicó una metodología de investigación e instrumentos para recolectar información de tipo cuantitativo y cualitativo, con los cuales se obtuvo información del modelo de negocio de la organización.
- Se ha logrado analizar la problemática actual de la organización y desarrollado una solución que cumple con los requisitos de la organización.
- Se logro definir el tipo de infraestructura que se utilizara para implementar la solución.
- Se completo el desarrollo de un prototipo de infraestructura en la nube que cumple con las necesidades de la organización para el almacenamiento y administración de copias de seguridad de archivos multimedia.

11. Recomendaciones

- Hacer uso de un clúster de Kubernetes para generar alta disponibilidad del aplicativo web para el manejo de archivos.
- Para la implementación en producción se recomienda desplegar los nodos del clúster Ceph con más capacidad de disco y a su vez agregar más nodos para aumentar la capacidad y mejorar la replicación de datos.
- Se recomienda el uso de una infraestructura de nube privada y administrada por la institución ya que presenta ventajas económicas ante otras soluciones administradas y desplegadas por terceros.

12. Bibliografía

- Aghayev, A., Weil, S., Kuchnik, M., Nelson, M., Ganger, G. R., & Amvrosiadis, G. (2019). File systems unfit as distributed storage backends: lessons from 10 years of Ceph evolution. In *Proceedings of the 27th ACM Symposium on Operating Systems Principles* (pp. 353-369).
- Burns, B., Beda, J., & Hightower, K. (2018). *Kubernetes*. Dpunkt.
- Burns, Brendan et al., “Borg, Omega, and Kubernetes: Lessons Learned from Three Container-Management Systems over a Decade,” *ACM Queue* 14 (2016): 70–93, Obtenido de: <http://bit.ly/2vIrL4S>.
- Citrix Systems. (2010). *Xen Cloud Platform – Advanced Virtualization Infrastructure for the Clouds*. Obtenido de: <http://www.xen.org/products/cloudxen.html>.
- Encinas, D; Kunysz, E; Galarza, B; Zaccardi, G; Morales, M. “Performance de arquitecturas multiprocesador: técnicas de modelado y simulación, plataformas reconfigurables y cloud computing.” XVII Workshop de Investigadores en Ciencias de la Computación (2015). Ciudad de Salta, Argentina.
- Fox, R., & Hao, W. (2018). *Internet Infrastructure Networking, Web Services, and Cloud Computing*.
- Fuentelsaz C. *Cálculo del tamaño de la muestra*. *Matronas Prof.* 2004;5(18):5-13.
- Garth A. Rodney V. (2000) *Communications of the ACM*, Volumen 43
- Harold F. Tipton, Micki Krause Nozak. (2012). *Information Security Management Handbook, Sixth Edition*, Volumen 6.
- Hoi C. Trieu C. (2012). *An Approach To High Availability For Cloud Servers With Snapshot Mechanism*
- Maltzahn, C., Molina-Estolano, E., Khurana, A., Nelson, A. J., Brandt, S. A., & Weil, S. (2010). *Ceph as a scalable alternative to the hadoop distributed file system*. *login: The USENIX Magazine*, 35, 38-49.

- Mesbahi, M. Rahmani, A. y Hosseinzadeh, M. (2018). Reliability and high availability in cloud computing environments: a reference roadmap
- OpenStack. (2018). Storage concepts. Obtenido de: <https://docs.openstack.org/arch-design/design-storage/design-storage-concepts.html>
- OpenStack. (2021). *Welcome to Kolla Ansible's Documentation!* OpenStack Documentation. Obtenido de: <https://docs.openstack.org/kolla-ansible/latest/>
- OpenStack. (2021b). External Ceph. Retrieved September 1, 2021, Obtenido de: <https://docs.openstack.org/kolla-ansible/latest/reference/storage/external-ceph-guide.html>
- Pepple, Ken July (2011). *Deploying OpenStack*. O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.
- Silva LC. Muestreo Para La Investigación En Ciencias De La Salud. Madrid: Díaz de Santos; 1993.
- Takako,P., Estcio,G., Kelner,J., Sadok,D. (2010) . *A Survey on Open-source Cloud Computing Solutions.WCGA - 8th Workshop on Clouds, Grids and Applications*.Gramado:28 May, 3-16. Teyssier,S. (2010).
- Velasco Rodríguez, V. M., Martínez O, Verónica Araceli, & Roiz Hernández, J. (2003). Muestreo y Tamaño de la Muestra. Una Guía Práctica Para Personal De Salud Que Realiza Investigación. El Cid Editor. Obtenido de: <http://site.ebrary.com/id/10022810>
- Weil, S. A., Brandt, S. A., Miller, E. L., Long, D. D., & Maltzahn, C. (2006). *Ceph: A scalable, high-performance distributed file system*. In Proceedings of the 7th symposium on Operating systems design and implementation (pp. 307-320).
- Weygant P. (2001). *Clustering For High Availability*.
- Zhang, Y., Xu, C., & Shen, X. S. (2020). *Data Security in Cloud Storage*.

13. Anexos



Anexo 1. Encuesta a empleados del área de IT

**UNIVERSIDAD DE EL SALVADOR
FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA DE INGENIERÍA DE SISTEMAS INFORMÁTICOS
CURSO DE ESPECIALIZACIÓN EN INFRAESTRUCTURA CLOUD**

Objetivo: Conocer del área de IT el efecto que posee los problemas de almacenamiento en la organización.

Marque con una X su respuesta.

Pregunta	Si	No
1. ¿Alguna vez se ha encontrado con algún mensaje de error cuando quiere acceder al sistema de almacenamiento?		
2. ¿Alguna vez el sistema de almacenamiento falló en su área de trabajo y le impidió trabajar durante horas o días?		
3. ¿Siente que los tiempos de respuesta del sistema es lento?		
4. ¿Alguna vez algún archivo que usa se corrompe o no puede tener acceso de nuevo?		
5. ¿Siente que es fácil usar y navegar el sistema de archivos actual?		
6. ¿Ha tenido que consultar con un técnico o superior por la razón que se presentó un tipo de mensaje de error o no respondía el sistema?		
7. ¿Siente que la información es accesible cuando usted lo desea?		
8. En caso de pérdida de credenciales, ¿La experiencia ha sido rápida para su resolución?		
9. ¿Siente que es fácil administrar o darle mantenimiento al sistema actual?		
10. ¿Qué tan fácil es agregar almacenamiento extra cuando se necesita al sistema actual?		

Anexo 2. Encuesta a empleados del área de Cámara

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA DE INGENIERÍA DE SISTEMAS INFORMÁTICOS
CURSO DE ESPECIALIZACIÓN EN INFRAESTRUCTURA
CLOUD



Objetivo: Conocer del área de Cámara el efecto que posee los problemas de almacenamiento en la organización.

Marque con una X su respuesta.

Pregunta	Si	No
1. ¿Alguna vez se ha encontrado con algún mensaje de error cuando quiere acceder al sistema de almacenamiento?		
2. ¿Alguna vez el sistema de almacenamiento falló en su área de trabajo y le impidió trabajar durante horas o días?		
3. ¿Siente que los tiempos de respuesta del sistema es lento?		
4. ¿Alguna vez algún archivo que usa se corrompe o no puede tener acceso de nuevo?		
5. ¿Siente que es fácil usar y navegar el sistema de archivos actual?		
6. ¿Ha tenido que consultar con un técnico o superior por la razón que se presentó un tipo de mensaje de error o no respondía el sistema?		
7. ¿Siente que la información es accesible cuando usted lo desea?		
8. En caso de pérdida de credenciales, ¿La experiencia ha sido rápida para su resolución?		
9. ¿Considera que el tiempo de subida de archivos es lento?		
10. ¿Considera que los archivos se guardan de forma correcta cuando usted lo solicita?		

Anexo 3. Encuesta a empleados del área de Edición

**UNIVERSIDAD DE EL SALVADOR
FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA DE INGENIERÍA DE SISTEMAS INFORMÁTICOS
CURSO DE ESPECIALIZACIÓN EN INFRAESTRUCTURA
CLOUD**



Objetivo: Conocer del área de edición el efecto que posee los problemas de almacenamiento en la organización.

Marque con una X su respuesta.

Pregunta	Si	No
1. ¿Alguna vez se ha encontrado con algún mensaje de error cuando quiere acceder al sistema de almacenamiento?		
2. ¿Alguna vez el sistema de almacenamiento falló en su área de trabajo y le impidió trabajar durante horas o días?		
3. ¿Siente que los tiempos de respuesta del sistema es lento?		
4. ¿Alguna vez algún archivo que usa se corrompe o no puede tener acceso de nuevo?		
5. ¿Siente que es fácil usar y navegar el sistema de archivos actual?		
6. ¿Ha tenido que consultar con un técnico o superior por la razón que se presentó un tipo de mensaje de error o no respondía el sistema?		
7. ¿Siente que la información es accesible cuando usted lo desea?		
8. En caso de pérdida de credenciales, ¿La experiencia ha sido rápida para su resolución?		
9. ¿Considera que la descarga de archivos del sistema actual es lenta?		
10. ¿Considera que el tiempo de subida de archivos es lento?		

Anexo 4. Entrevista a jefe del área de IT

**UNIVERSIDAD DE EL SALVADOR
FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA DE INGENIERÍA DE SISTEMAS INFORMÁTICOS
CURSO DE ESPECIALIZACIÓN EN INFRAESTRUCTURA CLOUD**



Entrevista a jefe del área de informática

Objetivo: Conocer la opinión de un alto mando de la empresa sobre la problemática planteada de resguardo y acceso de archivos multimedia en la corporación

1. ¿Cuál es el problema que considera más importante al momento acceder o hacer uso de algún tipo de contenido multimedia de la organización? (Almacenamiento, rápido acceso, entre otras)

Presentamos 2 grandes problemas, los que son los medios físicos de comunicación de los servidores como fibra óptica, cables ITP, aún la misma red refiriéndose al wifi y lo otro son los ISP, si no se cuenta con un buen servicio de internet y se quiere acceder a servicios en la red y dependiendo del tipo de servicio que requiera, si estos servicios son multimedia que requiere bastante margen de ancho de banda para video, audio tal vez no imagen, más que todo para video se necesita una conexión estable y tu servidor o proveedor de servicio no te garantiza esta conexión entonces no serviría de mucho.

2. ¿Ha recibido algún tipo de queja de parte de usuarios sobre el almacenamiento de información? ¿De qué sector se han generado la mayor cantidad de quejas?

Si hemos recibido algunas quejas tal vez no con frecuencia, pero te puedo resumir que el mayor problema son con los usuarios que acceden a los servidores de audio y video, nuestra infraestructura está diseñada con servidores NAS de almacenamiento de red, tenemos dos servidores NAS de video cada uno con su redundancia, un servidor de audio también NAS con su redundancia. La mayoría de usuarios los problemas que tienen son con el acceso, no pueden entrar, son por muchas razones la más común es porque pierden sus credenciales o no las introducen bien, otras que pasan difícilmente es porque el servidor se ha caído o necesita una actualización, eso por parte de radio. Por parte del canal, se divide en editores, cámaras, productores los servidores de ellos manejan suficientes TB de almacenamiento, han existido ciertas quejas, una de las más frecuentes es que se han quedado sin almacenamiento, nosotros contamos con 80 TB de almacenamiento, pero el contenido que ellos suben es en HD, ellos piensan que el almacenamiento es ilimitado.

3. ¿Qué datos considera más importante que actualmente no se está cubriendo de la forma más apropiada?

Te podría decir que por nuestro tipo de organización que somos podría hablar de videos y predicas del pastor general, que están almacenadas, pero me gustaría que tuvieran algún tipo de encriptación o una biblioteca especialmente como respaldo, que no se sabe lo que puede pasar y tal vez cierto material de documental, refiriéndonos más que todo a la cuestión de seguridad quiero tocar lo que es el acceso, si se pueden crear ciertas restricciones que cierto usuario no acceda a ciertas carpetas.

4. ¿Qué tan frecuente es su acceso a algún tipo de dato que se almacena en la organización en una semana? ¿Cuántas de esas veces se le presenta un error?


En mi caso como IT puede ser 2 o 3 veces a la semana tenemos una carpeta especial donde almacenamos las ISO de los sistemas operativos, librerías, paqueterías, en el caso de productores o editores es diario, su acceso a los servidores de almacenamiento, en la nube de ellos es a diario, aunque ellos editan sus archivos offline pero todos sus archivos están en la nube cuando ellos hacen la post producción o hacen edición final lo suben, igual en el canal las programadoras tienen que estar subiendo contenido para pasarlo en el canal señal al aire lo agarran directamente de los servidores y eso es a diario.

Hemos recibido algunas quejas de los usuarios del canal, programadores de pautas, editores, algunos productores el acceso, no pueden ingresar bien la contraseña. Al estar dentro del sistema se nos ha dado el caso de error, nuestro esquema es así, estos servidores están enlazados con otro servidor de DNS y un servidor DHCP, entonces, como son diferentes uno es Synology y el otro es Windows Server 2002, entonces la compatibilidad entre ambos es posible, sin embargo, cuando uno de los dos tiene actualizaciones se generan problemas, y ciertos usuarios por ejemplo, como tenemos a los usuarios por grupo, los usuarios editores no pueden pegar, porque cada usuario tiene sus permisos, entonces ahí nos damos cuenta de que hay problemas.

Anexo 5. Instalación de Ceph

Configuraciones De Cada Nodo

1. Configurar las IP de cada máquina usando netplan
2. Configurar el archivo `/etc/hosts` con las IP de los demás nodos a tal punto de poder hacer *ping* entre todas.
3. Para que el despliegue funcione, necesitaremos un usuario en todos nuestros nodos que tenga **sudo** sin contraseña y sea capaz de entrar con SSH en todos los nodos utilizando una clave SSH privada.



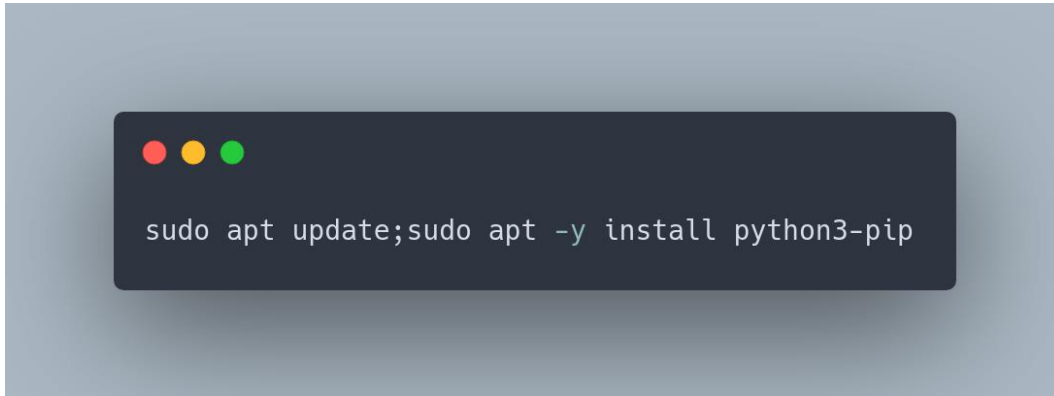
```
ubuntu@controller:~$ sudo useradd deploy-user -m -d /home/deploy-user -s /bin/bash
ubuntu@controller:~$ echo "deploy-user ALL=(ALL) NOPASSWD: ALL" | sudo tee /etc/sudoers.d/deploy-user
```

4. Crear llaves ssh en cada nodo

Depliegue de Cluster ceph

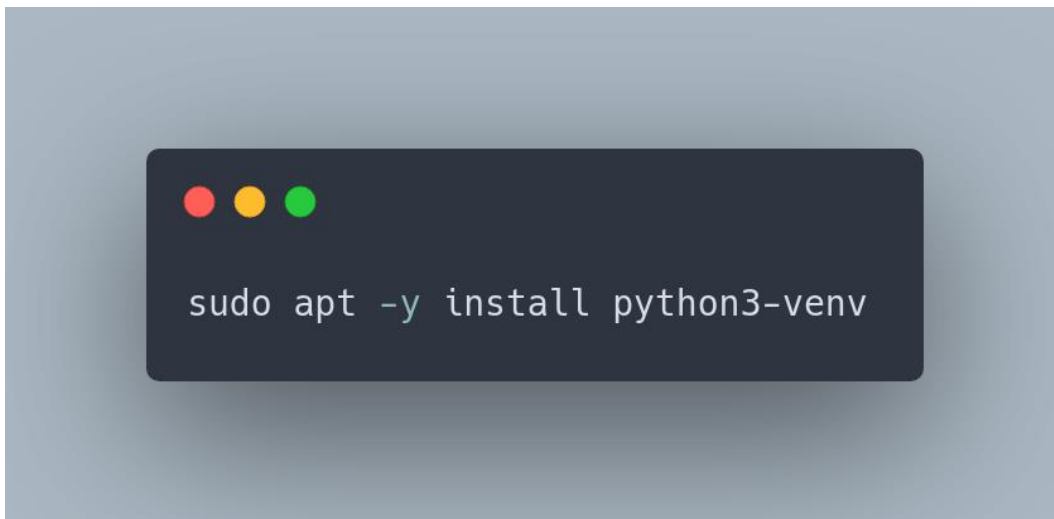
Configuración de Nodo Deployer

1. Instalar python3-pip:

A terminal window with a dark background and three colored window control buttons (red, yellow, green) in the top left corner. The terminal displays the command: `sudo apt update;sudo apt -y install python3-pip`

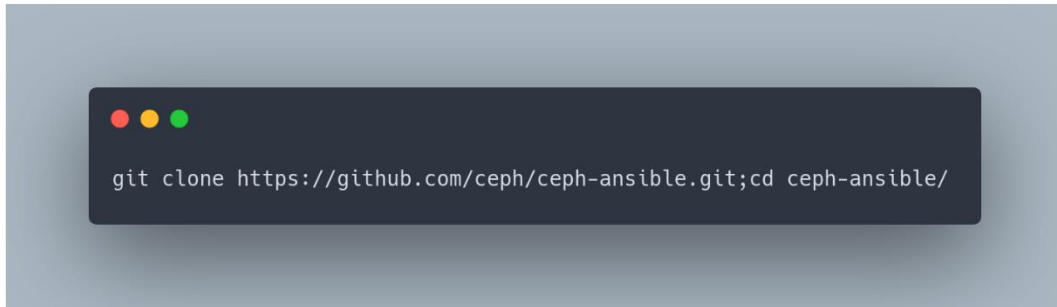
```
sudo apt update;sudo apt -y install python3-pip
```

2. Instalar python3-venv para crear entornos virtuales y evitar problemas entre ceph-ansible y kolla-ansible:

A terminal window with a dark background and three colored window control buttons (red, yellow, green) in the top left corner. The terminal displays the command: `sudo apt -y install python3-venv`

```
sudo apt -y install python3-venv
```

3. Clonar el repositorio de ceph-ansible (<https://github.com/ceph/ceph-ansible.git>) y cambiar al directorio:



```
git clone https://github.com/ceph/ceph-ansible.git;cd ceph-ansible/
```

4. Crear entorno virtual:



```
python3 -m venv ceph
```

5. Activar entorno virtual



```
source ceph/bin/activate
```

6. Para evitar errores en la instalación de requerimientos, instalar wheel:



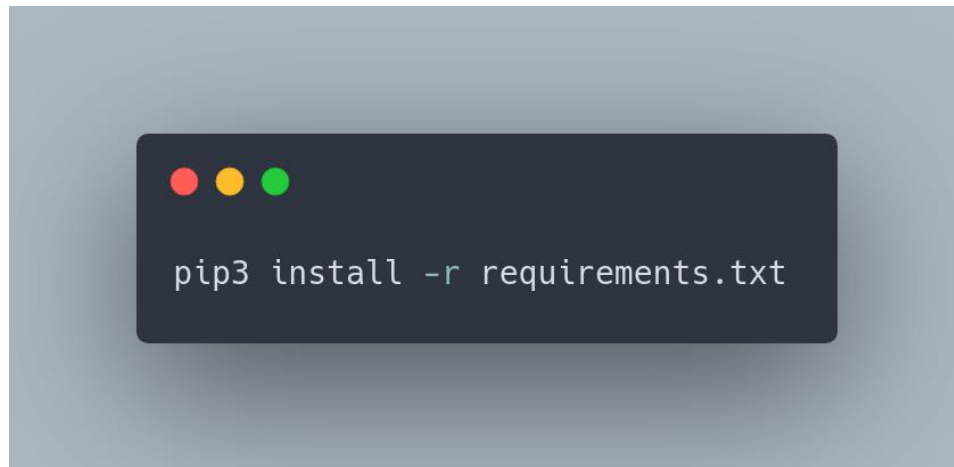
```
pip install wheel
```

7. Cambiar a branch stable-6.0



```
git checkout stable-6.0
```

8. Instalar requerimientos de ceph-ansible



```
pip3 install -r requirements.txt
```

9. Instalar Ansible



```
sudo add-apt-repository ppa:ansible/ansible  
sudo apt update  
sudo apt install ansible
```

10. Crear archivo inventory.ini con la configuración de acuerdo a la topología:

Importante especificar archivo de llave primaria para evitar errores, ya que el comando para empezar el deploy usa privilegios **sudo** y usar la llave localizada en */root/.ssh*

11. Copiar los archivos de configuración necesarios:

```
cativo@deployer:~/ceph-ansible$ mv group_vars/all.yml.sample group_vars/all.yml
cativo@deployer:~/ceph-ansible$ mv group_vars/osds.yml.sample group_vars/osds.yml
```

12. Editar el archivo group_vars/all.yml con las siguientes configuraciones:

```
dummy:
  cluster: ceph
  ntp_service_enabled: true
  ntp_daemon_type: chronyd
  ceph_origin: repository
  ceph_repository: community
  ceph_mirror: https://download.ceph.com
  ceph_stable_key: https://download.ceph.com/keys/release.asc
  ceph_stable_release: pacific
  ceph_stable_repo: "{{ ceph_mirror }}/debian-{{ ceph_stable_release }}"
  monitor_address_block: 192.168.122.0/24
  ip_version: ipv4
  public_network: 192.168.122.0/24
  cluster_network: 192.168.4.0/24
  radosgw_address_block: 192.168.122.0/24
  ceph_docker_image: "ceph/daemon"
  ceph_docker_image_tag: latest-pacific
  ceph_docker_registry: docker.io
  ceph_docker_registry_auth: false
  ceph_client_docker_image: "{{ ceph_docker_image }}"
  ceph_client_docker_image_tag: "{{ ceph_docker_image_tag }}"
  ceph_client_docker_registry: "{{ ceph_docker_registry }}"
  containerized_deployment: True
  dashboard_enabled: True
  dashboard_protocol: https
  dashboard_port: 8443
  dashboard_admin_user: admin
  dashboard_admin_user_ro: false
  dashboard_admin_password: p@ssw0rd
  grafana_admin_user: admin
  grafana_admin_password: admin
```

13. Editar el archivo group_vars/osds.yml con las siguientes configuraciones:

```
dummy:
copy_admin_key: true
devices:
- /dev/vdb
```

14. Hacer deploy de cluster ceph:

```
deployer-user@deployer:~/ceph-ansible$ ansible-playbook -i inventory.ini site-container.yml.sample
```

15. Verificar estado del clúster en el nodo controller:

```
cativo@controller:~$ sudo docker ps -a
CONTAINER ID   IMAGE                                COMMAND                  CREATED        STATUS        PORTS          NAMES
688296690e35   prom/prometheus:v2.7.2             "/bin/prometheus --c..." 3 minutes ago  Up 3 minutes  prometheus
b83e6a283ee2   prom/alertmanager:v0.16.2         "/bin/alertmanager -..." 3 minutes ago  Up 3 minutes  alertmanager
55e02b10118d   grafana/grafana:6.7.4              "/run.sh"               3 minutes ago  Up 3 minutes  grafana-server
2e64b283743a   ceph/daemon:latest-pacific         "/usr/bin/ceph-crash"    2 hours ago    Up 2 hours    ceph-crash-
controller
39d676cd30ff   prom/node-exporter:v0.17.0         "/bin/node_exporter ..." 2 hours ago    Up 2 hours    node-exporter
7a1a399c2e73   ceph/daemon:latest-pacific         "/opt/ceph-container..." 2 hours ago    Up 2 hours    ceph-mgr-
controller
0fdc41895076   ceph/daemon:latest-pacific         "/opt/ceph-container..." 2 hours ago    Up 2 hours    ceph-mon-
controller
cativo@controller:~$ sudo docker exec ceph-mon-controller ceph -s
cluster:
  id: a2f42f43-11e8-4e75-8d84-61b7f01bba25
  health: HEALTH_OK

services:
  mon: 1 daemons, quorum controller (age 2h)
  mgr: controller(active, since 2m)
  osd: 3 osds: 3 up (since 2h), 3 in (since 2h)
  rgw: 1 daemon active (1 hosts, 1 zones)

data:
  pools: 6 pools, 113 pgs
  objects: 207 objects, 6.2 KiB
  usage: 73 MiB used, 300 GiB / 300 GiB avail
  pgs: 113 active+clean
```

16. Importante revisar el nombre completo de la instancia de ceph-mon con:

docker ps -a

Verificar Funcionamiento

De acuerdo a lo definido en el archivo inventory.ini (paso x), se instaló el RADOS Gateway en el nodo **compute1** el cual nos proporciona la posibilidad de interactuar con el clúster por medio de una API REST compatible con AWS S3:

1. En el nodo **compute1**, crear usuario con credenciales para interactuar con el clúster, anotar **user**, **access_key** y **secret_key**:



```
cativo@compute1:~$ sudo docker exec -it ceph-rgw-compute1-rgw0 bash
[root@compute1 /]# sudo radosgw-admin user create --uid="cativo" --display-name="cativoS3"
{
  "user_id": "cativo",
  "display_name": "cativoS3",
  "email": "",
  "suspended": 0,
  "max_buckets": 1000,
  "subusers": [],
  "keys": [
    {
      "user": "cativo",
      "access_key": "QLV4AP0YIKZPUP191NF8",
      "secret_key": "W4TuCjDu4jDP09fgadCDAFiazbIffhsb14JeBpkv"
    }
  ],
  "swift_keys": [],
  "caps": [],
  "op_mask": "read, write, delete",
  "default_placement": "",
  "default_storage_class": "",
  "placement_tags": [],
  "bucket_quota": {
    "enabled": false,
    "check_on_raw": false,
    "max_size": -1,
    "max_size_kb": 0,
    "max_objects": -1
  },
  "user_quota": {
    "enabled": false,
    "check_on_raw": false,
    "max_size": -1,
    "max_size_kb": 0,
    "max_objects": -1
  },
  "temp_url_keys": [],
  "type": "rgw",
  "mfa_ids": []
}
```


2. En la maquina **host** con la herramienta de comando de AWS configuramos para que tome las credenciales:

```
>>> ~ aws configure --profile=ceph
AWS Access Key ID [None]: QLV4APOYIKZPUP191NF8
AWS Secret Access Key [None]: W4TuCjDu4jDP09fgadCDAFiazbIffhsb14JeBpkv
Default region name [None]:
Default output format [None]: json
```

3. Luego probando creando un bucket

```
>>> ~ aws --profile=ceph --endpoint=http://192.168.122.5:8080 s3 mb s3://test
make_bucket: test
```

4. Listamos todos los buckets

```
>>> ~ aws --profile=ceph --endpoint=http://192.168.122.5:8080 s3api list-buckets
{
  "Buckets": [
    {
      "Name": "test",
      "CreationDate": "2021-09-13T08:47:27.207Z"
    }
  ],
  "Owner": {
    "DisplayName": "cativoS3",
    "ID": "cativo"
  }
}
```

Anexo 6. Comparativo de Tecnologías Compatibles con Ceph

Ceph

Ceph es una solución de almacenamiento escalable que replica los datos en nodos de almacenamiento básicos. Hace uso de un mecanismo de almacenamiento de objetos para almacenar los datos y los expone a través de diferentes tipos de interfaces de almacenamiento para el usuario final:

- Almacenamiento de objetos
- Almacenamiento de bloques
- Interfaces de sistemas de archivos

Ceph ofrece soporte para la misma API de almacenamiento de objetos que swift y puede utilizarse como back-end para el servicio de almacenamiento en bloque (cinder), así como para el almacenamiento back-end para las imágenes de glance.

Ceph admite el aprovisionamiento ligero implementado mediante la copia en escritura. Esto puede ser útil cuando se arranca desde un volumen, porque un nuevo volumen puede ser aprovisionado muy rápidamente. Ceph también es compatible con la autenticación usando keystone, por lo que puede sustituir sin problemas a la implementación predeterminada de OpenStack swift.

Las ventajas de Ceph incluyen:

- El administrador tiene mayor control sobre la distribución de datos y las estrategias de replicación.
- Consolidación del almacenamiento de objetos y del almacenamiento de bloques.
- Rápido aprovisionamiento de instancias de arranque desde volumen mediante aprovisionamiento ligero.
- Compatibilidad con la interfaz del sistema de archivos distribuidos CephFS.
- Se debe considerar Ceph si desea gestionar su almacenamiento de objetos y de bloques dentro de un único sistema, o si desea soportar un rápido arranque desde volumen.

Gluster

Un sistema de archivos compartidos distribuidos. A partir de la versión 3.3 de Gluster, puede ser utilizado para consolidar su almacenamiento de objetos y archivos en una solución unificada de almacenamiento de archivos y objetos, que se denomina Gluster For OpenStack (GFO). GFO utiliza una versión personalizada de swift que permite utilizar Gluster como almacenamiento back-end.

La razón principal para utilizar GFO en lugar de swift es si también desea soportar un sistema de archivos distribuido, ya sea para apoyar la migración en vivo de almacenamiento compartido o para ofrecerlo como un servicio independiente a sus usuarios finales. Si desea gestionar su almacenamiento de objetos y archivos dentro de un único sistema, debería considerar GFO.

LVM

El gestor de volúmenes lógicos (LVM) es un sistema basado en Linux que proporciona una capa de abstracción sobre los discos físicos para exponer los volúmenes lógicos al sistema operativo. El back-end de LVM implementa el almacenamiento en bloque como particiones lógicas LVM.

En cada host que albergará el almacenamiento en bloque, un administrador debe crear inicialmente un grupo de volúmenes dedicado a los volúmenes de almacenamiento en bloque. Los bloques se crean a partir de volúmenes lógicos LVM.

Una desventaja de LVM es que no proporciona ninguna replicación. Normalmente, se debe configurar RAID en los nodos que utilizan LVM como almacenamiento en bloque para protegerse contra los fallos de los discos duros individuales. Sin embargo, RAID no protege contra un fallo de todo el host.

iSCSI

Internet Small Computer Systems Interface (iSCSI) es un protocolo de red que opera sobre el Protocolo de Control de Transporte (TCP) para enlazar dispositivos de almacenamiento de datos. Transporta datos entre un iniciador iSCSI en un servidor y un objetivo iSCSI en un dispositivo de almacenamiento.

iSCSI es adecuado para entornos de nube con servicio de almacenamiento en bloque para soportar aplicaciones o para sistemas de intercambio de archivos. La conectividad de red puede conseguirse a un coste menor en comparación con otras tecnologías de back-end de almacenamiento, ya que iSCSI no requiere adaptadores de bus de host (HBA) ni dispositivos de red específicos para el almacenamiento.

NFS

Network File System (NFS) es un protocolo de sistema de archivos que permite a un usuario o administrador montar un sistema de archivos en un servidor. Los clientes de archivos pueden acceder a los sistemas de archivos montados a través de llamadas a procedimientos remotos (RPC).

Las ventajas de NFS son el bajo coste de implementación debido a que se comparten NICs y componentes de red tradicionales, y un proceso de configuración y puesta en marcha más sencillo.

Para obtener más información sobre la configuración del almacenamiento en bloque para utilizar el almacenamiento NFS, consulte Configurar un back end de almacenamiento NFS en la Guía del administrador de OpenStack.

Sheepdog

Sheepdog es un sistema de almacenamiento distribuido de espacio de usuario. Sheepdog escala a varios cientos de nodos, y tiene potentes características de gestión de discos virtuales como snapshot, clonación, rollback y aprovisionamiento ligero.

Es esencialmente un sistema de almacenamiento de objetos que gestiona los discos y agrega el espacio y el rendimiento de los discos de forma lineal en hiperescala en hardware básico

de forma inteligente. Además de su almacén de objetos, Sheepdog proporciona un servicio de volumen elástico y un servicio http. Sheepdog requiere una versión específica del kernel y puede funcionar bien con sistemas de archivos compatibles con xattr.

ZFS

El controlador Solaris iSCSI para OpenStack Block Storage implementa los bloques como entidades ZFS. ZFS es un sistema de archivos que también tiene la funcionalidad de un gestor de volúmenes. Esto es diferente a lo que ocurre en un sistema Linux, donde hay una separación entre el gestor de volúmenes (LVM) y el sistema de archivos (como, ext3, ext4, xfs y btrfs). ZFS tiene una serie de ventajas sobre ext4, incluyendo una mejor comprobación de la integridad de los datos.

El back-end de ZFS para OpenStack Block Storage sólo es compatible con sistemas basados en Solaris, como Illumos. Aunque hay una versión Linux de ZFS, no está incluido en ninguna de las distribuciones Linux estándar, y no ha sido probado con OpenStack Block Storage. Al igual que con LVM, ZFS no proporciona la replicación a través de los hosts por sí mismo, es necesario añadir una solución de replicación en la parte superior de ZFS si su nube tiene que ser capaz de manejar los fallos de los nodos de almacenamiento.

Anexo 7. Comparativo de Soluciones de Infraestructura como Servicio

Tabla 8. Tabla Comparativa de Plataformas de Infraestructura como Servicio

Características	OpenStack	CloudStack	Eucalyptus	OpenNebula	Nimbus
Año de Lanzamiento	2010	2010	2008	2008	2009
Desarrollado por	Rackspace, NASA, Dell, Citrix, Cisco, Canonical etc.	Cloud.com	Santa Barbara university, Eucalyptus System Compan	European Union	University of Chicago
Filosofía	Ofrecer Servicios de Computación en la Nube		Imitar a Amazon EC2	Nube privada altamente Personalizable	Nube a la medida a investigadores científicos.
Adecuado para	Empresas, proveedores de servicios e investigadores	Empresas, proveedores de servicios e investigadores	Grandes empresas comerciales, instituciones de investigación	Grandes empresas comerciales e instituciones públicas.	Instituciones de investigación.
Arquitectura	Integración de varios componentes de Openstack.	Jerárquico con cuatro componentes principales: - Servidor de Gestión, - Zona de Disponibilidad,	Agrupados jerárquicamente desde el CLC, pasando por el CC, hasta el CN: - Jerárquico - Cinco componentes - Mínimo dos	Tres módulos contienen todos los componentes: - Centralizado - Tres componentes - Mínimo dos servidores	Tres módulos contienen todos los componentes: - Centralizado

		- Pod, - Nodos informáticos	servidores		- Tres componente s - Mínimo dos servidores
Soporte de la API	Native API, Amazon EC2 API, CloudFiles REST API.	Amazon EC2 API, S3.	Amazon EC2 API.	Native API in Ruby and JAVA. XML- RPC API for interfaces creation. OGF OCCI & Amazon EC2 APIs.	EC2 APIs, S3 APIs, JAVA client APIs.
Soporte de Amazon	EC2, S3.	EC2, S3.	EC2, S3, EBS, IAM, AMI.	EC2, EBS, AMI.	EC2, S3.
Tipo de Nube	Publica, privada e Híbrida	Publica, privada e Híbrida	Privada e Híbrida	Privada e Híbrida	Privada
Hypervisor	KVM, Xen, VMware ESX, ESXi, Hyper-v, LXC, QEMU, UML, PowerVM, Bare metal	VMware, Oracle VM, KVM, XEN	KVM, Xen, VMware	e KVM, Xen, VMware ESX, ESXi	Python, Bash, Ebttables, Libvirt, KVM, Xen
Lenguaje de Programación	Python	Java	Java, C, Python	Java, Ruby, C++	Java, Python

Comunidad	+++++	++++	+++	+++	++
Frecuencia de Lanzamiento	< 4 meses	4 meses	>4meses	>6meses	<4meses
Facilidad de Uso	+++++	+++++	++	+++	+++
Sistemas Operativos Soportados	Linux, Windows	Depende del Hypervisor and hardware - Mac OS X, Asianux, CentOS, Debian, DOS, Fedora, FreeBSD, Novell Netware, Oracle Enterprise Linux, Ubuntu, Red Hat Enterprise Linux, Sun Solaris, SUSE Linex Enterprise, Window	Linux (Ubuntu, Fedora, CentOS, OpenSUSE et Debian)	CentOS, Debian, Fedora, RHEL open-SUSE, SLES, and Ubuntu.	La mayoría de las distribuciones de Linux
Almacenamiento	Se admite el almacenamiento de objetos y de bloques. Los volúmenes son persistentes (los	Soporte para iSCSI, NFS, SMB/CIFS; compatibilidad con	Soporte para iSCSI, EBS, Amazon S3. Soporte de hardware para el	Soporte de hardware para Fibre Canal, iSCSI, NAS almacenamiento o compartido,	

	datos se conservan hasta que el volumen se volumen, independientemente de la máquina virtual). El almacenamiento de archivos es soportado a través de Swift u otros como Ceph	OpenStack Swift y Amazon S3	almacenamiento estándar de la industria.	SCSI / SAS / SATA. Sistemas de archivos compartidos y no compartidos (NFS, LVM con CoW, VMFS, etc.).	
Red	VLAN, NO VLAN, IPs públicas, IPs privadas, SDN, IDS, Load-balance, Firewalls, VPN, etc.	VLAN, IPs Públicas	VLAN, NO VLAN IPs públicas, IPs privadas; servidor DHCP en el controlador del clúster	VLAN, NO VLAN, IP's públicas, IP's privadas, Ebttables, OVSswitch; Configuración manual	Servidor DHCP instalado en nodos
Interfaz de Usuario	Interfaz web (es decir Dashboard) y la interfaz de línea de comandos de línea de comandos para desplegar VMs y una consola para	Interfaz web e Interfaz de línea de comandos (CLI)	euca2ools (CLI)	Interfaz web e Interfaz de línea de comandos (CLI)	Servicios web, específicamente: Nimbus Web

	gestionar las máquinas virtuales.				
Seguridad	API incluye protección contra ataques DoS o clientes defectuosos. El concepto de proyecto es introducido por Nova, permitiendo a los administradores gestionar otras cuentas de usuario y los recursos del proyecto. Keystone se utiliza para la gestión de la identidad.	CloudStack Security Groups	El controlador de la nube genera una clave pública/privada para la autenticación del usuario autenticación del usuario	Autenticación por contraseñas, shell seguro y pares de claves RSA Lightweight Directory Access Protocol ; Marco de autenticación basado en contraseñas, pares de claves RSA SSH o LDAP. Varios roles de administración . Multi-tenancy para nubes públicas.	Infraestructura de clave pública
Robustez ante Errores	Replicación	Replicación	Las agrupaciones separadas reducen la probabilidad de	Base de datos permanente para almacenar	Comprobación periódica y copia de seguridad de

			que se produzcan errores o relacionados con ellas; la separación del controlador de la agrupación	información sobre hosts, redes y máquinas virtuales máquinas virtuales; Base de datos backend (registra la información de información de las máquinas virtuales)	los nodos de los trabajadores ; Verificación verificación de los nodos nodos de la nube.
Balanceador de Carga	El controlador de la nube	Equilibrador de carga TCP	El controlador de la nube	Nginx	Le Context Broker
Licenciamiento	ApacheLicence Version2	ApacheLicence Version2	BSD-Licence	ApacheLicence Version2	ApacheLicence Version2
Documentación	+++++	+++++	+++	+++	++

Fuente: *A Comparative Study of Current Open-source Infrastructure as a Service Frameworks*

(<https://pdfs.semanticscholar.org/10f6/0c8f67621fcd90e90c9f1257dd8418f34a33.pdf>)

Anexo 8. Comparativo de nube pública vs Privada

Tabla 9. Tabla Comparativa de Nube Pública vs Nube Privada

	Nube pública	Nube privada
Costo	No se invierte en centros de datos físicos ni hardware especializado, la inversión inicial es menor, aunque a largo plazo represente un mayor costo.	Representa una mayor inversión inicial si no se cuenta con los recursos necesarios, aunque a largo plazo puede representar un ahorro.
Mantenimiento	Está en manos de los proveedores, no se necesita de personal TI especializado.	Se necesita de algún personal TI especializado.
Seguridad	Se depende en gran manera de la seguridad que posee el proveedor, dejando a cargo todo tipo de datos, incluyendo datos sensibles.	El cliente de la nube es el encargado de la seguridad y de garantizar.
Privacidad	Los datos son resguardados por el proveedor y salen del establecimiento físico del cliente de la nube.	El cliente de la nube tiene el control total de los datos que maneja.
Rendimiento	No hay lugar para personalizar soluciones en una nube pública, por eso es tan barato. Una nube pública puede resultar inadecuada para empresas con arquitectura o procesos complejos.	El cliente de la nube puede elegir el hardware en el que se ejecutará su infraestructura, de modo que pueda influir en el rendimiento.
Flexibilidad	Muy a menudo los proveedores no otorgan a los clientes la libertad de instalar los sistemas operativos de su elección. Esa es una de las principales razones por las que una nube pública no se ajusta a las necesidades de las empresas con regulaciones de cumplimiento.	Se pueden configurar para cumplir con precisión las expectativas de su organización.

Fuente: *The right kind of cloud? How to choose the cloud services that suit your company needs.* (<https://dsstream.com/how-to-choose-cloud-services>)