



**UNIVERSIDAD ESTATAL
PENÍNSULA DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES
CARRERA DE TECNOLOGÍAS DE LA INFORMACIÓN**

MODALIDAD: EXAMEN COMPLEXIVO

Componente Práctico, previo a la obtención del Título de:

**INGENIERA EN TECNOLOGÍAS
DE LA INFORMACIÓN**

TEMA:

**“PROPUESTA DE UNA GUÍA DE ADOPCIÓN DE UN DATA LAKE
ENMARCANDO LOS ASPECTOS IMPORTANTES DE CLOUD
COMPUTING. CASO DE ESTUDIO: FEDERACIÓN DE
ORGANIZACIONES POPULARES PENINSULARES (FOPOPE).”**

AUTOR:

BALÓN DE LA CRUZ MAYRA LILIBETH

**LA LIBERTAD – ECUADOR
PAO 2022-1**

DEDICATORIA

Recordemos que la investigación es una tarea en equipo, dentro de mi recorrido estudiantil entre batallas por cualquier cuestión siempre llegaron momentos en que la lucha cesó y se realizó una tregua para poder lograr metas conjuntas, por esa razón con mi corazón lleno de regocijo de amor, felicidad y esperanza dedico a toda persona que me brindó su apoyo físicamente y espiritualmente, especialmente a mi ángel del cielo que con sus recuerdos me motivaron a seguir adelante , a mi padre y mis hermanos por ser el motivo fundamental de crecer cada día, a mi novio por brindarme sus enseñanzas y compañía , y a mi familia por creer en mí.

Mayra Balón

AGRADECIMIENTO

Expreso mi agradecimiento y gratitud a Dios por bendecir y guiar mi vida, por ser ese apoyo y fortaleza en momentos de alegría y de dificultad.

En el transcurso de la vida encuentras personas que iluminan tu vida, sin lugar a duda agradezco a toda mi familia, a mi compañero de vida y colegas que a través de consejos, apoyo y paciencia me ayudaron a concluir esta meta.

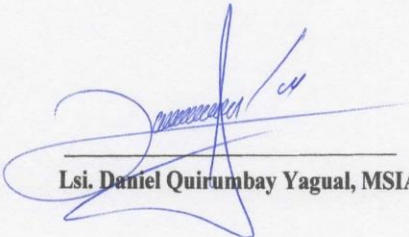
A mi tutor por brindarme sus enseñanzas, por guiarme, y no ser sólo mi tutor sino también un gran amigo.

A la Universidad Estatal Península de Santa Elena que en mi trayecto universitario se transformó en un segundo hogar, sin dejar atrás a quienes conforman la carrera de Tecnologías De La Información por la orientación y por permitirme concluir una etapa más de mi vida.

Mayra Balón

APROBACIÓN DEL TUTOR

En mi calidad de tutor del trabajo de componente práctico del examen de carácter Complexivo: **“Propuesta de una guía de adopción de un Data Lake enmarcando los aspectos importantes de Cloud Computing. Caso de estudio: Federación de Organizaciones Populares Peninsulares (FOPOPE)”**, elaborado por la estudiante Balón De La Cruz Mayra Lilibeth de la carrera de Tecnología de la Información, de la Universidad Estatal Península de Santa Elena, me permito declarar que luego de haber orientado, estudiado y revisado, la apruebo en todas sus partes.



Lsi. Daniel Quirumbay Yagual, MSIA

La Libertad, 16 de agosto del 2022

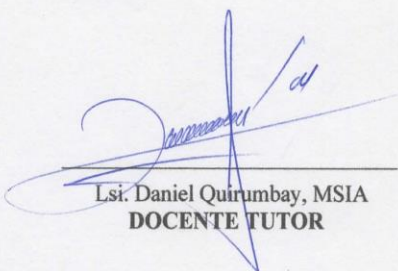
TRIBUNAL DE GRADO



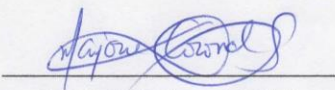
**Ing. Jaime Orozco I, MGT
DIRECTOR DE CARRERA**



**Ing. Iván Coronel S, MSIA
DOCENTE ESPECIALISTA**



**Lsi. Daniel Quirumbay, MSIA
DOCENTE TUTOR**



**Ing. Marjorie Coronel, MGTI
DOCENTE GUÍA UIC**

RESUMEN

La presente propuesta tecnológica estableció como finalidad la elaboración de una guía de adopción de un Data Lake enmarcando los aspectos importantes de Cloud Computing en la Federación de Organizaciones Populares Peninsulares (FOPOPE), debido a que la Federación y sus Organizaciones filiales efectuaban un el proceso dentro de la organización en el departamento de archivística y sus respectivos almacenamiento de datos de manera local, lo que causaba que al momento de analizar dichas bases de datos estructuradas así mismo no estructuradas que maneja la institución se tornaba difícil obtener reportes esenciales para una buena toma de decisiones ya que estas no le permitían visualizar de manera ágil, rápida y sencilla sus datos que por ende no brindaba soporte y control a los directivos, llevándolos a un manejo inapropiado de datos, incluso pérdidas de información relevantes e importantes para el correcto funcionamiento

Es por ello para lograr este objetivo y tener conocimiento de cómo viabilizar un estudio para el servicio de Cloud Computing de Data Lake se utilizó un conjunto de Servicios de diferentes proveedores y a su vez para la realizar una comparación entre las mismas para dar solución al problema en la federación. Se utilizó Microsoft Azure, como línea base Data Lake Storage Gen 2 así mismo Data Factory para la debida orquestación de almacenamiento de datos, Azure Data Bricks para el respectivo análisis de datos y Power Bi para su respectiva visualización. Como segundo proveedor se utilizó a Google Cloud Platform, como línea base Google Cloud Storage, para la correcta orquestación de datos se utilizó el servicio de Data Cloud Composer y Cloud Data Proc para el análisis. Como tercer proveedor se utilizó Amazon Web Service (AWS), Como línea base se utilizó a Data Lake S3, AWS Glue Con Data Pipeline para su respectiva orquestación de datos. Además, se consideró como base la metodología del Instituto Nacional de Estándares y Tecnología (NIST), ya que este enfoque se basa en la implementación de una guía para realizar y para personas que deseen adquirir dicho servicio que quiera en la nube, Así mismo la metodología está basada en 4 fases esta permite realizar un proceso de establecer el nivel de amenaza viable y sus respectivos riesgos con un sistema TIC.

Palabras claves: Análisis de datos, Almacenamiento en la nube, Cloud Computing, Data lake,

ABSTRACT

The present technological proposal established as its purpose the elaboration of a data lake adoption guide framing the important aspects of Cloud Computing in the federation of popular peninsular organizations (FOPOPE), due to the fact that the Federation and its affiliated organizations carried out the archiving and data storage process locally, which meant that when analyzing its structured databases and spreadsheets managed by the institution, it became difficult to obtain essential reports for good decision-making. decisions since these did not allow him to visualize his data in an agile, fast and simple way, which therefore did not provide support and control to managers, leading them to inappropriate data management, including loss of relevant and important information for organizational functioning.

That is why, in order to achieve this objective and to have knowledge of how to make a study for the Cloud Computing service in a Data Lake feasible, a set of Services from different providers was used and, in turn, to make a comparison between them to solve the problem. in the federation. Microsoft Azure was used as a Data Lake Storage Gen 2 baseline, Data Factory for proper data storage ingestion, Azure Data Bricks for data analysis and Power Bi for the respective visualization. Google Cloud Platform was used as a second provider, Google Cloud Storage as a baseline, the Data Cloud Composer service and Cloud Data Proc were used for data ingestion for analysis. Amazon Web Service (AWS) was used as a third provider, Data Lake S3 was used as a baseline, AWS Glue with Data Pipeline for its respective data ingestion.

In addition, the methodology of the National Institute of Standards and Technology (NIST) was considered as a basis, since this approach is based on an implementation guide for the entrepreneur who wants to acquire services in the cloud. Likewise, since it is based on 4 phases that in the generation of a Data Lake according to the guide, it allows the process of determining the level of potential threat and its risks associated with an ICT system.

Keywords: Data Analytics, Cloud Storage, Cloud Computing, Data Lake,

DECLARACIÓN

El contenido del presente trabajo de graduación es de mi responsabilidad; el patrimonio intelectual del mismo pertenece a la Universidad Estatal Península de Santa Elena.



MAYRA LILIBETH BALÓN DE LA CRUZ

TABLA DE CONTENIDO

AGRADECIMIENTO	II
APROBACIÓN DEL TUTOR	III
TRIBUNAL DE GRADO	IV
RESUMEN	IV
RESUMEN	V
ABSTRACT	VI
DECLARACIÓN	VII
INTRODUCCIÓN	XII
CAPÍTULO I	1
1. FUNDAMENTACIÓN	1
1.1. ANTECEDENTES	1
1.2. DESCRIPCIÓN DEL PROYECTO	4
1.3. OBJETIVOS DEL PROYECTO	7
1.3.1. OBJETIVO GENERAL	7
1.3.2. OBJETIVOS ESPECÍFICOS	7
1.5. ALCANCE DEL PROYECTO	10
1.6. METODOLOGÍA	13
1.6.1. METODOLOGÍA DE LA INVESTIGACIÓN	13
1.6.2. TÉCNICA DE INVESTIGACIÓN	14
1.6.3. METODOLOGÍA DE DESARROLLO	16
CAPÍTULO II	20
2.1.1 Amazon Web Services AWS	20
2.1.2 Google Cloud Platform GCP	20
2.1.3. Azure	20
2.1.4. Data Lake	20
2.1.5. Definición de Cloud Computing	20
2.1.6. TIPOS DE DATA LAKE	21
2.2. MARCO TEÓRICO	27
2.2.1. Agentes impulsores de la adopción de Cloud Computing en las empresas. ¿Quién mueve la nube?	27
2.2.2. Data Lake para la Dirección de Análisis Universitaria: Arquitectura y Metadata.	27
2.2.3. Data Lake: Aplicaciones, Herramientas y Arquitecturas.	28
2.3. COMPONENTES DE LA PROPUESTA	28
2.5. DISEÑO DE LA PROPUESTA	46

3.1.- RESULTADOS	60
CONCLUSIONES.	91
RECOMENDACIONES.	92
BIBLIOGRAFÍA	93
ANEXOS	96

ÍNDICE DE FIGURA

Figura 1: Fases adaptadas de la guía de NIST	17
Figura 2:Infraestructura de Data Lake Azure	21
Figura 3:Infraestructura Data Lake AWS	22
Figura 4:Infraestructura de Data Lake en Google	23
Figura 5:Infraestructura en Microsoft Azure	24
Figura 6: Infraestructura en AWS	25
Figura 7: Infraestructura en Google	25
Figura 8: Análisis de la factibilidad de la adopción de Data Lake en la Federación – Power-Bi desktop	30
Figura 9:Diseño del enlace del servidor local con el Cloud Computing	47
Figura 10: Propuesta de solución con Cloud Computing	48
Figura 11:Página principal Portal Azure	61
Figura 12: Visualización Portal	61
Figura 13: Creación de grupos en el portal de Azure	62
Figura 14: Visualización de grupos en el portal de Azure.	62
Figura 15: Creación de cuenta de almacenamiento en el Portal Azure.	63
Figura 16: Visualización Portal	63
Figura 17: Visualización cuenta de almacenamiento en el Portal	64
Figura 18: Creación Contenedores Portal	64
Figura 19: Visualización de contenedores en el Portal de Azure	65
Figura 20: Visualización del explorador de almacenamiento	65
Figura 21: Generación de firmas de acceso SaaS Portal Azure	66
Figura 22: Generación de tokens de acceso compartido Portal Azure	66
Figura 23: Generación de accesos y permisos ACL en el Portal Azure	67
Figura 24: Pagina de Google Cloud Platform	67
Figura 25: Registro de la Pagina de Google Cloud Platform	68
Figura 26: Panel principal de GCP	68
Figura 27: Panel de Google Cloud Storage	69
Figura 28: Creación de depósito en Google Cloud Storage	69
Figura 29: Configuración de depósito en Google Cloud Storage	70
Figura 30: Configuración de clase de almacenamiento en Google Cloud Storage	70
Figura 31: Configuración de control de acceso de objetos en Google Cloud Storage	71
Figura 32: Configuración de encriptación en Google Cloud Storage	71
Figura 33: Visualización del Bucket creado en Google Cloud Storage	72
Figura 34: Almacenamiento de datos dentro del Bucket creado en Google Cloud Storage	72
Figura 35: Configuración de permisos para el Bucket creado en Google Cloud Storage	72

Figura 36: Configuración de la protección para el Bucket creado en Google Cloud Storage	73
Figura 37: Página Principal de AWS	74
Figura 38: Creación de Bucket dentro de la plataforma	74
Figura 39: Panel de configuración de objetos de AWS	75
Figura 40: Panel de configuración de bloqueo de acceso público para Bucket	75
Figura 41: Panel de configuración y control de versiones de Bucket	76
Figura 42: Configuración en el panel de cifrados predeterminados	76
Figura 43: Visualización de la creación de Bucket	77
Figura 44: Conocimiento Del Data Lake	81
Figura 45: Conocimiento de los servicios del Data Lake	82
Figura 46: Despliegue de un Data Lake en la Organización	83
Figura 47: Motivación a la adopción de un Data Lake	84
Figura 48: Uso de nuevas tecnologías para ahorro de tiempo y dinero	85
Figura 49: Motivación a la adopción de un Data Lake	86
Figura 50: Beneficios del Data Lake	87
Figura 51: Área de la organización a implementar estas herramientas.	88
Figura 52: Capacitación a las organizaciones.	89
Figura 53: Frecuencia de la Capacitación a las Organizaciones.	90
Figura 54: Entrevista al presidente de la Federación FOPOPE.	97
Figura 55: Encuesta con Dirigentes de las organizaciones.	97
Figura 56: Socialización del proyecto a los directivos y de la organización que conforman la Federación de Organizaciones Populares Provinciales FOPOPE	97
Figura 57: Encuesta a directivas de la Federación FOPOPE.	97
Figura 58: Presentación de la simulación en el Portal de Azure	97
Figura 59: Presentación de la Visualización el Power BI	97

ÍNDICE DE TABLAS

Tabla 1: Características de computación en la nube	21
Tabla 2: Comparativa de los servicios de almacenamientos aplicando el pago de los servicios en los diferentes proveedores en la nube.	23
Tabla 3: Comparación rápida entre Azure Data Lake Gen1 y Azure Data Lake Gen2	26
Tabla 4: Caracterización Del Departamento De Administración.	31
Tabla 5: Caracterización Del Departamento de Archivística.	31
Tabla 6: Caracterización Del Departamento De TIC	33
Tabla 7: Infraestructura en hardware de la federación en la actualidad	34
Tabla 8: Infraestructura en software de la federación en la actualidad	35
Tabla 9: Escenario Cloud en base confidencialidad de la información	40
Tabla 10: Escenario Cloud en base al tipo de dato a tratar	40
Tabla 11: Escenario Cloud en base al foco del sistema a utilizar	41
Tabla 12: Comparativa entre Plataformas Cloud.	42
Tabla 13: Herramientas utilizadas en la infraestructura local	46
Tabla 14: Herramientas de la infraestructura de solución	48
Tabla 15: Infraestructura Requerimiento	49
Tabla 16: Arquitectura Lógica.	50

Tabla 17 Arquitectura Física	50
Tabla 18: Componentes de red	50
Tabla 19: Coto de Operación.	51
Tabla 20: Manteniendo Hardware y Software	51
Tabla 22: Factibilidad técnica sistema archivístico	52
Tabla 22: Factibilidad del servicio del portal web	52
Tabla 23: Factibilidad del servicio del correo electrónico	52
Tabla 24: Encuesta de aceptación del proyecto	53
Tabla 25: Costo de salario personal operativo	55
Tabla 26: Costo de instalación y mantenimiento	55
Tabla 27: Costo de instalación y mantenimiento	55
Tabla 28: Costo suplementario	56
Tabla 29: Costo de operación anual	56
Tabla 30: Flujo de caja infraestructura federación	57
Tabla 31: Flujo de caja solución en la nube	57
Tabla 32: Azure calculadora de precios.	58
Tabla 33: Costo total solución en la nube.	59
Tabla 34: Calculo TIR	60
Tabla 35: Costo	78
Tabla 36: Accesibilidad	78
Tabla 37: Servicios de Google Cloud, Azure y AWS	79
Tabla 38: Comparación entre los servicios que ofrece Servicios de Google Cloud Storage, Data lake Azure y AWS	80

LISTA DE ANEXOS

Anexos 1: Entrevista al presidente de la Federación de Organizaciones Populares Penínsulares (FOPOPE).	97
Anexos 2: Formato de encuesta - Federación de Organizaciones Populares Penínsulares (FOPOPE).	97

INTRODUCCIÓN

La Tecnología y sus avances de los últimos años hasta fecha actual se ha transformado en el almacenamiento de muchos volúmenes de datos y es por esto que se ha convertido en una gran problemática para diversas maquinarias en las empresas y organizaciones, para realizar el análisis de una gran cantidad de datos de una entidad condesciende a las organizaciones dominar y sobre todo poder realizar la reducción de sus costes y de esa manera obtener una mayor organización en sus datos.

Por llevar un mal manejo de datos tanto estructurados y no estructurados y a su vez realizando las hojas de cálculos que tiene la empresa respaldados en la nube contando también con el mal control para el manejo que con llevan a obtener pérdidas de diferentes tipos de información relevantes e importantes para el funcionamiento de la organización se realiza un estudio de factibilidad para dar solución y a su vez la utilización de sus datos para una posterior toma de decisión.

El propósito de dicha investigación es poder proporcionar a la Federación de Organizaciones Populares Peninsulares (FOPOPE), una guía de adopción de un data lake enmarcando los aspectos importantes de Cloud Computing, que impulse a una buena organización de sus datos, teniendo un mejor control en la salida o mal uso de información y la conservación de sus datos dentro sus tres departamentos, archivística, gerencia y TICS para obtener y realizar una excelente toma de decisiones en bien de la organización.

CAPÍTULO I

1. FUNDAMENTACIÓN

1.1. ANTECEDENTES

Con el acrecentamiento de investigación y la información en las empresas, y sobre todo la ciencia y el mismo comercio se retorna la penuria de realizar un gran cambio con relación a la debida conducción y manejo de diferentes informaciones, que hace muchos años abajo se venía efectuando en analogía a las bases de datos que se usan para la exploración de información solicitada así mismo se retorna gastador y caro o costoso al minuto de estudiar un sin número de volumen de información, en ambientes donde cada año y anualmente se almacenan datos a gran volumen [1].

Habitualmente que en muchas empresas la buena y perfecta toma de decisiones de los respectivos directores es vista que no es apoyada con respecto al tema de datos. En otros casos similar este tipo de información se ven en aglomeración con la respectiva simulación sin saber qué hacer con aquello ya que es de mucha importancia tener un excelente trabajo para una respectiva y buena organización si se habla de una empresa, en esta investigación y trabajo para la obtención de una aprovechamiento con grandes volúmenes de información se logra crear un buen lineamiento ya la magnitud y complejidad que tienen estos datos se han aprovechado de diferentes tecnologías que no son fundadas y legales. [2].

Es por este motivo que las diferentes empresas y las organizaciones que sobre todo muchas personas que se dan utilidad a la misma se han enfrentado a muchos desafíos y que tienen la debida información y conocimiento el cual ha sido de mucho esfuerzo y dedicación para alcanzar las mismas, por lo consiguiente para el correcto manejo de estas magnos conjuntos de información se hace guiar para la elaboración de diversas herramientas para la correcta administración y gestión de diferentes tipos de datos estos ya sean estos estructurados y no estructurados incluso hojas de cálculo para así obtener una buena toma de decisiones. En la actualidad con estas nuevas implementaciones ya dadas se han encontrados diversas tecnologías como lo concierne Hadoop por consiguiente se obtiene a los servicios de Data Lake, y bases de datos NoSQL, ya que las

mismas se utilizan para una implementación como lo es el BIG DATA [3].

La Federación de Organizaciones Populares Peninsular de la Provincia de Santa Elena. Se fundó el 26 de noviembre de 2.005 la Federación de Organizaciones Populares Peninsular de la Provincia de Santa Elena que tiene como objeto el estudio, el mejoramiento, la defensa y la búsqueda del bien común de sus miembros filiales, el bien público de diferentes sectores en general o de una humanidad en privativa, consta de 30 filiales en toda la provincia, las mismas que manejan un magno de conjuntos de información totalmente relevante para las organizaciones y estas mismas son manejada en forma manual y en bases de datos estructuradas almacenadas en servidores locales, lo que está expuesta a posibles pérdidas de información y obtener una mal manejo de los datos a nivel general.

Mediante entrevista realizada al presidente de la Federación de Organizaciones Populares Peninsulares de la Provincia de Santa Elena FOPOPE ([ver anexo 1](#)) se obtuvo que la federación y sus organizaciones filiales realizan en este año y los consiguientes el proceso de almacenar información en el departamento de archivística de manera local mediante bases de datos estructuradas y hojas de cálculos en Excel, esto a su vez al momento de analizar datos y obtener reportes esenciales para la toma de decisiones, es complicado ya que las mismas no le permite visualizar de manera ágil y sencilla, que brinde soporte y control a los directivos de dicha federación, llevándolos a un manejo inapropiado de datos el cual como consecuencia se obtiene pérdidas de información relevantes y la mala organización dentro de la empresa.

También manifiesta que dentro de sus organizaciones se ha mantenido un constante compromiso al momento de realizar el uso de nuevas tecnologías, pero al no tener conocimientos o nociones de estas tecnologías no logran una implementación de la misma manera actualmente no figuran con los recursos para dicha ejecución ya que piden diferentes costos tanto de infraestructura y a su vez de una buena capacitación debido a que la mayoría de las personas no tienen nociones del uso de herramientas en la nube ni el manejo de datos, pero que tienen nociones de la existencia de alternativas más económicas incluso gratuitas.

En este tiempo actual concurren muchas herramientas que admiten realizar este tipo de sistematizaciones como lo es la arquitectura ya que esta menciona a diferentes plataformas que conllevan al manejo y diferentes servicios que mencionan a datos es por eso se llevó a cabo la realización de un centro médico para la hospitalización en la ciudad de Colombia Medellín ya que al realizar aquello el resultado final fue que las herramientas necesarias y tecnológicas se lleva acabo con el gobierno que conlleva a obtener diferentes tipos de datos y a sus vez no generan un buena sunción de estas tecnologías en cuestión a herramientas [4].

Existe un ejemplo como lo es del mercado de valores llamado Análisis Real Time que conllevan a las tecnologías en diferentes ambientes, si hablamos de proyectos y de Big Data se debe adaptar a diferentes especialidades con mayor mercado en el sector tomando en cuenta el tiempo actual y el futuro , el hecho es que diferentes alumnos de diferentes áreas más si se habla de ingeniería convivan con este mundo de especialidades actuales y su vez tengan la experiencia necesaria [5].

Otro ejemplo muy claro es hablar sobre el gran Beneficio de un Data Lake de manera automática que conlleve con diferentes metadatos ya que hay una gran finalidad como lo es la debida gestión de un gran bulto de datos para en un futuro realizar un buen análisis y obtener buenos resultados de la misma para así logran un buen objetivo como lo es en esta investigación ya que en cierta formas se dota la emancipación de datos de diferentes clientes que conlleven a muchas a plataforma [6].

Así mismo para dicha implementación se tiene a un Data Lake en la ciudad de México y la Zmg se obtuvo una análisis para la debida infestación ya que esta investigación contiene servicios de la nube basado en su computación para obtener una infraestructura adecuada para la misma que para esto se utilizó el maravilloso servicio que contiene Amazon Web Services con el fin de poder tener un análisis de movilidad urbana y así poder realizar, cargar y analizar estos datos que ofrecen estos mismos servicios [7].

Un ejemplo en base Data Lake con su respectivo diseño en diferentes sectores como lo

son los servicios eco turísticos que a su vez diseñan una guía exploratoria y conceptual con el fin de realizar un estudio técnico basado en las penurias de una empresa o negocio para dar una respectiva solución esto a su vez en la gran Colombia sector eco turístico [8].

Dichos ejemplos dados están básicamente encaminados a la adopción y sobre todo a el despliegue de un Data Lake para gestionar los repositorios de datos incluso no se ha propuesto una definición formalidad debido a los temas investigados ya que dichos trabajos existentes no están completos y faltan partes importantes del tema que incluso no se centran en el dominio de dificultad de los datos, el papel de infraestructura de esas soluciones.

En termino para lo mencionado anteriormente se expone un buen estudio de investigación sobre un Data Lake con dicho servicio en la nube para el debido y posterior análisis de datos para la Federación de Organizaciones Populares Peninsulares, ya que trae consigo una gran mejora en diversos procesos para de esa manera llevar a cabo este tipo de análisis en el departamento de archivos y TICS, que es son realmente indispensable en diferentes áreas al manejo de datos totalmente fuertes y a gran volumen.

1.2.DESCRIPCIÓN DEL PROYECTO

Un punto importante es que los diversos tipos de datos son muy importantes para diversa empresa y para la economía totalmente mundial. El Big Data al momento de realizar el respectivo análisis de la misma intenta conseguir una acumulación de diversos tipos de datos y a su vez estos datos sean netamente de la empresa que a su vez trabaje en diversos tipos de escritura y lectura en sus esquemas continuos que para de esa manera se pueda lograr el alcance propuesto. [9].

Este proyecto de investigación netamente se centra en la propuesta de proveer una guía para organizaciones, empresas o entidades e incluso a personas que actúen y brinden el servicio de Data Lake en Cloud Computing que como resultado de este trabajo investigativo se logrará la documentación de la guía o manual y el conocimiento de cómo

posibilitar o manejar un estudio para el servicio que ofrece Cloud Computing en un Data Lake.

A continuación, se presenta las 4 fases que se exhibe en la procreación de un Data Lake de acuerdo con el manual que ofrece el Instituto Nacional de Estándares y Tecnología en sus siglas en inglés (NIST), este proceso permite determinar el nivel de amenaza potencial y sus riesgos asociados con un sistema TIC.

FASE 1: Levantamiento de Información.

Se realiza la documentación que permitirá organizar de forma apropiada toda la información correspondiente en el tema investigativo de estudio, dentro de las actividades referentes a esta fase están las siguientes:

- Establecer diferentes bibliografías netamente actualizadas y a su vez estructuradas sobre guías de adopción de despliegue de Data Lake dentro de la nube.
- Establecer diferentes bibliografías actualizadas y estructuradas sobre el proceso de Cloud Computing que conlleven al manejo de un Data Lake.
- Primer acercamiento a las autoridades y directores de las diferentes organizaciones que conforman la federación para dar a conocer la propuesta para un posterior levantamiento de información a través de encuesta.
- Elaboración de encuestas y entrevistas a realizar a los dirigentes organizacionales para conocer la realidad que presentan dentro de sus organizaciones con respecto a las tecnologías basadas a un Data Lake.

FASE 2: Caracterización de la empresa.

Se realiza la visita a las organizaciones filiales a la federación, con el propósito de realizar los estudios correspondientes a la estructura y formación de dichas empresas, conocer su área de tecnología y observar cual ha sido la influencia que ha tenido el modelo de Cloud Computing.

- Visita a las instituciones:

La visita a las organizaciones tiene como propósito conocer las instalaciones, la arquitectura que por medio de la caracterización permite conocer cómo funcionan los

procesos internos a través de identificación de elementos esenciales que facilitan la gestión y control de dichos conocimientos.

- Análisis de los resultados obtenidos de la fase anterior por medio de formularios de Google y Power BI.

FASE 3: Establecer el estudio de Factibilidad.

Se tendrá en cuenta la información tomada en las actividades de la fase anterior con el fin de poder analizarlas y dar origen a las siguientes actividades correspondientes a esta fase de la metodología:

- Organizar la información recolectada en la visita a las organizaciones.
- Realizar una comparación entre la información encontrada en el levantamiento de información con la encontrada en las visitas a cada una de las empresas.
- Establecer el estudio de la factibilidad para determinar el grado de estabilidad de los diversos servicios que contiene el Data Lake.

FASE 4: Elaboración de la guía.

Se tendrá en cuenta toda la información obtenida hasta el momento, para elaborar el manual que es una herramienta muy importante ya que facilitará la generación de un documento de apoyo para los directivos de la federación al patrocinar tecnologías de Data Lake dentro de Cloud en dicha institución como lo es la federación con un gran propósito que es el de presentar recomendaciones encuadrando los diversos aspectos de tecnologías dentro del área de gerencia de la federación.

Teniendo en cuenta la información sobre la elaboración de la guía de procesos de adopción se inicia con su elaboración.

Las actividades para esta fase son:

- Aclaraciones Legales
- Estrategia Cloud
- Diseño de solución
- Contratación de Cloud

El presente proyecto investigativo se ayuda mediante una importante línea de investigación con temas netamente relacionada con infraestructuras y de seguridad, a su vez con temas de virtualización y así mismos referentes a la computación en la nube, que por lo consiguiente ayuden a los accesos a información indispensable para una buena toma de decisiones a un futuro [10].

Basado a su vez en una sub línea de investigación basado en Tecnologías y sistemas de información dentro del cometido de conocimiento en las organizaciones que tiene como intención la realización de una investigación en el impacto de habilidades y diversas estrategias si se habla del desempeño de una organización o empresa que así también incluye una gran investigación sobre la mejora e implementación de modelos de gestión explicando su impacto en las empresas, organizaciones y la humanidad existente en la sociedad [11].

1.3.OBJETIVOS DEL PROYECTO

1.3.1. OBJETIVO GENERAL

Propuesta de una guía de adopción de despliegue de un data lake dentro de la nube para la Federación de Organizaciones Populares Peninsulares, basado en un modelo tecnológico de Cloud Computing para proporcionar una visión general de la situación actual de la investigación sobre los servicios dentro de la nube y las principales teorías sobre adopción aplicadas.

1.3.2. OBJETIVOS ESPECÍFICOS

- Realizar un levantamiento de información referente a la construcción de una guía de proceso de adopción, los lineamientos de uso y manejo de servicios en Cloud Computing.
- Establecer una caracterización de la federación y diversas organizaciones y a su vez una visión frente al modelo de Data Lake de Cloud Computing.
- Definir la factibilidad de adopción en la organización para tomar los servicios en la nube, a partir del análisis y caracterización.
- Elaborar una guía para la adopción y uso de Data Lake para la Federación enmarcando los aspectos de importancia de Cloud Computing.

- Realizar simulador para proporcionar un guía replicable para la elaboración y creación de un Data Lake en las instituciones que manejen cantidad masivas de información.

1.4. JUSTIFICACIÓN DEL PROYECTO

En años antiguos hasta la actualidad se ha logrado una gran identificación de muchos campillos que aprecian una multiplicación inesperada de datos el cual por aquello esto realiza el aumento de una gran cantidad e magna de datos que se deben sentenciar y la complejidad de conformaciones en los que se amontonan las reseñas que esto a su vez forma coacción sobre las subestructuras automatizadas de informática y los métodos de análisis de datos existentes, ya que cada vez más datos se consideran una germen ventajoso de información para poder así tener una buena toma de decisiones críticas y acertadas en diferentes tipos de campos únicos [12].

Esta justificación teórica está basada en presentar los beneficios para la organización de utilizar las ciencias de datos como un Data Lake en un Cloud Computing. En este caso es bajo los servicios de Microsoft Azure y diferentes plataformas conocidas que ofrecen los servicios de almacenamientos de lago de datos a nivel empresarial, el cual proporciona muchas formas de desplegar las aplicaciones corporativas de manera eficaz [13].

Este proyecto de investigación en primera petición beneficiará a los directivos de la federación y encargado del departamento de archivista, en resolver problemas prácticos como el tener acceso a la administración de forma local y con base de datos estructurados, no estructurados y hojas de Excel o dicho cualquier tipo de datos almacenados e importantes de la institución, así como reportes de importancia y relevancia que permite el correcto funcionamiento organizacional.

Y como beneficiarios indirectos se encuentra las organizaciones asociadas a la federación, sus coordinadores y presidentes que son encargados de llevar de manera manual la información almacenada en documentos y hojas los cuales la federación lo digitaliza y almacena en su sistema local de archivos para luego usar estos datos para posteriores reportes al momento de una rendición de cuenta o cambios de directiva, entre otros. Este

tipo de tecnologías es útil para recopilación de datos de grandes escalas, el análisis de datos, con visualización y capacidades de bucle de retroalimentación.

Es muy importante tener en cuenta de que muchas empresas, organizaciones y sobre todo las federaciones tengan un gran interés sobre diversos tipos de tecnologías y sobre todo poder implementarlas más que todo si se trata de Big Data para que así se pueda obtener un buen método administrativos y sobre todo tratables y concisos para dicha entidad, los cuales deben ayudar al fortalecimiento de una organización y así mismo a su vez al desarrollo que proporcione el delicado uso de estos procesos. [14].

La resolución de este estudio en la Federación de Organizaciones Populares Peninsulares (FOPOPE) accederá el uso de la nueva tecnología que esta será utilizada mediante herramientas como lo es el Open Source ya que esta permite calzar muchos puntos como lo es la construcción de infraestructura para lograr un buen almacenamiento de datos de la federación, esto ayudará mucho a la federación y al gran almacenamiento que se tiene en la nube que como anteriormente eso era muy escaso y así mismo como un buen resultado se permitirá analizar estos datos de manera más fácil y entendible para el usuario.

Por medio de la realización de esta guía se podrá conocer los conceptos muy claves para el buen intelecto si se habla de las tecnología en nube ya que estas buscan de manera inmediata poder realizar una consumación en la federación, el cual permitirá que cualquier persona con poco conocimiento o de conocimiento intermedio puedan tener como referencia este documento o despejar sus dudas ante alguna inquietud y así mismo podrán hacer uso de los datos de manera correcta cuando se requiera.

Cabe destacar que el resultado de la guía servirá como base para todas aquellas instituciones con los mismos procesos o que estén inmersas en obtener un gran cambio tecnológico y que sobre todo que deseen adquirir este tipo de innovaciones y requieren de alguna asesoría específica para el tipo de servicios que ofrecen.

La presente propuesta esta direccionado al Plan de creación de oportunidades, haciendo énfasis en 3 ejes relevantes, en el cual se detallan lo siguiente:

Eje 1: Económico [15].

Objetivo 1:

Incrementar y fomentar, de manera inclusiva, las oportunidades de empleo y las condiciones laborales.

Objetivo 4:

Eje 1: Económico Garantizar la gestión de las finanzas públicas de manera sostenible y transparente.

Eje 2: Social [15].

Objetivo 5: Proteger a las familias, garanticas sus derechos y servicios, erradicar la pobreza y promover la inclusión social.

Eje 5: Institucional [15].

Objetivo 14: Fortalecer las capacidades del Estado con énfasis en la administración de justicia y eficiencia en los métodos de regulación y control, con independencia y autonomía.

1.5. ALCANCE DEL PROYECTO

El propósito de este documento es proporcionar una guía de evaluación tanto para organizaciones o empresas que ofrezcan servicios de Cloud Computing o que deseen estar al tanto en el diseño y elaboración de un Data Lake dentro de la nube así como entidades que vayan a pactar servicios bajo este modelo de dependencia aunque al basarse en las buenas prácticas de COBIT 5, puede emplear a cualquier tipo de entidad, enfocándose en decretar el nivel de seguridad de su servicio tomando en cuenta los principales procesos de seguridad y sus respectivos catalizadores. El propósito de esta guía es poder obtener un gran apoyo en cuanto al inicio hasta el final para así evaluar los diferentes elementos de seguridad de una organización.

La propuesta prometerá el buen soporte ineludible a la Federación de Organizaciones Populares Peninsulares de la Provincia de Santa Elena y al uso de estas nuevas tecnologías de Data Lake que se encuentran dentro de la nube y para así se pueda llevar a cabo dicho

expansión y servicios de estas divergentes tecnologías de una manera oportuna con pedestal a las penurias de las organizaciones.

El presente proyecto contará de las siguientes fases:

- Fase de Levantamiento de Información.
 - Establecer la bibliografía netamente actualizadas y estructuradas sobre guías o manual de procesos de adopción de Data Lake.
 - Establecer bibliografía únicamente actualizadas y estructuradas sobre Cloud Computing con orientación hacia de un Data Lake.
 - Elaboración de encuestas y una entrevista para el levantamiento de información para conocer y visualizar la realidad que se presenta con respecto a la computación en la nube dentro de la federación.

- Fase de Caracterización de la federación y sus organizaciones.
 - Mediante la caracterización consentirá conocer el funcionamiento de los procesos netamente internos de la federación, a través de la respectiva identificación de elementos que con ayuda se facilita a la gestión y control de las técnicas a utilizarse en la ayuda para la respectiva adopción de tecnologías de Data Lake dentro de nube.
 - El análisis de los resultados de la fase de levantamiento de información es realizado mediante la herramienta Power Bi, a través de datos de las encuestas antes realizadas, esto permite conocer mediante graficas de datos simplificados, los resultados obtenidos en la opinión de cada una de las personas encuestadas y conocer la factibilidad en la adopción de esta guía para la federación.

- Fase de estudio de factibilidad.
 - Organizar la respectiva investigación y entrega de información recolectada en la visita a la organización respectiva.
 - Realizar una comparación entre la información encontrada en el levantamiento de información con la encontrada en las visitas a cada una

de las empresas.

- Establecer el estudio de factibilidad para determinar el valor del grado de persistencia de los servicios de Cloud Computing para el Data Lake.

- Fase de Elaboración de la guía.

Paso 1 Aclaraciones legales

Existe incertidumbre respecto a qué limitantes legales existen en la utilización de servicios Cloud por parte de los organismos. Las organizaciones públicas en muchos casos manejan información de carácter sensible, la cual se encuentra regida por ciertas leyes que deben ser consideradas al momento de realizar una implementación Cloud. Este primer punto busca aclarar estas dudas, entregando un resumen de las leyes que involucran en el uso de esta nueva tecnología Cloud.

Paso 2 Estrategia Cloud

Las organizaciones poseen un nivel de madurez insipiente en relación con el uso de tecnologías Cloud. Por lo tanto, si existe una intención de promover el uso de tecnologías Cloud, es necesario introducir a las organizaciones en el tema, y los potenciales beneficios asociados a su utilización.

La estrategia que busca fomentar esta guía postula una adopción de modelos Cloud y Lata Lake por parte de los organismos, mediante una migración paulatina de sus servicios, en base a proyectos específicos y la utilización de tecnologías Cloud para los mismos.

Paso 3 Diseño de Solución.

El diseño de la solución nace en respuesta a un análisis de los requerimientos del sistema que la organización en cuestión desea implementar. A partir de estos requisitos, y contrastándolos con las variables Cloud involucradas, es posible diseñar una posible solución que se aproxime de optima manera a lo que el organismo busca.

Escenarios Cloud por considerar

1. Arquitectura Cloud (Nube pública, privada, Híbrida)
2. Proveedor Cloud (Nacional, Extranjero)

Pasos para diseñar una solución Cloud

- Análisis de información a tratar
- Identificar al usuario
- Niveles de servicio.
- Análisis de costo.

Paso 4 Contratación Cloud.

Con los puntos anteriores cubiertos, es momento de comenzar a establecer algún posible proveedor Cloud que satisfaga los puntos establecidos en el diseño de la solución. Muchas veces el mismo equipo asesor debiese ser capaz de ofrecer, ya sea como resellar o dueño de la infraestructura, los servicios a los que se quiere optar.

La propuesta de estudio manifestará con las diferentes evidencias de rasgos únicos a esta nueva tecnología de la ciencia de datos y su vez sobre la computación en la nube en cuanto a sus peculiaridades características, sus diferentes ventajas al momento de utilizar herramientas y así mismo al implementar dicho servicio. El examen oportuno sobre las tecnologías que tiene la federación nos abastece mucha información de mucha importancia para poder identificar las mejores procedencias en relación con los apoderados de las diversas misiones y gestiones que están dentro de la organización, al momento de obtener esta nueva tecnología.

1.6. METODOLOGÍA

1.6.1. METODOLOGÍA DE LA INVESTIGACIÓN

El análisis exploratorio de la metodología utilizada proporciona al investigador un buen ambiente investigativo debido a que obtengas diversas bases y conocer los diferentes lineamientos ineludibles que admitan que lo investigado actualmente sea lo más audaz y verídico posible [16]. La propuesta que se ha realizado en esta dicha investigación de manera tecnológica aprovechará de una manera importante y atenderá su mano al estudio de estas nuevas tecnologías, por lo tanto, se usará dicha exploración para averiguar y hablar sobre el tema con una amplia visión. Esta investigación se verificará mediante la investigación de trabajos que contengan diversos párrafos bibliográficos referentes al tema a realizar y a su vez habituando lo propuesto a realizar y conocer nuevas

perspectivas.

La investigación diagnóstica es aquella en la que ante un problema específico se pretende realizar un análisis de las actitudes, opiniones y motivaciones que lo han ocasionado con el objeto de tener información que sirva de ayuda a su resolución y permita proponer alternativas para tomar decisiones o corregir el curso de una acción [17]. La investigación diagnóstica se realizará a través de encuestas a los directivos de la federación y organismos que lo conforman. Con esta información se podrá identificar las necesidades y mejoras a emplear en el proyecto.

Este tema de investigación estudiará aproximadamente del 50% de la información de datos que se facilitó a través de las personas entidades que forman parte del medio a analizar.

1.6.2. TÉCNICA DE INVESTIGACIÓN

En técnicas de la investigación se detallan las diferentes técnicas e instrumentos de recolección de información que se manejarán y utilizarán en el transcurso del desarrollo de este tema investigativo:

➤ **Técnicas:**

Estado del arte, encuesta, entrevista y fuentes bibliográficas.

➤ **Instrumento:**

La entrevista estará dirigida al presidente de la federación, con el objetivo para obtener información específica y relevante de la organización y conocer sobre los conocimientos de nuevas tecnologías en dicho entorno y su impacto sobre la adquisición. Además, se realizarán encuestas al personal administrativo de la organización con el objetivo de identificar su nivel de conocimiento y manejo con respecto a las tecnologías.

Las fuentes bibliográficas aportarán al conocimiento para el buen análisis de información especializada vinculadas al tema de investigación.

➤ **Población:**

La población ente de estudio la conforman los entes de la Federación de Organizaciones Populares Peninsulares (FOPOPE).

Análisis de entrevista

Análisis de entrevista realizada al Sr. Sergio González Tigrero presidente de la Federación de Organizaciones Populares Peninsulares (FOPOPE) de la provincia de Santa Elena.

El día 16 de diciembre del año 2021 se realizó una entrevista al presidente de la federación (FOPOPE), con el objetivo de conocer la situación actual sobre importante manejo que tienen estas tecnologías y sobre todo para conocer la percepción sobre el Data Lake en la nube.

En primer lugar, el Sr. Sergio González mencionó que la tecnología cada día va avanzando, es de mucha importancia en la vida personal laboral y profesional ya que nos ayuda a organizarnos mejor y aprender cosas nuevas. Así mismo manifestó que el avance de cada día de la tecnología es positivo para las organizaciones ya que ayuda al desarrollo social y lo más importantes es en el entorno empresarial más que todo a las organizaciones ayuda a su innovación con respecto a su sistema. El considera que sería muy bueno aplicar las tecnologías ya que esto nos ayuda a simplificar el trabajo que anteriormente era muy demoroso y mientras más avance la tecnología más se debe innovar y actualizar y por ende invertir en aquella.

Así mismo con respecto a la adaptación de almacenamiento y procesamiento en la nube en la organización manifestó; si, tenemos personal capacitado que ya tienen experiencia en el manejo de almacenamiento en la nube y que en los últimos años se han manejado base de datos en lo que comprende estatutos reglamentales, base de datos de sistema de gestión documental y sistema de archivo el cual se ha almacenado por 5 años hasta la fecha actual. Sostuvo también que no tiene conocimientos sobre computación en la nube, datos e información y lago de datos, pero si se tienen un departamento de TICS que si tiene conocimiento que manejan en la actualidad almacenamiento en la nube y que se encuentra preparado para ese tipo de actualizaciones y que sobre todo con capacitaciones se lograría captar los temas. También sobre la terminología de almacenamiento de datos dio un ejemplo, resguardar la información con seguridad a posibles robos.

El Sr. Sergio González dijo nuestros datos están respaldados en la nube y en servidores que lo manejan los encargados de TICS y que si estaría dispuesto a efectuar un cambio hacia la innovación dentro de la institución porque toda empresa que no innova se queda en la vanguardia y no progresa y puede dejar de existir. Es por ello dijo: Tenemos una línea base en la cual podemos realizar mejoras con respecto a la toma de decisiones a

través de programas que analicen estos datos almacenados.

Una vez realizada la entrevista se pudo concluir que efectivamente el presidente da cabida para la adopción de esta tecnología como una guía para una futura aplicación ya que consta con los recursos y herramientas necesarias para aplicar estas tecnologías para un buen beneficio de la federación.

INDICADOR

Reducción del tiempo de respuesta mediante la utilización de una guía o manual de adopción de una arquitectura en el área de TICS de la federación FOPOPE para así que el área de archivística obtenga un repositorio en la nube para la respectiva creación de un Data Lake y así redirigir una buena organización de sus datos de dicha entidad.

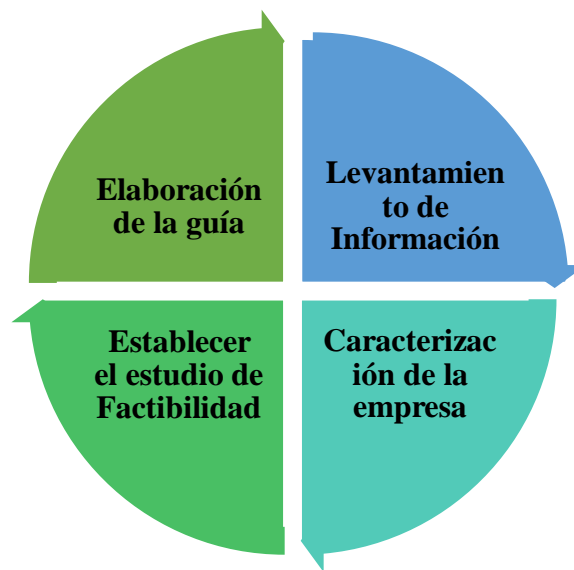
Esto valdrá como puntal para la respectiva evaluación del tiempo que se tarda en poder hacerlo de manera intuitiva y a su vez al realizarlo por medio de una guía haciéndola comparación de manera variable cuantitativa y ver el tiempo que se tarda entre ambos.

1.6.3. METODOLOGÍA DE DESARROLLO

La presente investigación se adecuará como lineamiento a la metodología NIST que tiene como orientación el fundamento en una guía o manual de ejecución para la entidad que desee adquirir dicho servicio en la nube que así mismo desliga los pasos netamente necesaria para una culta misión en su aplicación [18].

La metodología de NIST se sintetiza en cuatro diferentes fases y se dividen de la siguiente manera:

Figura 1: Fases adaptadas de la guía de NIST



Fuente: Instituto Nacional de Estándares y Tecnologías

FASE 1: Levantamiento de Información [18].

- Establecer un marco referencial con diferentes argumentos sobre el proceso de adopción de Data Lake Storage con dirección hacia la nube.
- Elaboración del proceso de manuales diagnósticos como es el levantamiento de información mediante encuestas y entrevistas, para el procesamiento y análisis de los resultados de estudio de adopción de un Data Lake en la nube para la Federación FOPOPE

FASE 2: Caracterización de las empresas [18].

- Mediante la caracterización permitirá conocer cómo funciona el proceso a través de la identificación de elementos esenciales que permiten la gestión y control de los procesos dentro de la federación FOPOPE.
- El estudio de análisis de los resultados de la primera fase como es el de levantamiento de información es realizado mediante la herramienta Power Bi que a su vez esto sirve para el análisis de los resultados siendo esto auténtico para el alcance del proceso de estudio e implementación dentro de la federación.

FASE 3: Establecer el estudio de factibilidades [18].

- Realizar el importante estudio de factibilidad Técnica para establecer los recursos netamente necesarios para la adopción de los servicios de Cloud Computing para

Data Lake.

- Realizar el estudio de factibilidad Operativa para decretar el grado de concurrencia que tiene la entidad en paralelo con el procedimiento a un Data Lake.
- Establecer el estudio de factibilidad Económica para determinar la calidad de grado de estabilidad financiera que posee la federación en el área de TICS que establezcan la aprobación y aceptación económica.

FASE 4: Elaboración de la guía [18].

Mediante la guía se establece los pasos a seguir para proporcionar el proceso de caminos de adopción de proveedor de Cloud Computing que brinden el servicio de Data Lake. Estos importantes pasos para seguir también suelen ser de mayor utilidad para la debida identificación de ser o no ser posible utilizar un conjunto Cloud para una viable solución que se anhele topar. La sucesiva guía está edificada a partir de basto bibliográfico relativa a la materia complementada con entrevistas a involucrados en el tema.

Paso 1 Aclaraciones legales

Existe irresolución respecto a qué existen diversos restrictivos lógicos en la utilización del servicio Cloud por parte de las entidades ya que aquellas son netamente públicas y que en muchas cuestiones manipulan información de carácter sensitivo, la cual se halla gobernada por ciertos códigos que incumben ser respetadas al instante de cumplir una ejecución Cloud. Este primer punto busca esclarecer estas incertidumbres, adjudicando una sinopsis de los códigos que implican el uso de un conjunto de tecnologías Cloud.

Paso 2 Estrategia Cloud

Las entidades poseen un nivel de conocimiento ignorante en correlación con la rutina de un conjunto de técnicas Cloud. Por lo tanto, si está una intención de suscitar el uso de tecnologías Cloud, es forzoso embutir a las entidades en el tema y los importantes beneficios que tiene al realizar su utilización.

Paso 3 Diseño de Solución.

El diseño de la solución aparece en contestación a un análisis de los requerimientos del sistema que la entidad en cuestión desea realizar. Partiendo de estas obligaciones, y

contrastándolos con las inconstantes Cloud es posible trazar una posible solución que se aproxime a lo que la entidad busca y necesita.

Escenarios Cloud

- Arquitectura Cloud consiste en (Nube pública, Nube privada y Nube Híbrida)
- Proveedor Cloud (Sean estas Nacionales o Extranjeras)

Pasos para el diseño de una solución Cloud

- Análisis de información.
- Identificar al usuario
- Niveles de servicio.
- Análisis de costo.

Paso 4 Contratación Cloud.

Es este punto se establece algún proveedor Cloud ya por el pasar de todo logre cumplir los diferentes puntos del diseño de la solución ya que en ocasiones es la misma entidad financiera es quien debe dar y ofrecer la infraestructura y los servicios que requiera en dicha institución

CAPÍTULO II

2. LA PROPUESTA

2.1. MARCO CONCEPTUAL

2.1.1 Amazon Web Services AWS

Ofrece un ligado de equipos y productos de servicios de Cloud que en unido forman una plataforma de computación ofrecidas a través de web ya que cuenta con un conjunto de productos y de peculiaridades ofreciendo desde conjunto de técnicas de fundamentos en donde destaca el almacenamiento de datos[19].

2.1.2 Google Cloud Platform GCP

Es un servicio de informática indudable con el que consigues elaborar aparatos tácticos en la subestructura de Google [20].

2.1.3. Azure

Destaca como el principal servicio en la nube donde puede desenvolver y elaborar presentaciones de proceso de peta bytes de identificaciones de datos de grafía con los significativos y evolutivos lenguajes como lo es el U-SQL, R, Python y .NET obteniendo la infraestructura para disponer y así se logre resolver los datos a postulación.[21].

2.1.4. Data Lake

Repositorio único en donde se pueden acumular indivisibles los datos de una compañía estas pueden ser estructurados, no estructurados, hojas de cálculos el hecho es que estén sin ningún tipo de proceso y facilitar en el proceso de ser almacenados hasta que puedan ser netamente analizados [22]. Esta ayuda está edificado para la nube y es obviamente y únicamente concurrente con HDFS ya que este se caracteriza por tener diferentes capacidades empresariales [22].

2.1.5. Definición de Cloud Computing

NIST concreta la sistematización en la nube como tal modelo que facilita mediante la red un debido camino hacia un conjunto de recursos y servicios únicos de automatización configurables ubicados en Data Centers que a su vez define como un modelo para la

retribución que puede proveer ágilmente para lo consiguiente seguir que tenga un buen funcionamiento por parte del proveedor del servicio. [23].

Tabla 1: Características de computación en la nube

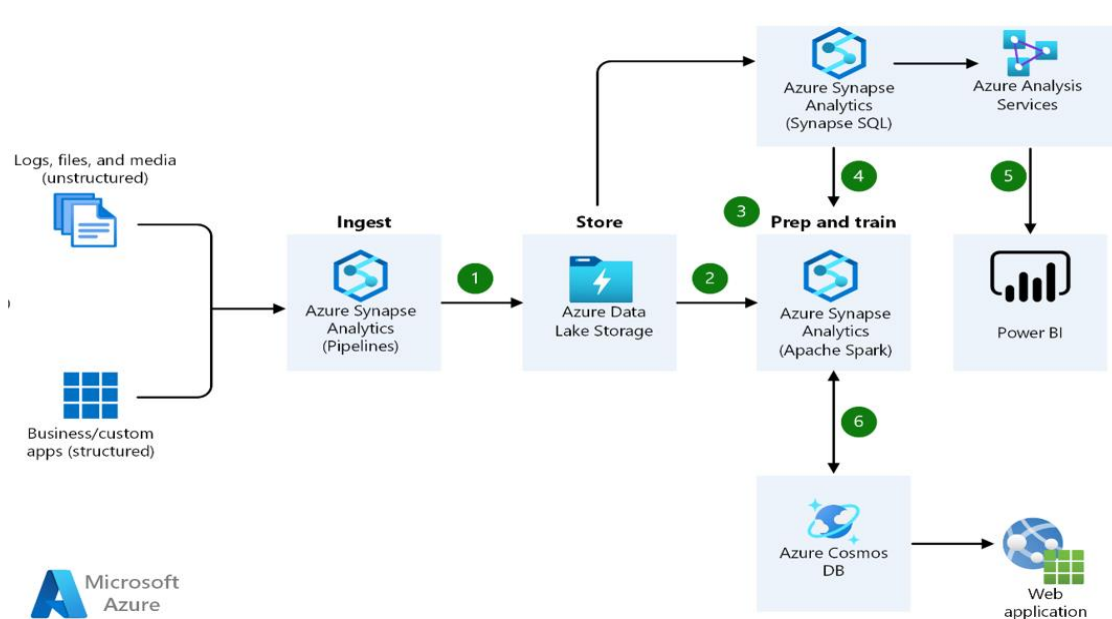
COMPUTACIÓN EN LA NUBE	
CARACTERÍSTICAS COMUNES	CARACTERÍSTICAS ESENCIALES
❖ Ampliación Intensiva	❖ Autoservicio en Demanda
❖ Uniformidad	❖ Extenso acceso a la red
❖ Virtualización	❖ Puesta común de Recursos
❖ Software bajo costo	❖ Rápida Elasticidad
❖ Cómputo Flexible	❖ Servicio Medido

2.1.6. TIPOS DE DATA LAKE

2.1.6.1 Azure Data Lake

El origen de datos de Azure para diferentes Data Lake es conmemorativamente y único a ADLS y sobre todo la generación 2 ya que esta desarrolla la funcionalidad del acopio de cosas de Azure Blob Storage y lo empapa a otro sistema que en este caso es en un procedimiento de ficheros netamente concurrente con los terminales de Hadoop y a su vez tiene seguridad, escalabilidad y disponibilidad que a diferencia Posee un valor importante como lo es por usanza y una escalabilidad infinita [24].

Figura 2: Infraestructura de Data Lake Azure

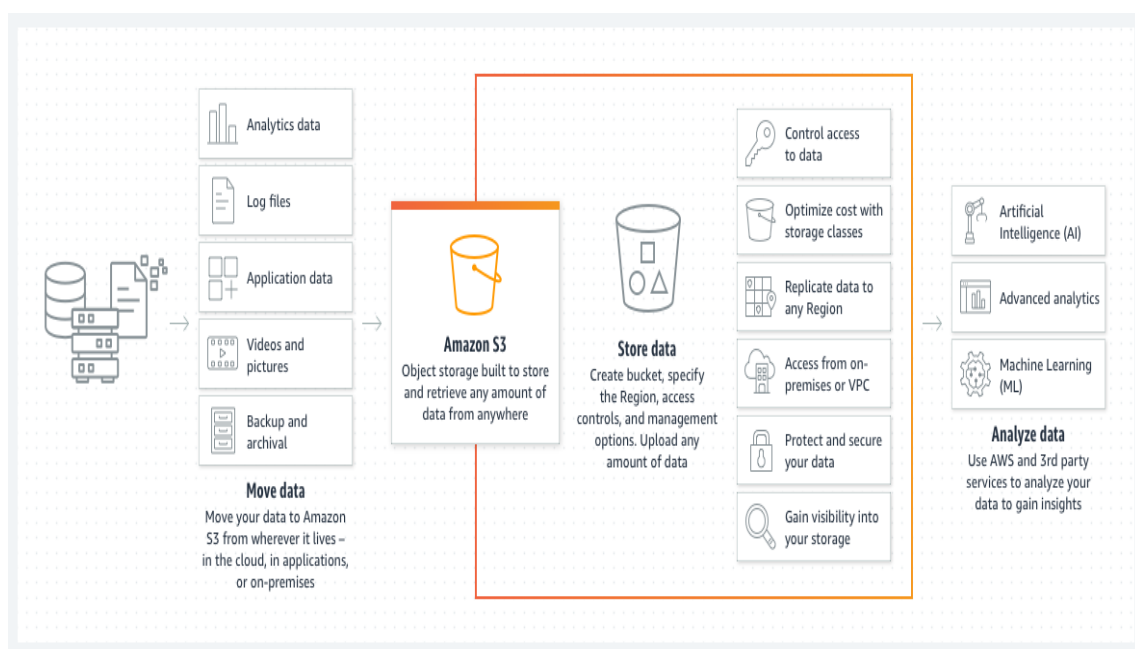


Fuente: <https://docs.microsoft.com/>

2.1.6.1.1 AWS Data Lake

Suministra diferentes iniciativas para realizar un Data Lake ya que se conciertan de algunos conjuntos de técnicas que logramos dividir por clases como lo es la fuente de datos ya que este servicio de acumulación de objetos suministra la constancia necesaria y también es el fragmento semejante a ADLS en Azure que en su importancia brinda una buena escalabilidad y se caracteriza por tener disponibilidad, da una excelente seguridad y sobre todo un buen rendimiento[25].

Figura 3: Infraestructura Data Lake AWS

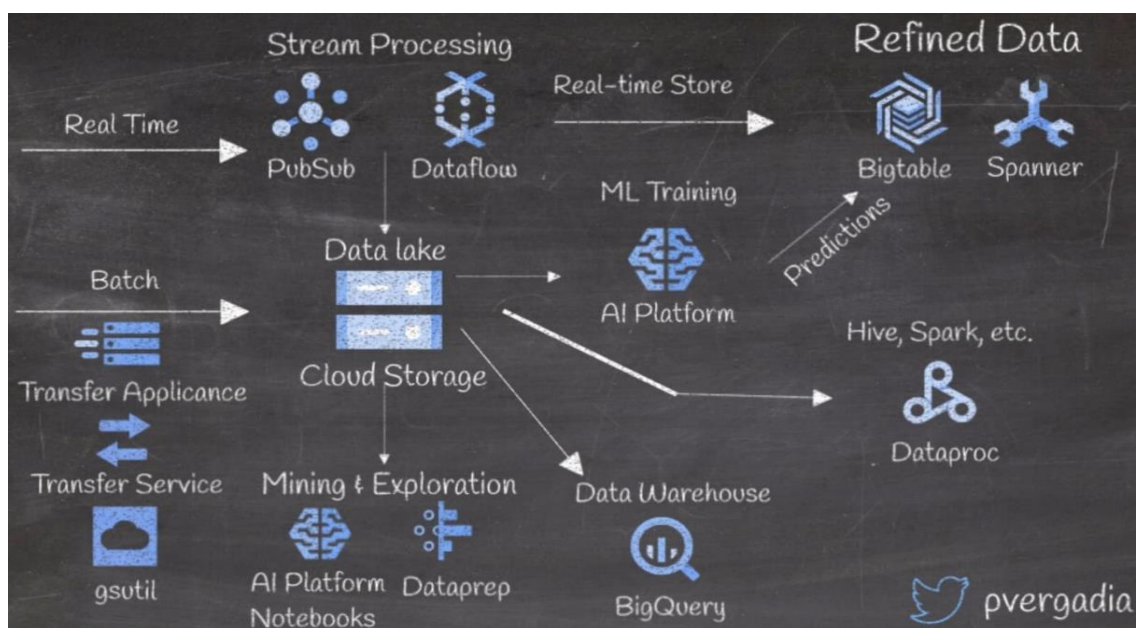


Fuente: <https://aws.amazon.com/>

2.1.6.1.1.1. Google Data Lake

Otra representación de extender el tema de un data lake es siguiendo una buena comercialización de Hadoop ya que estas tecnologías acceden en la expansión de Data Lake y está asentada en el conjunto de técnicas de Hadoop y en el almacenamiento de datos que realiza en HDFS que a su vez puede ser recuperable cuando se requiera o sea posible.[26].

Figura 4: Infraestructura de Data Lake en Google



Fuente: <https://cloud.google.com/>

Tabla 2: Comparativa de los servicios de almacenamientos aplicando el pago de los servicios en los diferentes proveedores en la nube.

Moneda	AWS S3 Standard	Azure DL Store Gen2	Google Cloud Storage
USD	11,88	156,5	11,27
EURO	9,80	129,14	9,30

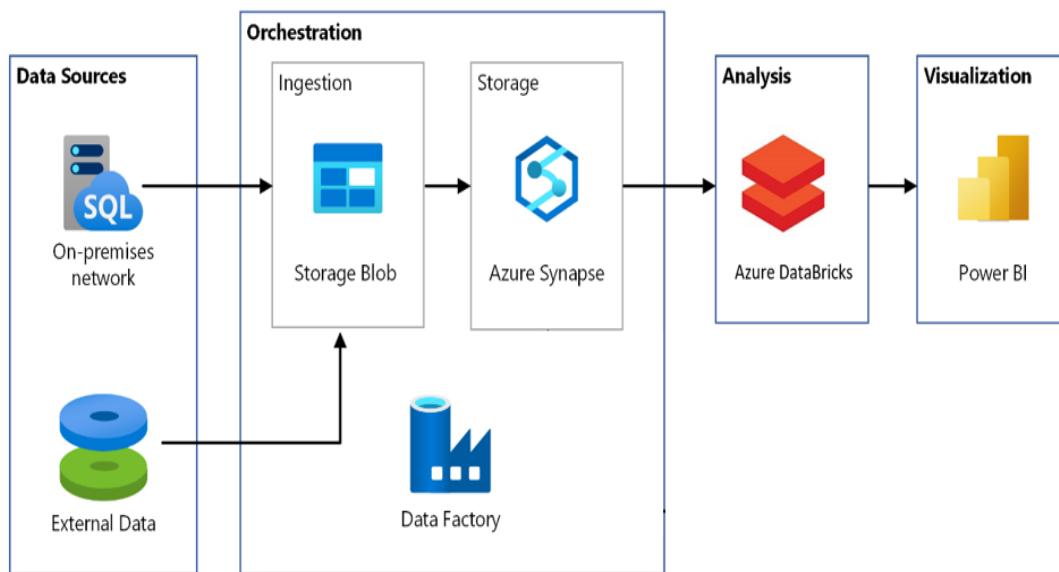
Fuente: explodat.cl

Los diferentes vendedores en el proveedor en la nube tienen una gran diversidad si se habla de los precios que ofrecen dentro del servicio en la nube, en esta semejante visualización de los costos en el cuadro comparativo en la aplicación de la producción del costo del servicio pagados como envoltorios, los mismos que dan una principal cabida y afán y más peculiaridades, beneficios se presenta a la cantidad de acumulación y seguridad de datos intrínsecamente de los sistemas de archivos es por aquello que cada quien de ellos brindan de afín a la necesidad del usuario que por lo consiguiente obtiene los siguientes precios de categoría standard que comprende desde los \$11,27 mensuales hasta los \$156,50.

Los servicios de Data Lake que ofrece Microsoft son:

- Azure Data Lake Store O Storage (ADLS) Gen1 / Gen2.
- Azure Data Lake Data Factory.
- Azure Databricks.
- Power Bi.

Figura 5: Infraestructura en Microsoft Azure

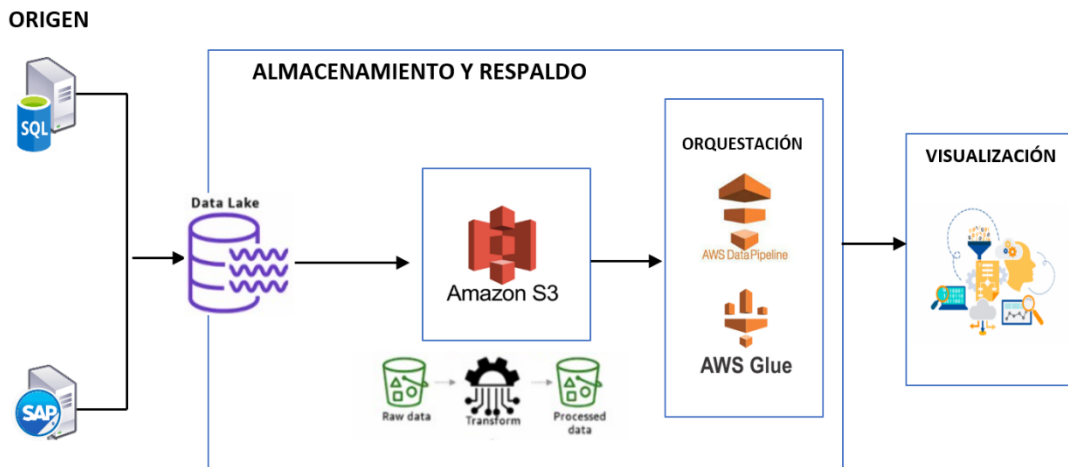


Para la solución en Azure tenemos el ingreso de datos con el origen de SQL o Excel, pasando por la orquestación en Data Factory, seguido de la ingestión de los datos por medio del almacenamiento dentro de los contenedores y repositorios de data lake Storage, el análisis se realiza a través de Azure Databricks, una vez analizados los destinos de los datos son visualizados a través de desbordad en Power Bi.

Los servicios de Data Lake que ofrece AWS son:

- Amazon Simple Storage Services (S3).
- Amazon Data Pipeline.
- Amazon Glue.

Figura 6: Infraestructura en AWS

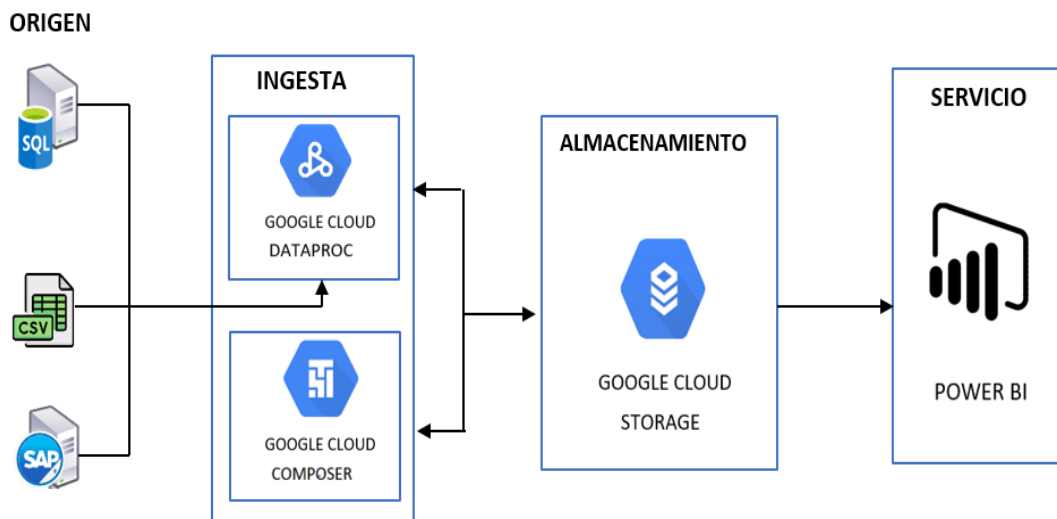


El proceso de data lake en Amazon Web Service es similar a todos los proveedores anteriores comenzando con un origen de datos pasando por el almacenamiento en el Storage y orquestando los datos para ser luego analizados y visualizados.

Los servicios de Data Lake que ofrece Google son:

- Google Cloud Storage.
- Google Cloud DataProc.
- Google Cloud Composer.

Figura 7: Infraestructura en Google



El origen de los datos comienza desde los datos estructurados, atraviesa por una ingesta de datos en Google DataProc, y Google Cloud Composer de los datos almacenados en los repositorios de Cloud Storage estos datos luego son analizados y mostrados en un visualizador de cuadros de mandos.

2.1.6.1.1.2. AZURE DATA LAKE STORAGE.

Es un repositorio de una masiva cuantía de datos en forma estructurada y no estructurada que se acopia en su formato nativo, forma parte de la plataforma de nube pública de Microsoft Azure que a su vez esta es considerada para el respectivo análisis y almacenamiento de Big data. [27].

En tiempos anteriores el uso de Azure Data Lake Gen1 era muy habitual y utilizado de forma concurrente pero en la actualidad eso quedo atrás, ahora Azure Data Lake Gen2 es uno de los manejados y que ahora principalmente se informa que el 29 de febrero de 2024 este sistema antiguo como lo es Gen1 se retirará por el cual toda persona que hace uso a este sistema debe migrar sus datos a la nueva generación.[27]

Tabla 3: Comparación rápida entre Azure Data Lake Gen1 y Azure Data Lake Gen2

<p>El almacenamiento de Data Lake Azure Gen1 logra almacenar exactamente un sin número de cuantía de datos de diferentes tipos de tamaño durante cualquier período de tiempo. Que por tanto generalmente se elaborará trabajos metódicos esgrimiendo exámenes de Data Lake en los datos almacenados en el comercio</p>	<p>Mientras que hablando de Azure Data Lake Storage Gen2, esta versión avanzada tendrá tanto las opciones de almacenamiento que es el almacenamiento del sistema de archivos como el almacenamiento del sistema de objetos.</p>
<p>De acuerdo con Azure Data Lake Storage Gen1, el usuario podrá almacenar los datos en forma de almacenamiento de archivos que se separa en bloques que le brindan rendimiento y seguridad. Esto puede depender de la jerarquía de almacenamiento del lago de datos de Azure</p>	<p>Mientras que Azure Data Lake Gen2 le brinda seguridad, así como un rendimiento saludable y sin problemas. Cuando hablamos de almacenamiento de objetos, está ahí para la escalabilidad.</p>

Azure Data Lake Gen1 no admite almacenamiento en caliente / frío y redundante	Considerando que Azure Data Lake Gen2 admite almacenamiento en caliente / frío y redundante
---	---

Fuente: <https://docs.microsoft.com>

Data Lake Storage Gen1 en cuanto a Azure tomando en cuenta que ya existe la generación 2 es un repositorio de hiperescala corporativo para obligaciones de compromiso en una entidad cuando se trabaje con Big Data. Nos aprueba detener datos de diferente tamaño, que así mismo el tipo y la importante velocidad de ingesta en un solitario lugar que de manera obligatoria debe realizar el facilita miento para el respectivo análisis operativo y exploratorio [27].

2.2. MARCO TEÓRICO

2.2.1. Agentes impulsores de la adopción de Cloud Computing en las empresas.

¿Quién mueve la nube?

El saliente trabajo investigativo examina la atribución sobre el amparo de Cloud Computing en las empresas. Los resultados que se obtuvo en el estudio principalmente se basó en diversas encuesta entregada a las diferentes entidades de las nuevas tecnologías ubicadas en los Jardines Especializados en tecnologías del país de España ya que estas por si marcan que son los Despenseros de agentes promotores de la adopción institucional de Cloud Computing, declarando su dominio sobre la ayuda tanto cuando operan de representación desierta tal cuando lo hacen en unión con diferentes agentes que se conforman entre ellas que por ultimo existen algunos resultados que a su vez admiten eliminar terminaciones e implicaciones en el área de la Dirección de Empresas. [28].

2.2.2. Data Lake para la Dirección de Análisis Universitaria: Arquitectura y Metadata.

El capítulo investigativo muestra el bosquejo de un Data Lake para una universidad el cual esta consiste en un método de acumulación para la misión y análisis de datos que siempre tenga el mismo objetivo de dicha entidad. También se obtiene que el diseño utiliza diferentes bandas para acumular datos en la zona primera acaudala los datos en riguroso por consiguiente como segundo punto acumula los datos procesados y por último la tercera y única que es la zona de oro el cual es el que realiza la colocación de los datos

que pueden ser analizados y utilizados por los indiscutibles de datos ,si bien es cierto que por otro lado también se debe tomar en cuenta que tener un gran volumen de datos personifican una enorme conformidad para otras entidades y que estos pavimenten venir en corpulencia, velocidad que sobrepasen de las capacidades de los sobrantes de gestiones habituales para su cogida, acopio y procesamiento en un lapso sensato [29].

2.2.3. Data Lake: Aplicaciones, Herramientas y Arquitecturas.

Avanzando hacia el creación Cloud muchas de las sociedades o entidades ya están recogiendo productos de lo anteriormente cosechado, como lo es la rapidez de abastecimiento, la importante y valida agilidad, también tenemos a la escalabilidad de los consumos corrientes de automatización, en sacar conclusiones muchas de las empresas no están actualidades referente a estar nuevas tecnologías que a su vez de tener dichos conocimientos en diferentes temas como lo es el Big Data, no pueden tener el mejor de los resultados es por esto que gracias al Data Lake existe un gran y masiva información dentro de las tecnologías de la IOT [30].

2.3. COMPONENTES DE LA PROPUESTA

Se creó cuentas gratuitas en los diversos proveedores de servicios en la nube como lo es en Azure, Google y AWS para el uso adecuado de su servicio de Data Lake y sus respectivos componentes para dicha creación de repositorios para el almacenamiento.

Las cuentas por utilizar en el proyecto investigativo son las siguientes:

- Cuenta en Portal Azure
- Cuenta en Google Cloud Platform
- Cuenta en Amazon Web Services

Portal Azure

- Creación de recursos
- Creación de almacenamientos
- Creación de contenedores
- Añadir recursos o datos al sistema de archivos

Google Cloud Platform

- Creación de depósitos
- Configuración de Cloud Storage

- Almacenamiento de datos dentro del sistema de archivos

Amazon Web Services

- Creación de Bucket de Amazon
- Adjuntar daos dentro del sistema de archivos.

Sistemas de Archivos.

En cada sistema de archivos de los servicios de data lake de los diferentes proveedores trabajan de una manera muy peculiar y similar en el almacenamiento de datos tomando en cuenta que son de diferentes tipos de configuraciones y distribuciones tanto para seguridad y a su vez como lo es los permisos acceso.

Seguridad

Azure Data lake Storage

- Firmas de accesos compartidos SAS: Se concede una URI en donde permite el acceso respectivo a los recursos de Azure.
- Token de acceso compartido: Permite dar autorizaciones precisas de lectura, escrituras y ejecución a los datos acopiados en el método de archivos.
- Administración dirección ACL: Instituir y gestionar todos los permisos para los registros fundados en el servidor de almacenamiento.

Google Storage.

- Protección de los datos de objetos: Fundada mediante la norma ISO/IEC 27002 en donde está centrado en los diversos servicios de la nube.
- Encriptación de datos: Datos netamente almacenados están encriptados mediante AES256 y AES128 que se guardan en la misma.
- Permisos de lectura escritura y ejecución: Permite dar autorizaciones precisas de lectura, escrituras y ejecución a los datos acopiados en el método de archivos.

AWS

- Permisos de lectura y escritura.
- Acceso de control de listas ACL.

- Configuración de Access Ponts.

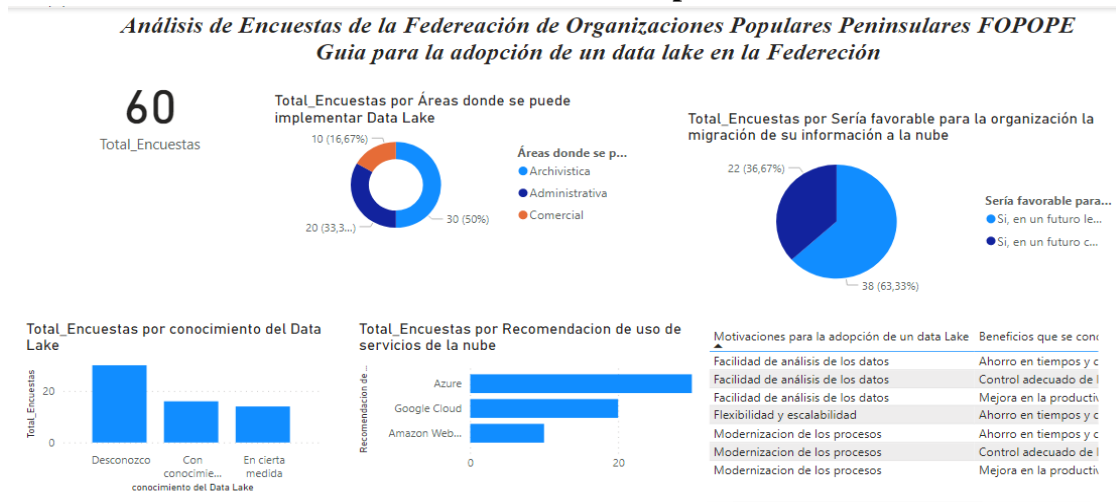
FASE 1

LEVANTAMIENTO DE INFORMACIÓN

Para la obtención de una excelente planeación ante los servicios en la nube y Data Lake, que ofrece los diferentes proveedores que se deben considerar para la institución, es necesario la recopilación de información mediante comparativas de sus aspectos relevantes que conciernen los tipos de servicio y modelo de despliegue de esta nueva tecnología para la federación de Organizaciones Populares Peninsulares FOPOPE. Se analizó el resultado de las encuestas que se abordó a las diferentes oficinas y personal dentro de la Federación, analizándolos de una forma detallada en Power BI para tener un análisis más completo de los resultados.

TABLEROS DE CONTROL REALIZADOS EN POWER BI

Figura 8: Análisis de la factibilidad de la adopción de Data Lake en la Federación – Power-Bi desktop



A través de las sesenta encuestas realizadas a los diferentes dirigentes de las organizaciones que conforman la federación pudimos obtener los siguientes resultados, En la ilustración se muestra un dashboard que consta de graficas de los servidores más conocidos y utilizados en cierta medida por parte de los dirigentes, esto conlleva a una idea posible de adopción a estas tecnología y usarlas en cierto periodo de tiempo ya que tienen varios beneficios para la optimización de sus procesos archivísticos y de esa manera obtener resultados más favorables en favor de la mejora de la federación.

FASE 2 CARACTERIZACIÓN DE LA EMPRESA

Referente a los temas tratados anteriormente se establece la caracterización interna de la federación para conocer el grado de factibilidad y arquitectura que compone federación y de esa manera conocer con que infraestructura cuenta o no y analizar el grado de perspectiva para la respectiva adopción de un Data Lake y sobre todo poder realizar el análisis de cuál es la adecuada para la federación según el manejo y grado de información procesada dentro de aquella.

La caracterización de la empresa permitirá conocer cómo se encuentra actualmente la institución y conocerla factibilidad y requisitos necesarios para el estudio de adopción de lago de datos y continuar con la fase siguiente.

Tabla 4: Caracterización Del Departamento De Administración.

Directiva Administrativa		
RESPONSABLE DEL PROCESO: presidente de la Federación.		
OBJETIVO: Realizar la planeación estratégica, establecer el plan operativo que permite lograr los objetivos y metas, asegurando el correcto funcionamiento de los procesos organizacionales con el fin de alcanzar el éxito de la federación.		
ALCANCE: Dirigir y entregar los lineamientos de acción de todos los procesos, hacer seguimiento y revisión a la eficacia, conveniencia y adecuación de los procesos que se piensan implementar.		
ENTRADA	ACTIVIDADES	SALIDA
- Estados financieros de la entidad. -Informe de avance y resultados. - Lineamientos de la junta directiva de la entidad.	1-Definir plan estratégico. 2-Definir y asegurar el cumplimiento de las políticas y objetivos de la calidad. 3-Definir responsabilidades.	- Plan estratégico. - Misión y política de calidad. - Estructura organizacional y competencias. - Objetivos de calidad.
RECURSOS REQUERIDOS		SEGUIMIENTO
HUMANOS: Personas totalmente capacitadas para asumir grandes proyectos (Gerencia general, coordinador de proyectos, coordinador administrativo y financiero).		-Auditoría interna. -Revisión por la dirección anualmente de cada proceso.

Tabla 5: Caracterización Del Departamento de Archivística.

Archivísticas		
RESPONSABLE DEL PROCESO: presidente de la Federación.		
OBJETIVO: Organizar la información de la federación y sus entes organizacionales y hacer que sea útil y accesible para todos de una manera rápida y sencilla.		
ALCANCE: Llevar el control y orden de los archivos, oficios, estatutos, útiles de la federación, así como almacenarlo de una manera digital y optar por su conservación a largo plazo,		
ENTRADA	ACTIVIDADES	SALIDA
- Archivos documentos	1-Archivar documentos 2.- Organizar Documentos 3- Conservar documentos a través de un sistema de gestión de archivos	- Digitalización de documentos -Archivar documentos -Almacenamiento de datos en base de datos locales
RECURSOS REQUERIDOS		SEGUIMIENTO
HUMANOS: personas totalmente capacitadas para asumir el control y almacenamiento de los archivos y cómo administrarlo de manera digital.		-Revisión por la dirección anualmente de cada proceso.

Tabla 6: Caracterización Del Departamento De TIC

Tecnología de la Información y Comunicación		
RESPONSABLE DEL PROCESO: jefe del Departamento de Tecnología de la información.		
OBJETIVO: Generar soluciones estratégicas en el área de tic para el cumplimiento de los fines misionales del Ministerio del Trabajo y formular lineamientos de estándares y buenas prácticas para el manejo de la información del departamento y la federación.		
ALCANCE: Incluye la identificación de necesidades TIC, elaboración del Plan estratégico de TIC.		
ENTRADA	ACTIVIDADES	SALIDA
<ul style="list-style-type: none"> • Lineamientos de Archivo. • Programa de Gestión Documental- Archivos documentos • Plan Estratégico de Tecnologías de Información 	1- Procedimiento Gestión de Soluciones de TIC 2.- Procedimiento Gestión de Seguridad 3- Procedimiento Gestión de Conocimiento 4.- Soporte Backup información	<ul style="list-style-type: none"> - Arquitectura empresarial TI - Informes del Proceso de Gestión de TI - Política de TIC - Política de Seguridad - Base de datos de conocimiento
RECURSOS REQUERIDOS		SEGUIMIENTO
HUMANOS: personas totalmente capacitadas y profesional en gestión del área de tic.		-Revisión por la dirección y presidente de la federación.

N° inventario	Tipo de bien	Descripción del bien			Estado del bien	Precio
		Marca	Modelo	Serie		
2	Impresora	CANON, XP-8600, Series – AIO			Bueno	\$180
1	Impresora	Epson, nXp 2101 2100			Bueno	\$150
1	Laptop	HP, corei3, J4025 2.0 GHZ-4GB-1TB-DVDRW-BLANCO-20.7"FHD-			Bueno	\$600
1	Laptop	Lenovo negro, V330 (14", Intel)			Bueno	\$550
2	Computadoras de escritorio	LG, Core i5, Dell Inspiron 15 Serie 3000			Bueno	\$1500
1	Computadoras de escritorio	Intel Core i5 serie P de ThinkStation			Bueno	\$1200
1	Servidor	Servidor Dell PowerEdge T140 Xeon Quad Core 16gb 2tb DVD			Bueno	\$1479
3	Teclados	Casio, DS7447E			Bueno	\$15
3	Soportes de CPU	Dual Socket Lga1366			Bueno	\$52.85
4	Reguladores de voltajes	Steren. Cdp 1000va 8 Tomas			Bueno	\$26
3	Repuestos de ventiladores de CPU	Acer nitro 5			Bueno	\$30
1	Tóner de fotocopidora	Ricoh Mpc3003			Bueno	\$1599
3	Mouse	Gamenote, serie Xpt			Bueno	\$ 8
3	Parlantes	Genius, serie pant1			Bueno	\$10
3	Cámaras web	Xtratech, serie camq			Bueno	\$20
4	Cartuchos	HP 305XL			Bueno	\$10
2	Cable USB	Anker			Bueno	\$2.50
1	Router	Huawei, Modelo B310s-518			Bueno	\$25
1	Switch	Telemecanique, Swich 2 CNT Block 240VAC 3A, serie XE2			Bueno	\$1400
Total						\$8914,34

Tabla 7: Infraestructura en hardware de la federación en la actualidad

Oficina matriz posee una oficina de tecnología de la información y comunicación

Nombre de la Institución: Federación de Organizaciones Populares Peninsulares

Departamento: Tecnología de la Información y Comunicación.

Tabla 8: Infraestructura en software de la federación en la actualidad

N°	Descripción	Paquete	Proveedor	Operatividad	Precio
1	Office Hogar y Estudiantes 2021	Perfecto para un máximo de 6 personas. Hasta 6 TB de almacenamiento en la nube, lo que equivale a 1 TB por usuario. Funciones de seguridad premium en la aplicación móvil.	Microsoft	Si	USD\$79,99al año
1	Solución NOD32 Antivirus 1 equipo, 1 año	Análisis de seguridad remota de sus dispositivos Fácil uso compartido de las licencias Aplicación móvil para iOS y Android Gestión de licencias centralizada	Eset NOD32	Si	\$ 25.19 + IVA
1	Sistema Operativo a 6 máquinas	Windows 10.	Microsoft	Si	USD\$289,00 al año
1	Sistema web de archivos	Programado en Php, MySQL, montado en su servidor.	Ambiente libre	Si	\$0,00
1	Página Web	Ambiente web, código abierto	Ambiente libre	Si	\$0,00

		HTML.			
--	--	-------	--	--	--

FASE 3 ESTABLECER EL ESTUDIO DE FACTIBILIDAD

Para tomar en cuenta que un proyecto de investigación sea factible se debe lograr netamente resultados únicos y reales en todos las ilustraciones planteadas cada uno de estos estudios son estimados de manera igualada ya que el análisis de posibilidad corresponderá tener superioridades enormes en los exteriores económicos, expertos, operativa, económico ya que tomando en cuenta que si los resultados son verdaderamente favorables se reflexiona que la propuesta es realizable para dicha institución.

En dicha fase se despliega el análisis de factibilidad para la respectiva ejecución la misma que simbolizan una gran superioridad competitiva por el ahorro de costos, en la solución en la nube debe ser minucioso para establecer qué ejemplo de nube y servicio son los que se van a elegir e realizar en la federación.

En este proyecto investigativo se tomará en cuenta un análisis de factibilidad de las diferentes de empresas del ecuador referentes a temas de Computación en la nube [31], en donde se realiza el análisis respectivo de los 4 aspectos primordiales a mencionar.

1. Estudio de Factibilidad Técnica.
2. Estudio de Factibilidad Operativa.
3. Estudio de Factibilidad Económica Financiera.

FASE 4: Elaboración de la guía

PASO 1 ACLARACIONES LEGALES

En Cloud Computing existen métodos y reglas totalmente legales encaminadas al proveedor de servicios en la nube en la que abastece fortunas de equipamientos con relación a la informática que se debe guardar la pesquisa del cliente en integridad del servicio adquirido ya que es el cliente el cual debe cancelar todo lo adquirido.

En lo establecido de diferentes artículos en las Leyes ecuatorianas de la república del Ecuador con información referente a la computación en la nube hace referencia a la “Ley de Comercio Electrónico Firmas y mensajes de Datos” en el cual nos podemos referir al

capítulo II los siguientes artículos [33].

Art. 45.- Validez de los contratos electrónicos. – Los contratos podrán ser instrumentados mediante mensajes de datos. No se negará validez o fuerza obligatoria a un contrato por la razón de haberse utilizado en su formación uno más mensajes de datos [33].

Art. 46.- Perfeccionamiento y aceptación de los contratos electrónicos. – El Perfeccionamiento y aceptación de los contratos electrónicos se someterá a los requisitos previstos en las leyes y se tendrá como lugar de perfeccionamiento el que acordaren las partes. La recepción la confirmación de recepción o apertura de mensaje de datos, no implica aceptación del contrato electrónico, excepto acuerdo de las partes [33].

Art. 47.- Jurisdicción. – En caso de controversias las partes se someterán a la jurisdicción estipulada en el contrato, a falta de esta se sujetarán a las normas previstas por el Código de procedimiento Civil Ecuatoriano y esta ley siempre y cuando no se trate de un contrato sometido a la ley orgánica de defensas del consumidor en cuyo caso se determinará de domicilio el del consumidor o usuario [33].

Tanto en el Artículo 45 y 46 se encuentra de por medio la garantía de respeto a la voluntad de las partes, ante todo. Nótese, en este sentido que la contratación electrónica y temática es el revestimiento tecnológico y contemporáneo del antiguo principio de la voluntad de las partes.

También se hace referencia a disposiciones legales que tratan sobre las comunicaciones en general y su regulación, tomando en cuenta que el internet es parte de los servicios públicos de telecomunicaciones.

Según la ley orgánica de telecomunicación [34] es en el capítulo II sobre Prestadores de Servicios de Telecomunicaciones, artículo 24 dispone:

Son deberes de los prestadores de servicios de telecomunicaciones, con independencia del título habilitante del cual se derive tal carácter, lo siguiente:

Art 3. Prestar el servicio de forma obligatoria, general, uniforme, eficiente, continua, regular, accesible y responsable, en cumplimiento de los parámetros de calidad establecidos por la Agencia de Regulación y Control de las Telecomunicaciones y de lo estipulado en los títulos habilitantes [34].

Art 14. Adoptar las medidas necesarias para la protección de los datos personales de sus usuarios y abonados, de conformidad con esta Ley, su Reglamento General y las normas

técnicas y regulaciones respectivas [34].

PASO 2 ESTRATEGIA CLOUD

Entregado que el objetivo a extenso plazo es fructificar los variados beneficios que recibe la federación gracias a la tecnología que a su vez es totalmente indispensable instituir una habilidad que se conserve para lograr diferentes objetivos planteados anteriormente. La habilidad mostrada se ocasiona en base al enfoque propio que ha sido determinado en este manual en plenitud con las energías del régimen digital que al momento de llegar a una difusión sobre el automatismo de sistemas Cloud a nivel público. Esta habilidad puede recomendada para cualquier entidad el cual se establece una gran recomendaciones momento de establecer cierto tipo de procedimiento que domine diligencias al breve, mediano y largo plazo.

La habilidad que se tiene este manual es poder obtener una gran adopción acompasada de sus servicios en base a proyectos totalmente únicos y específicos en base a la utilización de tecnologías Cloud para la federación ya que esta se entiende que una adopción perfeccionista de construcción es un trabajo complicado. El camino que busca esta guía va en función de comenzar a realizar proyectos específicos ya sea una nueva solución o la emigración de un equipamiento de servicios mediante la utilización de SaaS que aplica Data Lake. Dentro de las estrategias a considerar tenemos:

1. Estrategia Cloud alineada con la estrategia institucional

Las habilidades de acción varían elocuentemente en referente a las entidades como lo es crucial que los envueltos se afirmen que su habilidad en la nube apoye rápidamente la táctica de negocio de su entidad

2. Evaluar peligros

Al instituir una pericia en la nube es fundamental evaluar riesgos concernientes con la nube como el riesgo de prontitud, muy importante el riesgo a una mala seguridad y como último el riesgo de disponibilidad.

3. Cuestionar la reducción de costes como principal impulsor para la adopción de la nube

El coste total estriba del tipo de ayuda en la nube y las tipologías de los impuestos de

trabajo, y de los contextos específicas de la entidad. Como la mayoría de las distribuciones han adquirido práctica con ejecuciones reales en la nube, los beneficios se ven más en invención y velocidad en lugar de precios.

PASO 3 DISEÑO DE LA SOLUCIÓN

El siguiente punto se orienta en la investigación de un esbozo de tramitación a la inseguridad planteada a partir de los temas el tema a tratar en este punto es el tipo de dato, los diversos horizontes de servicio citados, sin dejar afuera las habilidades internas de las entidades visitadas que se puedan lograr educar al lector para que este logre trazar una tramitación.

Antes de establecer los pasos para aproximar una viable solución a horizonte de diseño, es necesario instituir los escenas o variables Cloud que se busca explicar a partir de estos puntos.

Análisis de la información a tratar.

Dentro de este análisis se busca y se encuentra una solución Cloud ya sea para mejorar su accesibilidad, así como también su debida disponibilidad o más aún si hablamos de seguridad ya que es una clave importante en la misma, son muchos y diferentes aspectos que se debe considerar ya que se debe elegir la correcta para la debida implementación y adopción en la federación.

Confidencialidad de la información a tratar

Aquí se evalúa dos diversos tipos de acontecimientos como lo tenemos en el primer tiene que ver con información de datos que a su vez tiene como finalidad ser utilizada por la ciudadanía en libre albedrío.

El segundo tipo de averiguación es de representación íntimo ya que son múltiples tipos de averiguaciones los que pueden abatir en esta condición, ya sea pesquisa confidencial de diferentes tipos de personas.

A continuación, se presenta la Tabla donde se establecen los posibles escenarios Cloud en base a la confidencialidad de la información.

Tabla 9: Escenario Cloud en base confidencialidad de la información

	Arquitectura Cloud	Proveedor Cloud
Información Confidencial	Privada o hibrida	Nacional
Información no Confidencial	Publica	Nacional o Internacional

Identificar el tipo de dato a tratar y su comportamiento

Con la respectiva información a negociar en los métodos es de mucha importancia para poder trazar una tramitación con compasivos tiempos de respuesta. El uso de conjunto de técnicas Cloud envuelve diferentes elementos que transgreden en las cosechas de respuesta, por lo cual desplegar un intelecto del tipo de dato y relacionarlo a las escenas Cloud es de gran calidad. A discrepancia de un vecino local, en un contexto Cloud se utilizando la escritura de internet para la entrega de datos y a su vez esto implica que compuestos componentes correspondan ser empañados en miramiento para que el trabajo sea muy bueno.

Tabla 10: Escenario Cloud en base al tipo de dato a tratar

	Proveedor Cloud
Dato multimedia (imágenes, sonido, documentos)	Nacional
Datos Informativo (texto)	Nacional o Internacional

Identificar al usuario

Este punto es muy importante resaltar los usuarios o personas encargadas de la administración y uso de todas las herramientas ya que es de mucha importancia poder dar cabida a los usuarios finales para la debida explicación y entendimiento del de las mismas para obtener un buen aplicativo y poder disminuir el tiempo de respuesta en las búsquedas correspondientes.

Usuario Objetivo de la solución

Desde el punto de vista de la arquitectura, nos encontramos con 2 situaciones: donde el sistema es de uso público y transparente para cualquiera a consultar, y cuando la información a consultar por los ciudadanos es privada y exclusiva para cada individuo.

Dada esta situación, se considera el uso de nubes públicas, y en caso de necesitar mayor seguridad, complementar con el uso de alguna arquitectura híbrida que resguarde la información confidencial específica.

Niveles de servicio

A nivel global son varias las corporaciones oficiales que piden de manera interna los diversos y altos niveles de servicio. A pesar de que esta es la reconstituyente a nivel organizacional, el contexto indica que viven diferentes recursos que no solicitan de horizontes de productos tan solicitantes. Es forzoso concebir esta contexto entregado que los superiores horizontes de favor por lo corriente son ofrecidos por vendedores, por lo cual es forzoso colocar en la ponderación todos los exteriores que conmueven una solución y prevalecer los de mayor categoría para así poder convergir a una mejor medio a la incierta. }

Tabla 11: Escenario Cloud en base al foco del sistema a utilizar

	Proveedor Cloud
Dato multimedia (imágenes, sonido, documentos)	Nacional
Datos Informativo (texto)	Nacional o Internacional

Análisis de costo y presupuestos

La dificultad para los cuerpos públicos es como consignar presupuesto para un costo que es variable. A partir de esta incierta cantidad para sumarse a un servicio de Cloud pública se requiere realizar una apreciación en un periodo fijo de los recursos a utilizar. Dado que los modelos Cloud públicas son comparativamente nuevos, la gran generalidad de los servicios Cloud prestados no manipulan costos variables debido que el modelo de percepción es en base a un sueldo y a una variable en puesto a acrecentamiento de recursos [35].

Tabla 12: Comparativa entre Plataformas Cloud.

PLATAFORMA	TIPO DE CLIENTE	COSTE	SERVICIOS	Hardware y Software
Google Cloud Platform	Pequeñas y medianas empresas. Se asocia con Cisco para llegar a las empresas	Gartner reconoce a GCP por sus precios centrados en el cliente, incluyendo grandes descuentos y contratos flexibles.	Menos servicios, pero capacidades técnicas avanzadas, con IA y herramientas de análisis de datos por delante de la competencia.	<ul style="list-style-type: none"> • Google Cloud Storage. • Google Cloud DataProc. • Google Cloud Composer.
Microsoft Azure	Empresas	Los descuentos basados en condiciones particulares requieren de ayuda para su evaluación.	Infraestructura robusta con capacidades probadas de nube híbrida.	<ul style="list-style-type: none"> • Azure Data Lake Store O Storage (ADLS) Gen1 / Gen2. • Azure Data Lake Data Factory. • Azure Databricks. • Power Bi.
Amazon Web Services	Cualquiera, pero con un enfoque no invasivo debido a su enorme tamaño.	Gartner recomienda el uso de una herramienta de terceros para estimar los precios de manera efectiva.	Grandes opciones, pero centradas en soluciones de nube pública, por lo que la implementación híbrida con centros de datos locales puede ser un reto.	<ul style="list-style-type: none"> • Amazon Simple Storage Services (S3). • Amazon Data Pipeline. • Amazon Glue.

Fuente: Acronis International GmbH

PASO 4 CONTRATACIÓN CLOUD

En este paso sobre la contratación Cloud lo primordial es realizar una respectiva revisión sobre que proveedores para revisar el debido cumplimiento de los pasos anteriores en base al diseño de la solución así mismo se establece diferentes aspectos forzosos a reflexionar en la convención vinculante con el proveedor. Ambos exteriores se hallan afines a lo proponente por parte de los proveedores ya que dentro de ellos se encuentran los requerimientos necesarios a la adopción.

Confidencialidad de la información

La confidencialidad de los datos como tal, que se almacenará en cualquier clase de nube. Es necesario revisar las características previas otorgados por el proveedor Cloud con respecto a esta temática, aportando políticas necesarias para uso como, por ejemplo, quienes tienen acceso a la misma. Bajo esta política se pretende asegurar que los datos no se utilizaran con ningún otro fin más que la operación del sistema que administra la institución. y, ningún otro tercero debiese tener acceso a la misma.

Cláusulas de confidencialidad a considerar en contrato con respecto a la confidencialidad de la información.

- Acceso a terceros: Se establece mediante estipulación, que el acceso a la información almacenada en el sistema Cloud tendrán un manejo exclusivo entre proveedor y el cliente.
- Uso de la información por parte del proveedor: La información será utilizada bajo lo estipulado del contrato.
- Establecer actividades del proveedor: Se conservará bajo contrato las acciones específicas establecidas que el proveedor ejecutará con la información almacenada en el sistema.

Seguridad de la información

Debe formar el contrato a través de las medidas que el proveedor efectúa para poder cumplir la seguridad e integridad de la información, enfocadas por protocolos y estándares de seguridad de la información.

- Implementación de medidas de seguridad en base a alguna certificación: El proveedor Cloud debe contar con certificación de seguridad e integridad de información, para realizar la contratación, la cual debe estar especificada en el

contrato vinculante.

- Protección tránsito de data: Las políticas de seguridad deben enfocarse en los protocolos de seguridad para la transferencia de información es decir poseer canales seguros de donde se traslada la información, para así contrarrestar que algún tercero mal intencionado, intercepte la información, ya sea para adquirir o modificar la información.
- Encriptación de información almacenada: los niveles sugeridos de encriptación deben ser superiores a DES, tales como AES (NIST) de 128, 192 o 256 (bits) para el almacenamiento de información.

Propiedad Intelectual

Es importante detallar que la propiedad intelectual de la información almacenada pertenece a la institución que lo maneja y no deben ser utilizados por el prestador de servicio. Existen 3 tipos de información que pueden ser consideradas como propiedad intelectual de la institución.

Primero se refiere al software, que en el caso de ser un desarrollo específico de la institución, el código y la propiedad intelectual del mismo es de la institución dueña del mismo sistema.

La segunda es la información como tal, que pueden ser subidos o que se encuentran dentro del aplicativo o infraestructura Cloud, como estos serán utilizados dentro de la plataforma o almacenamiento de datos.

La tercera sería la información originada a partir del uso del sistema de almacenamiento o base de datos, tomando en cuenta que todo tipo de información en la institución se considerada como propiedad intelectual de la misma.

Terminación del contrato

Es importante considerar las condiciones para la terminación de servicio o contrato en caso de que la institución así lo requiera, por lo cual el contrato debe la libertad de terminación de servicios y contrato por parte del cliente.

2.4. REQUERIMIENTOS

Objetivo Especifico	Requerimiento del proyecto
Realizar un levantamiento de información referente a la construcción de una Guía de proceso de adopción, los lineamientos de uso y manejo de servicios en Cloud Computing.	REQ01: Investigar proveedores de soluciones data lake en la nube temas relacionados de información bibliográficas referente al tema.
	REQ02: Levantar Información mediante encuesta para conocer el estado actual de la federación con respecto a infraestructura ti.
	REQ03: Analizar los resultados mediante Power Bi para ver la factibilidad de aprobación del proyecto investigativo a trabajar dentro de la federación
Establecer una caracterización de las organizaciones y su visión frente al modelo de data lake de Cloud Computing.	REQ04: Realizar la caracterización de los diferentes departamentos e infraestructura interna que lo compone
	REQ05: Realizar la caracterización del hardware y software e infraestructura del departamento de TI que contempla la organización para una futura solución a Cloud Computing
	REQ06: Comparar la infraestructura actual con una solución en la nube para verificar características, beneficios y una posible elección de proveedor y solución en la arquitectura tanto local como en la solución en la nube.
Definir la factibilidad de adopción en la organización para tomar los servicios en la nube, a partir del análisis y caracterización.	REQ07: Realizar la factibilidad Técnica, para determinar si los servicios de TI ofrecidos pueden ser compatibles o adaptables a los que ofrecen data lake en Cloud Computing y el grado de estabilidad que tendrán cuando estos sean implementados
	REQ08: Realizar la factibilidad operacional a través del análisis de las encuestas realizada al encargado del departamento de ti cuyos resultados nos ayuda a predecir si se debe poner en marcha la nueva solución, puntualizando en los beneficios que ofrece la misma.

	REQ09: Realizar la factibilidad económica. Se considera un análisis de costo de adquirir y operar la solución sugerida y se establece una comparación entre la plataforma actual y la nueva. Este análisis permite seleccionar el escenario más conveniente económicamente y viabilidad y rentabilidad para la federación.
Elaborar una guía para la adopción y uso de Data Lake para la federación enmarcando los aspectos de importancia de Cloud Computing	REQ10: Investigar las leyes y normativas que rigen en Ecuador con respecto a computación a la nube
	REQ11: Realizar la búsqueda de un diseño de solución a la problemática planteada a partir de las temáticas mencionadas anteriormente, consideran dos principales variantes como es arquitectura y proveedor iCloud.
	REQ12: Presentación de la guía completa para la federación.
Realizar simulador para proporcionar un guía replicable para la elaboración y creación de un data lake en las instituciones que manejen cantidad masivas de información.	REQ13: Simulación en la adopción por medio de una guía de pasos a seguir para la creación y uso de un Data Lake en Azure.

2.5. DISEÑO DE LA PROPUESTA

Tabla 13: Herramientas utilizadas en la infraestructura local

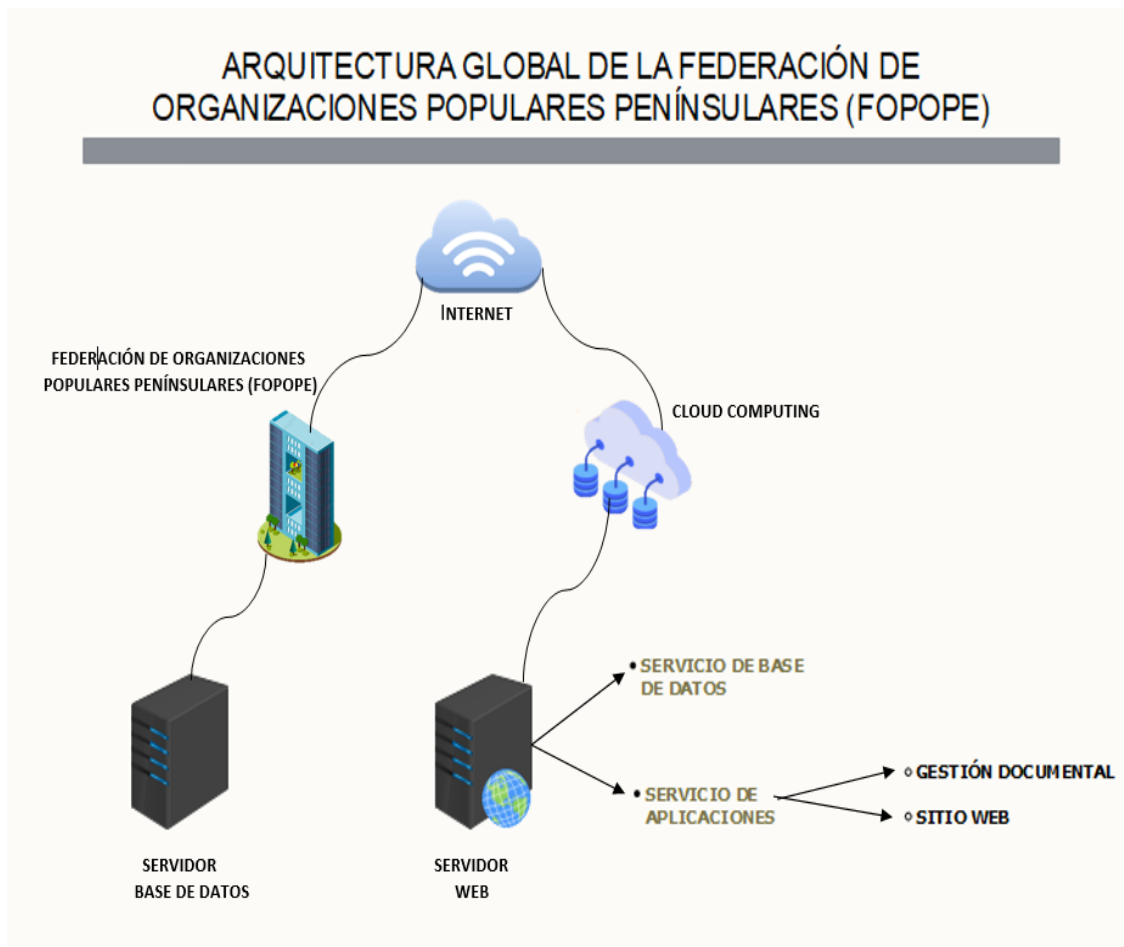
1	Router	Huawei, Modelo B310s-518	Bueno	\$25
1	Internet CNT Fibra Óptica	80 MBPS */ Velocidad simétrica	Bueno	\$37.99* incl. imp.
1	Servicio de nube	Data Lake Azure Storage gen2	Bueno	\$89.28
2	Servidor	Servidor Dell PowerEdge T140 Xeon Quad Core 16gb 2tb DVD	Bueno	\$2958
1	Firewall DMZ	Firewall Fortinet FG-60F 10 x GE RJ45 Puertos (Incluyendo 7 Puertos Internos 2 Puertos WAN 1 Puerto DMZ)	Bueno	\$120.99

SOLUCIÓN LOCAL Y EN LA NUBE

Arquitectura Global.

En la gráfica se presenta la solución posibles resultados de los pasos de la guía tanto a nivel local y Cloud Computing.

Figura 9: Diseño del enlace del servidor local con el Cloud Computing

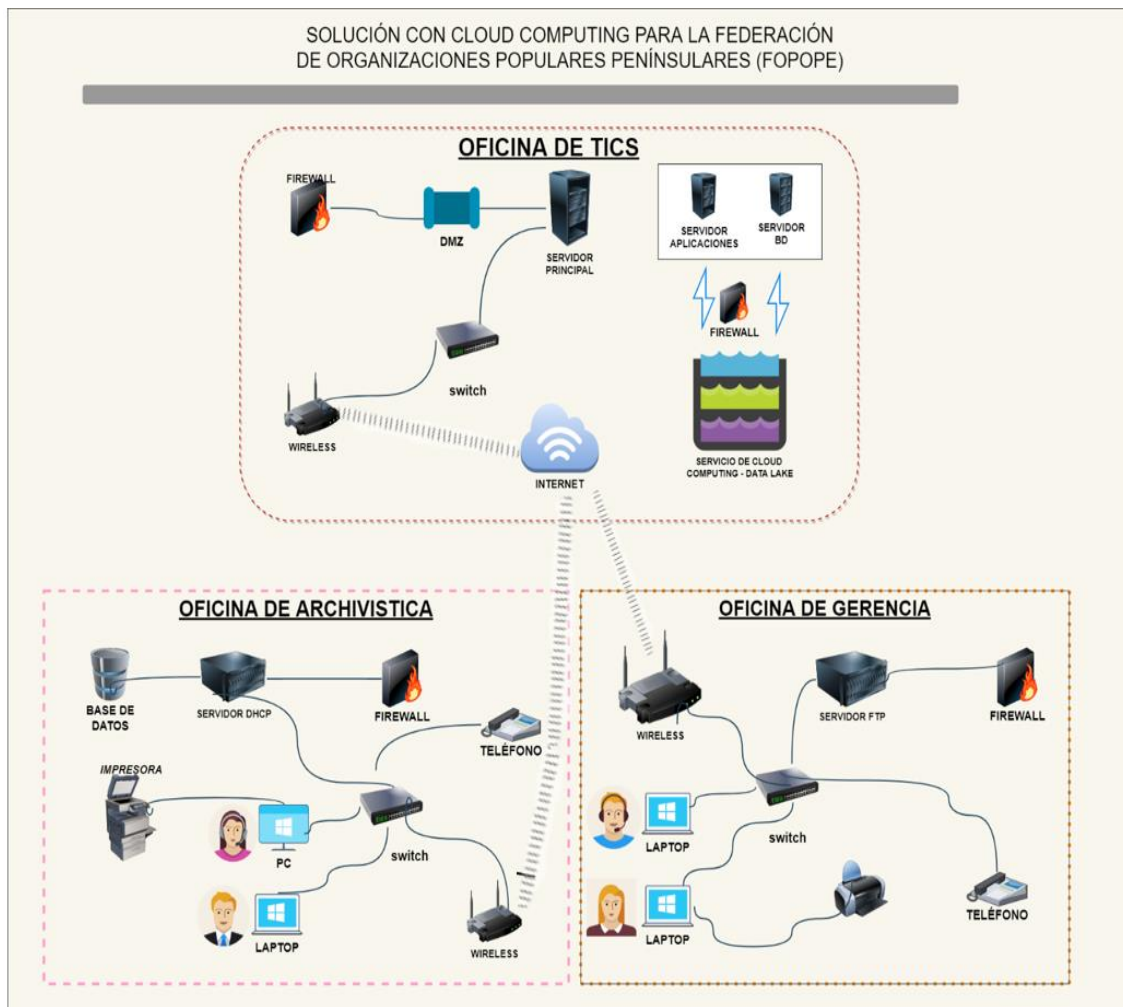


Arquitectura de la solución utilizando Data Lake Azure.

Tabla 14: Herramientas de la infraestructura de solución

1	Servidor	Servidor Dell PowerEdge T140 Xeon Quad Core 16gb 2tb DVD	Bueno	\$1479
1	Internet CNT Fibra Óptica	80 MBPS */ Velocidad simétrica	Bueno	\$37.99* incl. imp.
1	Servicio de nube	Data Lake Azure Storage Gen2	Bueno	\$89.28

Figura 10: Propuesta de solución con Cloud Computing



Se muestra el diseño de la red como solución contratando el servicio de Cloud Computing, donde alojan los servidores virtuales como el de Base de Datos y de Aplicaciones. La solución en la Nube estará trabajando en todo momento de manera que la función es respaldar automáticamente toda la información que maneje el servidor local cuando deje de funcionar el servidor físico.

El recurso para nuestro análisis se basa en la tecnología Cloud Computing Data Lake, la cual utilizará un proveedor de IAAS (Infraestructura como servidor). En base a las anteriores comparaciones de empresas que nos brindan el servicio de Cloud.

3.6. ESTUDIO DE FACTIBILIDAD.

Factibilidad Técnica

Consiste en determinar si los servicios de TI ofrecidos pueden ser compatibles o adaptables a los que ofrecen Data Lake en Cloud Computing y el grado de estabilidad que tendrán cuando estos sean implementados.

La disponibilidad de la infraestructura de la red tecnológica que se necesita es de 24/7 en un 100% los días laborables, en caso de la federación FOPOPE, por los múltiples proyectos que maneja en un determinado tiempo.

Aspectos Técnicos

Para realizar una comparación en ambiente Tradicional y en la Nube es importante considerar los siguientes requerimientos:

Tabla 15: Infraestructura Requerimiento

Tipo de Servidor	Observación	Descripción del Requerimiento
Programación Máquinas Virtuales	No Existente	Nuevo
Máquinas Virtuales	No Existente	Nuevo
Servidor de escritorio	Desarrollo Pruebas MAIA	Mejorar Capacidad de Almacenamiento actualmente 500 GB

Red	Topología estrella	Cambiar infraestructura a topología anillo
Base de datos	María DB	Mejorar la capacidad de almacenamiento actualmente 120 GB
Internet	CNT	Nuevo

Tabla 16: Arquitectura Lógica.

	Tecnología Cloud	Tradicional
Arquitectura	En la nube	Propia Infraestructura
Tecnología	Sql Azure	Microsoft SQL Server
Internet	Si	No
Aplicaciones	SI Azure	Sql

Tabla 17 Arquitectura Física

	Tecnología Cloud	Tradicional
Sistema Operativo	Windows 10	Windows 10
Fabricante	Microsoft	Microsoft
Licenciamiento	Renovar cada año	Por meses
Aplicaciones	Sql Azure	Sql
Versión	Actualizada	2013 beta

Tabla 18: Componentes de red

Componente de red y comunicaciones	Tecnología Cloud	Infraestructura propia
Router, Modem	Windows 10	Router, Switch, Access Point
Protocolos de comunicación	Ipv4	IPv4

Costos de Operación

Aquí se referencia el personal técnico que tiene la empresa, los costos en las arquitecturas y el tiempo que se va a demorar en la implementación insumos y materiales físicos que se utilice.

Tabla 19: Costo de Operación.

	Tecnología Cloud	Infraestructura propia
Insumos y Materiales Físicos	No se necesita estimar	Costo mensual
Recurso Humano	No se necesita estimar	Costo personal de soporte

Mantenimiento de Hardware y Software

En este apartado referenciamos el software y hardware que debe incluirse dentro del análisis técnico, esto nos servirá para identificar de manera entendible, los costos de Cloud versus la infraestructura propia.

Por lo general el personal delegado de la infraestructura y base tecnológica se hace cargo de verificar las conexiones mecánicas eléctricas, entre otros. Al mismo tiempo en equipamiento de hardware y software, en este sentido el departamento de desarrollo informa a infraestructura o arquitectura si va a realizar cambios o mantenimiento en base de datos.

Tabla 20: Manteniendo Hardware y Software

Hardware	Software	Tecnología Cloud
Temperatura	Actualización y mantenimiento de licencias	No se necesita estimar
Ventilación	Parche al software	No se necesita estimar
Cableado y Electricidad	Parches al sistema operativos	No se necesita estimar

Estado Físico del equipo	Parches al firewall/ antivirus	No se necesita estimar
Incremento de características	Conectividad con aplicativos externos	No se necesita estimar

Factibilidad técnica del sistema archivístico

Tabla 21: Factibilidad técnica sistema archivístico

Parámetros	Capacidad Instalada	Proveedor
Números de procesadores	3	X
Espacio de almacenamiento	450 GB	X
Memoria	8 GB	X
Sistema Operativo	Linux Debian versión 8	X
Base de datos	My Sql versión 5	X
Lenguaje de Desarrollo	Php versión 5.0	X
Nivel de Seguridad	-	Alta

Factibilidad técnica del servicio del portal web

Tabla 22: Factibilidad del servicio del portal web

Parámetros	Capacidad Instalada	Proveedor
Números de procesadores	3	X
Espacio de almacenamiento	100 GB	X
Memoria	8 GB	X
Sistema Operativo	Linux Debian versión 8	X
Base de datos	PostgreSQL versión 9.6	X
Lenguaje de Desarrollo	Laravel versión 5.4	X
Nivel de Seguridad	-	Alta

Factibilidad técnica del correo electrónico

Tabla 23: Factibilidad del servicio del correo electrónico

Parámetros	Capacidad Instalada	Proveedor
Números de procesadores	3	X
Espacio de almacenamiento	500 GB	X
Memoria	8 GB	X

Sistema Operativo	Linux Ubuntu 16.04	X
Base de datos	MySQL versión 5.0	X
Lenguaje de Desarrollo	Java versión 1.7	X
Nivel de Seguridad	-	Alta
Plataforma	Zimbra	-

Existe un nivel alto de factibilidad técnica en la adopción de los servicios debido a la compatibilidad de versiones de sus características de sus servicios, esto permite evaluar si determinado proyecto o nueva tecnología generara beneficios o pérdidas al negocio, reduciendo esfuerzo laboral del personal involucrado en la administración y operación de sus servicios e infraestructura de TI actual.

Factibilidad Operacional.

Permite en cierta manera, predecir si se debe poner en marcha la nueva solución, destacando los beneficios que ofrece en comparación con la producción de los actuales procesos realizados en la federación con respecto al procesamiento de datos dentro de la institución.

Al valorar la factibilidad operativa, se tomó como referencia del estudio de Grijalva [32], para eso se mantuvo conversación con el jefe de Tecnología de la Información de la federación. Y se manejó mediante la siguiente encuesta comparativa

Tabla 24: Encuesta de aceptación del proyecto

Pregunta	Solución Actual
¿En qué medida usted conoce el Data Lake?	En cierta medida
¿Qué servicios que ofrecen lago de datos de la nube conoce y piensa que sería más recomendable de usar?	Azure Data Lake
¿Piensa usted que sería favorable para la organización la migración de su información a la nube y el análisis en un Data Lake?	Si en un futuro
¿Cuáles considera que sean las principales motivaciones para la adopción de un Data Lake?	Modernización de los procesos Facilidad de análisis de los datos Flexibilidad y escalabilidad
¿Considera que, al usar estas nuevas tecnologías como lago de datos, se puede ahorrar tiempo y dinero utilizando estos servicios?	Si

¿Cuál es la preocupación que usted tendría sobre la implementación servicios en la nube de un Data Lake?	Insuficiente seguridad y confidencialidad Pérdida de control sobre los servicios Dependencia por parte del proveedor
¿Cuáles cree que sean los principales beneficios identificados al usar un Data Lake?	Ahorro en tiempos y costo Control adecuado de los datos Mejora en la productividad
¿En qué áreas de la organización usted considera que se utilizan más las soluciones de análisis de datos?	Comercial Archivística TIC Administrativa
¿Se realizan dentro de la organización capacitaciones sobre nuevas tecnologías?	No
¿Con que frecuencia se realizan las capacitaciones?	Pocas veces

A través de las encuestas realizadas a los cargos correspondientes mencionados anteriormente, se concluye que existe un apoyo de la directiva y del personal de la Federación De Organización Populares Peninsulares para que se lleve a cabo la implementación de este Estudio de Factibilidad para la adopción de esta guía de procesos de Cloud Computing en Data Lake; además existe un completo entendimiento sobre los beneficios que ofrece la Computación en la Nube, como la reducción de procesos administrativos como recursos y costos asignados a estos, también aumenta la disponibilidad, escalabilidad, rapidez y elasticidad de los servicios, dando como resultado un crecimiento rápido del negocio.

FACTIBILIDAD ECONÓMICA FINANCIERA

Se considera un análisis de costo de adquirir y operar la solución sugerida y se mantiene una comparación entre la plataforma actual y la nueva. Este análisis permite seleccionar el escenario más conveniente económicamente y viabilidad y rentabilidad para la federación.

Para realizar el análisis es imprescindible analizar de manera individual las soluciones y con la comparación de indicadores económicos como:

VAN (Valor Actual Neto).

TIR (Taza Interna de Retorno).

Para realizar una comparación en ambiente Tradicional y en la Nube es primordial tomar

en consideración los siguientes aspectos:

Costo de salario personal operativo

Tabla 25: Costo de salario personal operativo

Ingeniero en Sistemas	Mensual	Año
N°	1	1
Salario Ingeniero	\$500	\$6000
Total	\$500	\$6000
	Observación: los costos del personal se realizaron al departamento de TI y se efectuó en base al salario del 2021 como fecha del levantamiento de información y no incluye en la actualidad el aumento salarial nacional	

Costo de instalación y mantenimiento

Tabla 26: Costo de instalación y mantenimiento

Hardware	Año
Tiempo total de máquinas y servidor e impresoras	\$671
Total	\$671
Software	Año
Tiempo total de software en las maquinas físicas	\$235.60
Total	\$235.60
Observación:	Se realiza limpieza de los componentes internos tanto para las computadoras, impresoras y servidor. De la misma manera se estima la instalación de nuevo software en cada hardware mencionado anteriormente y los costos de cada uno se obtuvo anualmente debido a que estos costos de mantenimientos se realizan anual o cada tres años respectivamente.

Costos de consumo Electricidad

Tabla 27: Costo de instalación y mantenimiento

	Mensual	Año
Consumo electricidad	\$120	\$1440
Total	\$120	\$1440

	Observación: los costos del consumo de energía es un aproximado mensual debido a que diferentes meses varía el precio a más o menos el valor anterior.
--	--

Costos suplementarios

Tabla 28: Costo suplementario

	Mensual	Año
Mantenimiento	\$250	\$3000
Licenciamiento	\$32,85	\$394,20
Total	\$282,85	\$3394,20
	Observación: los costos de licenciamiento abarcan los de office, antivirus, y sistema operativo por un año, los cuales se renuevan anualmente con el mismo precio, o dependiendo de la mejora del servicio o infraestructura que se necesita al pasar del tiempo en la federación.	

Costo de Operación anual

Tabla 29: Costo de operación anual

Ingeniero en Sistemas	Mensual	Año
N°	1	1
Costo de salario personal operativo	\$500	\$6000
Costo de instalación y mantenimiento	\$75,30	\$903.6
Costos de consumo Electricidad	\$120	\$1440
Costos suplementarios	\$282.85	\$3394.20
Total	\$983.28	\$11737.80

Valor Neto Actualizado VAN.

Al trabajar con el análisis del VAN se compara en los ambientes tradicionales y de Cloud Computing. Para que nuestro análisis resulte satisfactorio el cálculo de VAN del escenario de Cloud Computing debe ser menor para que exista beneficio en el departamento de TI de la federación. Es decir, al beneficio que hemos obtenido de una inversión (o que planeamos obtener) le restamos el costo de inversión realizada. Luego eso lo dividimos entre el costo de la inversión y el resultado.

Para el cálculo del VAN se considera principalmente el presupuesto para el área

informática anual con el flujo de caja para cada año el número de años es 3. El presupuesto anual es de \$9700, con un incremento del 10% anual. Los egresos son el costo por mes de Windows Azure \$ 89,28 x 12 = 1071,36 este valor es cada año.

$$VAN = -I + \sum_{n=1}^N \frac{Q_n}{(1+r)^n}$$

Valor Neto Actualizado VAN escenario tradicional.

Flujo de Caja Infraestructura Federación

Tabla 30: Flujo de caja infraestructura federación

Flujo	Inversión Inicial	Año 1	Año 2	Año 3	Año 4	Año 5
Ingresos		\$9,700	106.70	\$117.37	\$ 12,910.7	\$ 14,201.77
Egresos	\$9700	\$638.16	\$667.12	\$704.14	\$ 749.22	\$ 802.36
TOTAL	\$9700	\$10338.16	\$11337.12	\$12441.14	\$13,704.92	\$15,004.13

Van = \$31,420.46

Valor Neto Actualizado VAN escenario Cloud Computing.

Flujo de Caja para Solución en la Nube

Tabla 31: Flujo de caja solución en la nube

Flujo	Año 1	Año 2	Año 3	Año 4	Año 5
Ingresos	\$9,700	10,670	\$11,737	\$12,910.7	\$14,201.77
Egresos	\$1,071.36	\$1,071.36	\$10,71.36	\$,1071,36	\$1,071,36
TOTAL	\$10,771.36	\$11,741.36	\$12,808.36	\$13,982.06	\$15,273.13

Van = \$30,029.61

Tasa Interna de Retorno (TIR)

Es un promedio de los futuros rendimientos esperados de la inversión, en términos generales o comunes se conceptualiza como el valor neto igual a cero.

Estos flujos de caja no tienen por ser constantes. Sin embargo, los flujos de caja deben ocurrir en intervalos regulares, como meses o años. La tasa interna de retorno equivale a la tasa de interés producida por un proyecto de inversión con pagos (valores negativos) e ingresos (valores positivos) que se producen en períodos regulares

$$VAN = \sum_{t=1}^N \frac{V_F T}{(1 + TIR)^t} - I_0 = 0$$

SOLUCIÓN EN LA NUBE

Por consecuencia a la indagación se realizó una colación con Windows Azure por no tenerlas todas consigo mayor beneficio con la empresa por las diferentes habilidades de operar en el manejo de sus servicios.

En la calculadora de precios de Microsoft Azure encontramos el servicio que vamos a utilizar dando el siguiente cuadro.

Estimación de costos del servicio de Data Lake Storage Gen2 en Windows Azure

Tabla 32: Azure calculadora de precios.

MICROSOFT AZURE					
ESTIMACIÓN					
Tipo De Servicio	Nombre personalizado	Región	Descripción	Costo Mensual Estimado	Costo Inicial Estimado
Cuentas De Almacenamiento		Central US	Data Lake Storage Gen2, estándar, redundancia LRS, nivel de acceso dinámico, estructura de archivo de espacio de nombres plana, capacidad de 500 GB: pago por uso, operaciones de escritura: 4 MB x 100 000 operaciones, operaciones de lectura: 4 MB x 100 000	\$60,28	\$0,00

			operaciones, 100 000 operaciones de lectura iterativa, 100.000 Operaciones de escritura iterativa, 100.000 Otras operaciones. Recuperación de datos de 1000 GB, escritura de datos de 1000 GB		
Soporte			Soporte	\$29,00	\$0,00
			Licencia de Programas	Contrato de cliente de Microsoft (MCA)	
			Perfil De Facturación		
Descargo de Responsabilidad			Total	\$89,28	\$0,00

Se identifican todos los servicios que ofrecen el servicio de Data Lake en Azure, cuyo Se identifican todos los servicios y plataformas que proporcionan el servicio de Data Lake en Azure, cuyo costo anual están incluidos el soporte de licencias y varios beneficios que contiene esta promoción, siendo ventajoso para la federación para sus actividades de almacenamiento y procesamiento de datos.

Podemos observar la tabla donde se detalla el costo por año de la solución en la nube
Costo

Costo total solución en la nube.

Se presenta los costos por tres años en la adquisición de este servicio.

Tabla 33: Costo total solución en la nube.

Producto y Servicio	Costo por mes	Costo Año 1	Costo Año 2	Costo Año 3	Costo total
Windows Azure Data Lake	\$89.28	\$1071.36	\$1071.36	\$1071.36	\$3214.08
TOTAL	\$89.28	\$1071.36	\$1071.36	\$1071.36	\$3214.08

Calculo TIR.

Se obtiene la tasa interna de retorno realizando una comparación de servicio de la nube con la comparativa de la infraestructura de la federación.

Tabla 34: Calculo TIR

	Cloud Computing	Infraestructura Propia
VAN	13790.23	14794.03
Ahorro proyecto	1003.80	
TIR (10%)	79%	TIR>10%

El resultado del cálculo es mayor al índice inicial, cuyo porcentaje es el 10% por tanto se considera aceptable el proyecto, la razón, que el proyecto da una rentabilidad mayor que la rentabilidad mínima requerida.

3.1.- RESULTADOS

Pruebas

Elaboración de un data lake en Microsoft Azure

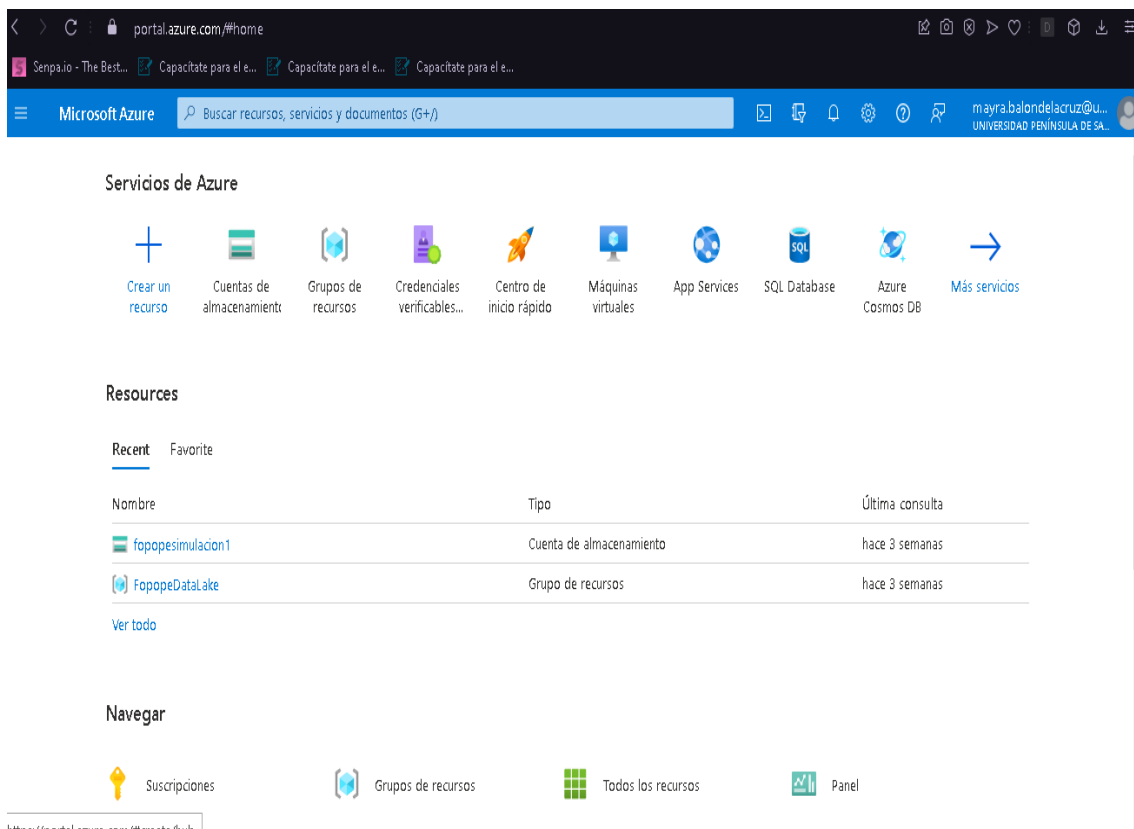
Entre los primeros pasos a seguir para la creación de un data lake es el registro en el portal Azure, con este proveedor nos proporciona un saldo de \$200 por un mes.

Figura 11: Página principal Portal Azure



Una vez registrado y logeado nos ubicamos en el Portal de Azure donde encontramos las herramientas a trabajar.

Figura 12: Visualización Portal



Creación de un recurso

Figura 13: Creación de grupos en el portal de Azure

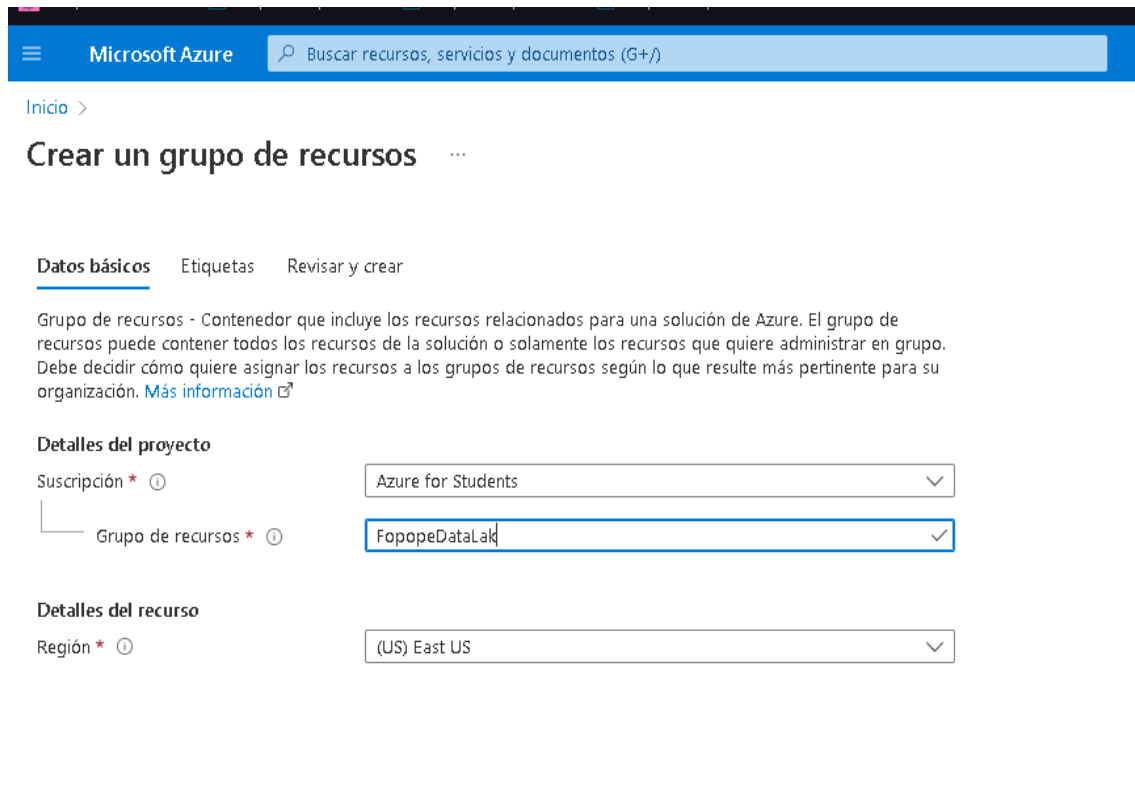
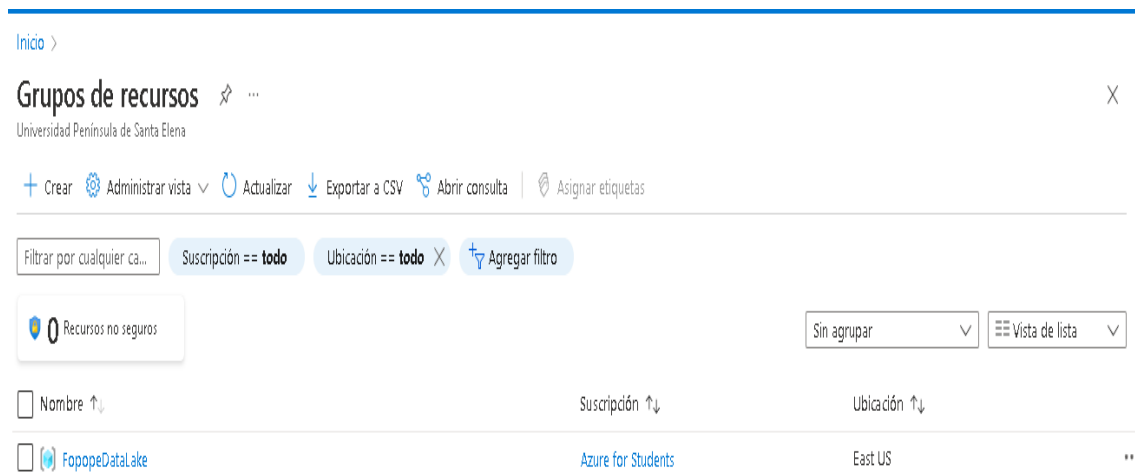


Figura 14: Visualización de grupos en el portal de Azure.



Creación de una cuenta de almacenamiento.

Figura 15: Creación de cuenta de almacenamiento en el Portal Azure.

[Inicio](#) > [Cuentas de almacenamiento](#) >

Crear una cuenta de almacenamiento ...

Datos básicos Opciones avanzadas Redes Protección de datos Cifrado Etiquetas Revisar y crear

Azure Storage es un servicio administrado por Microsoft que proporciona almacenamiento en la nube altamente disponible, seguro, duradero, escalable y redundante. Azure Storage incluye Azure Blob (objetos), Azure Data Lake Storage Gen2, Azure Files, Azure Queues y Azure Tables. El costo de una cuenta de Storage depende del uso y de las opciones que elija a continuación. [Más información sobre las cuentas de almacenamiento de Azure](#)

Detalles del proyecto

Selección la suscripción en la que se creará la nueva cuenta de almacenamiento. Elija un grupo de recursos nuevo o uno ya existente para organizar y administrar la cuenta de almacenamiento junto con otros recursos.

Suscripción *

Grupo de recursos *
[Crear nuevo](#)

Figura 16: Visualización Portal

Microsoft Azure | Buscar recursos, servicios y documentos (G+)

Inicio >

Cuentas de almacenamiento

Universidad Península de Santa Elena

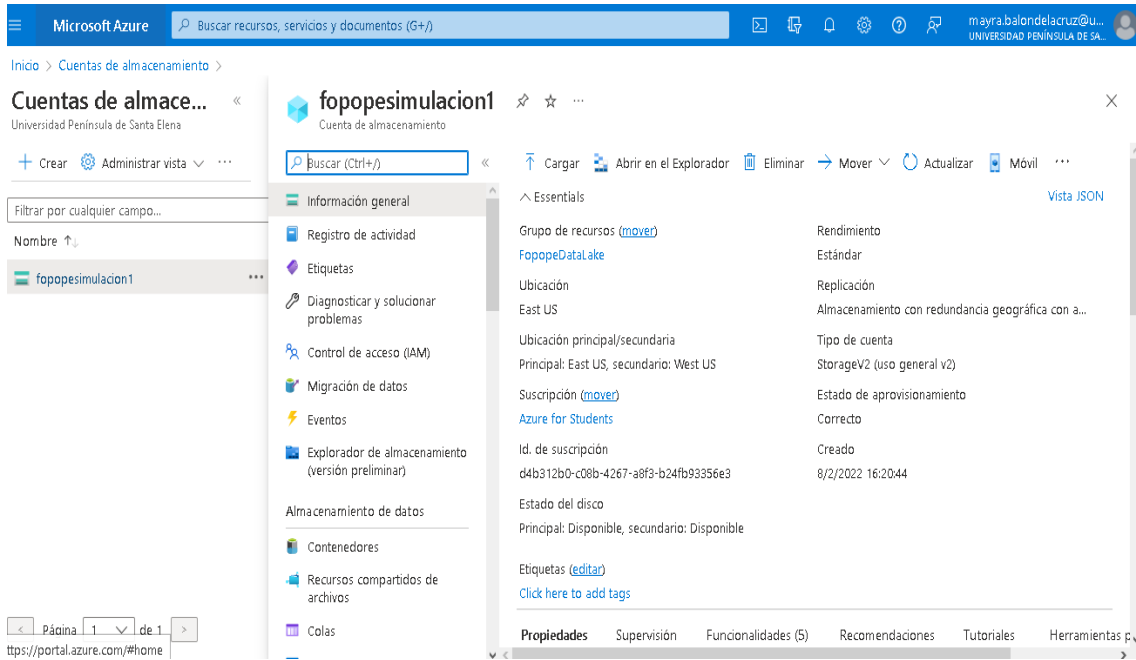
+ Crear | Administrar vista | Actualizar | Exportar a CSV | Abrir consulta | Asignar etiquetas | Eliminar

Filtrar por cualquier campo | Suscripción == todo | Grupo de recursos == todo X | Ubicación == todo X | Agregar filtro

Sin agrupar | Vista de lista

Nombre ↑↓	Tipo ↑↓	Variante ↑↓	Grupo de recursos ↑↓	Ubicación ↑↓	Suscripción ↑↓
fopopesimulación1	Cuenta de almacenamiento	StorageV2	FopopeDataLake	East US	Azure for Students

Figura 17: Visualización cuenta de almacenamiento en el Portal



Creación de contenedores

Figura 18: Creación Contenedores Portal

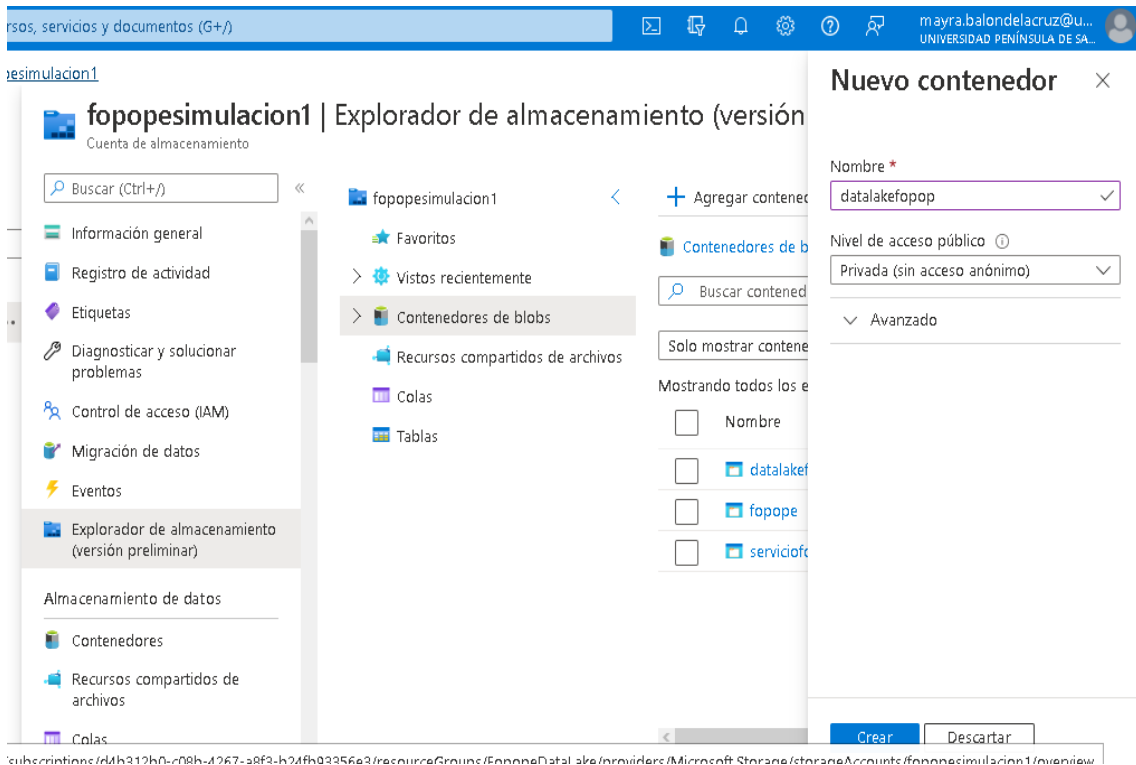
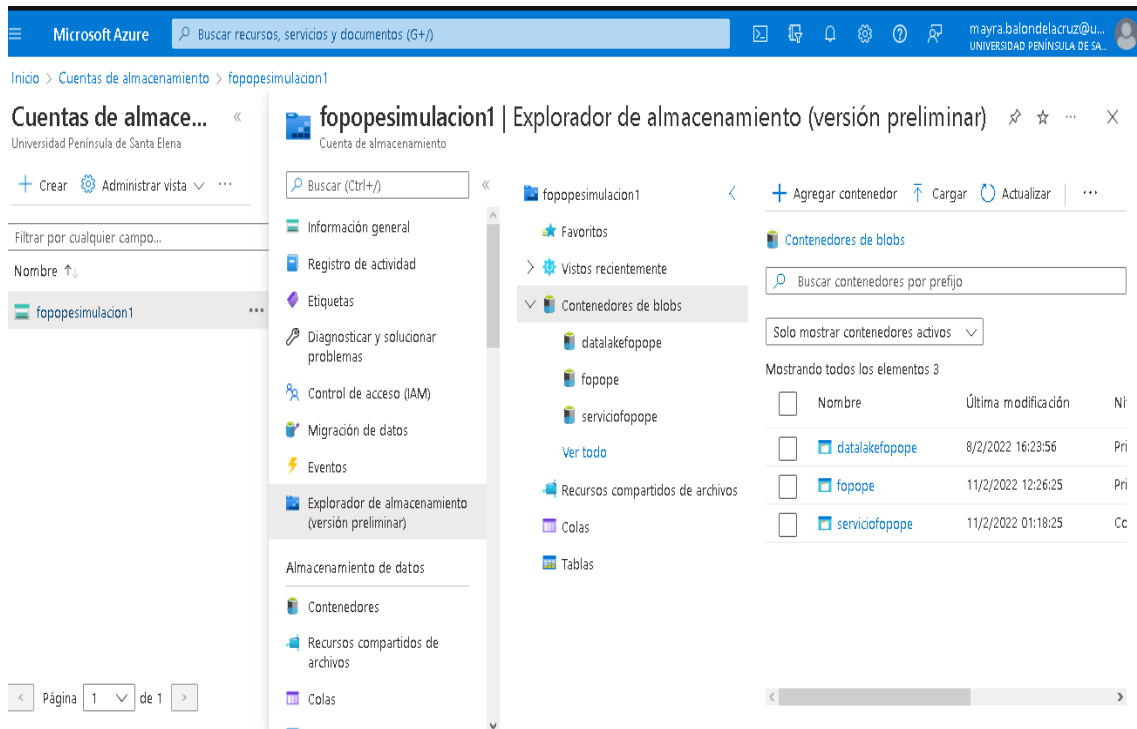
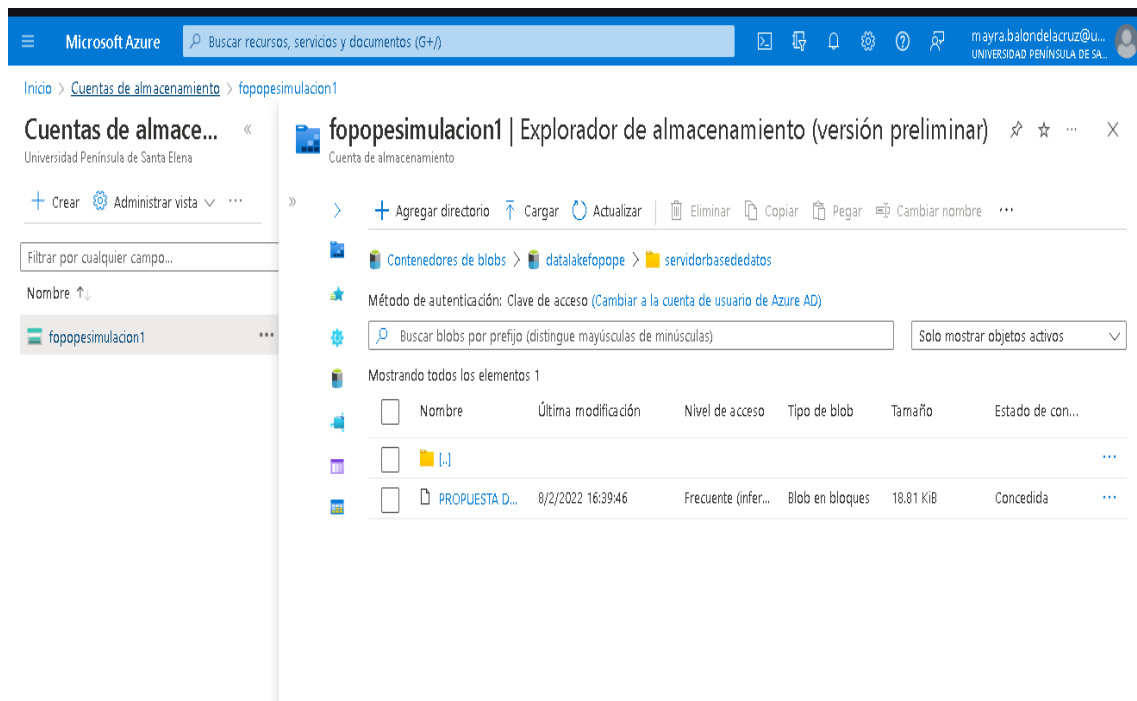


Figura 19: Visualización de contenedores en el Portal de Azure



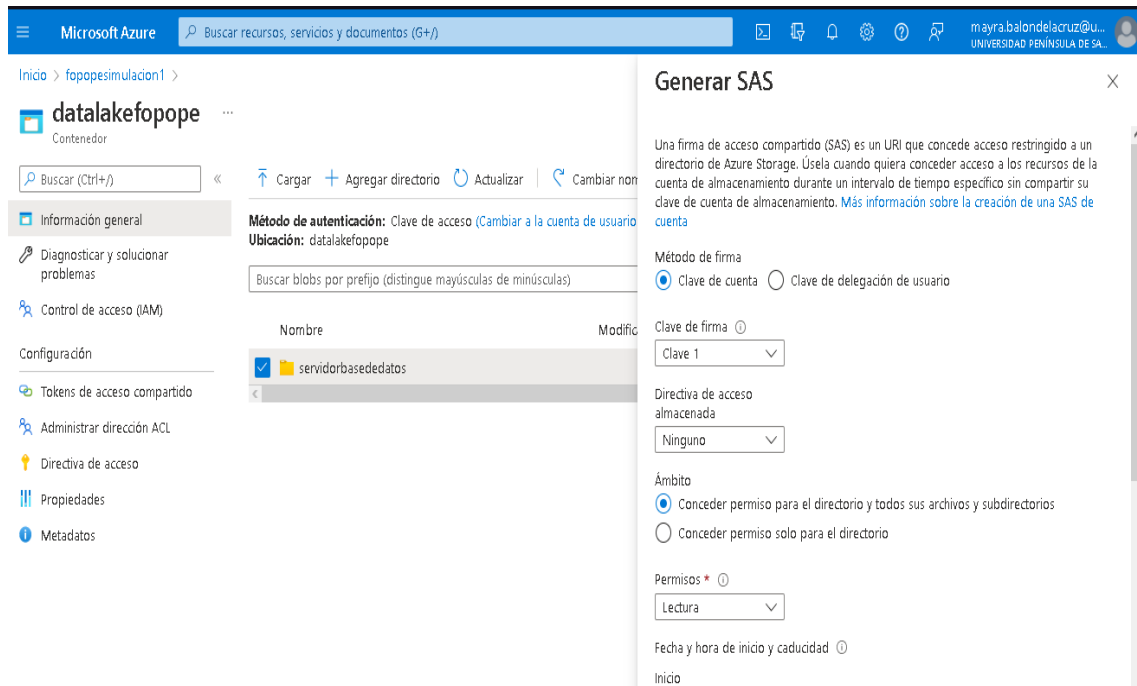
A partir de este punto se visualiza un sistema de archivo donde se puede almacenar todos los archivos documentos datos estructurados semiestructurados y no estructurados.

Figura 20: Visualización del explorador de almacenamiento



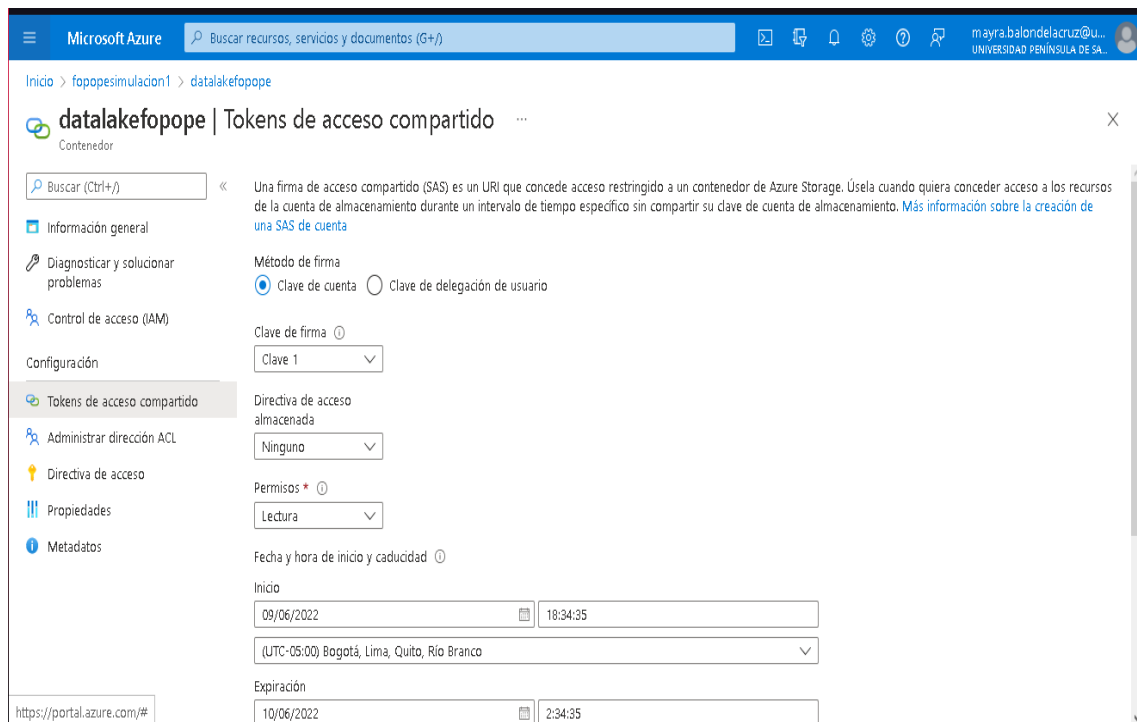
Se configura la firma de acceso compartido o restringido según dado el caso de los administradores.

Figura 21: Generación de firmas de acceso SaaS Portal Azure



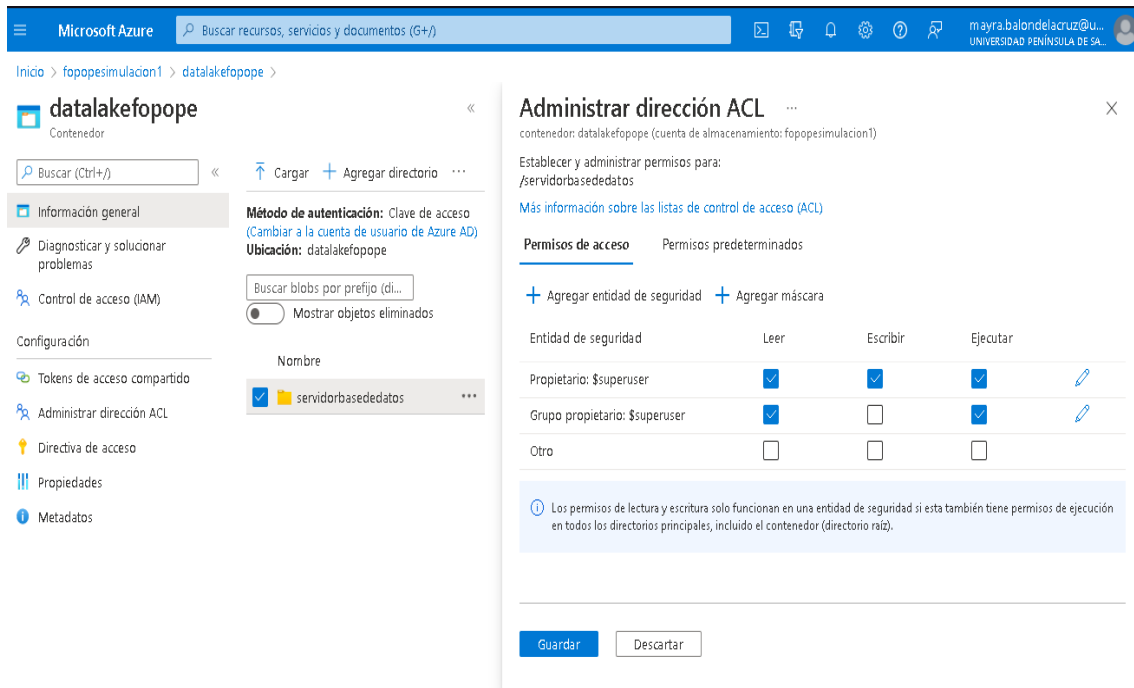
Configuración de Token de Acceso compartidos.

Figura 22: Generación de tokens de acceso compartido Portal Azure



Con figuramos los permisos de ACL como lectura, escritura y ejecución.

Figura 23: Generación de accesos y permisos ACL en el Portal Azure



Construcción un Data Lake en Google Cloud.

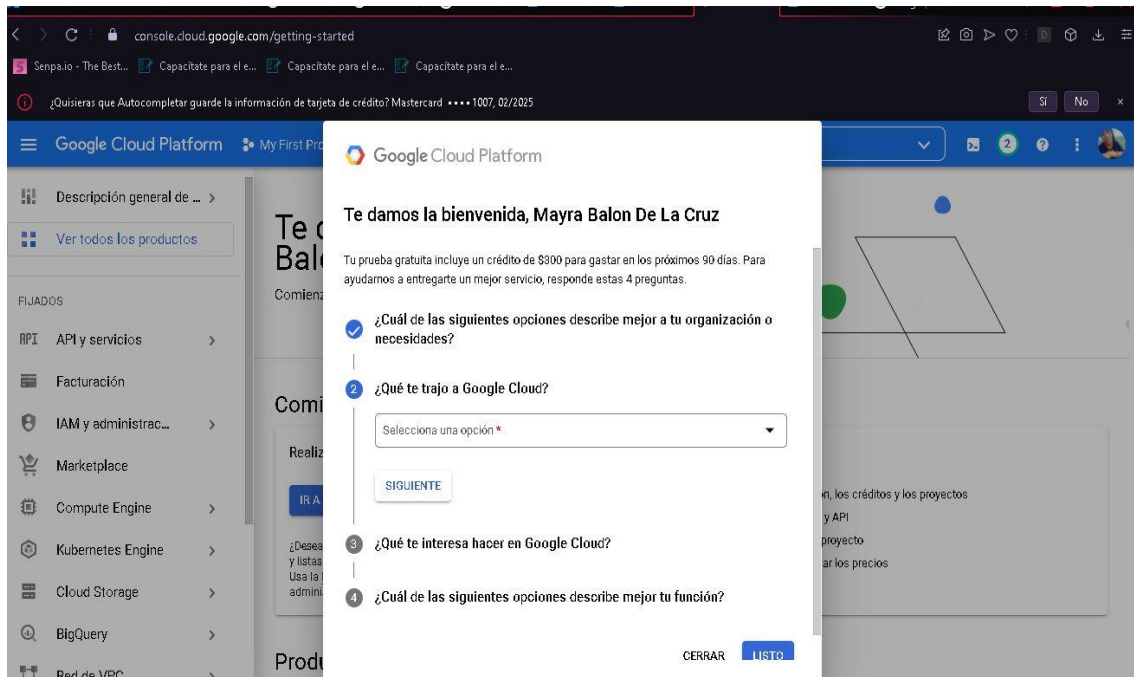
Mediante este proveedor nos proporciona \$300 por 20 productos en un lapso de dos meses

Figura 24: Pagina de Google Cloud Platform



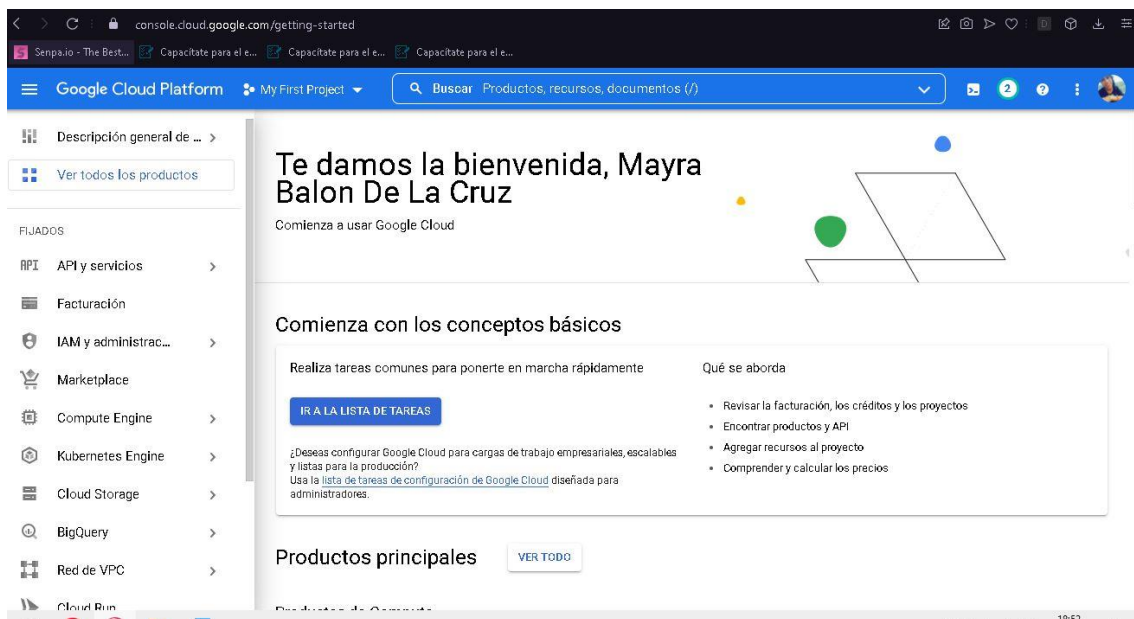
Registro con la cuenta d Google

Figura 25: Registro de la Pagina de Google Cloud Platform



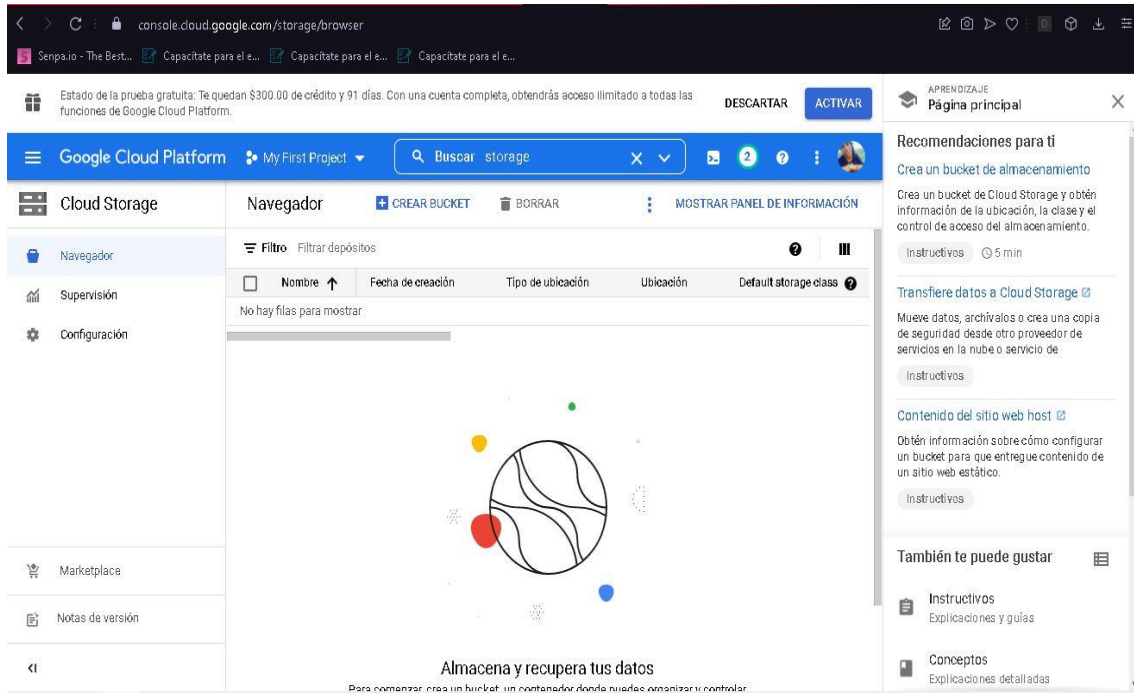
Panel de Google Cloud Platform

Figura 26: Panel principal de GCP



Abrir la opción de Cloud Storage

Figura 27: Panel de Google Cloud Storage



Creación un depósito

Figura 28: Creación de depósito en Google Cloud Storage

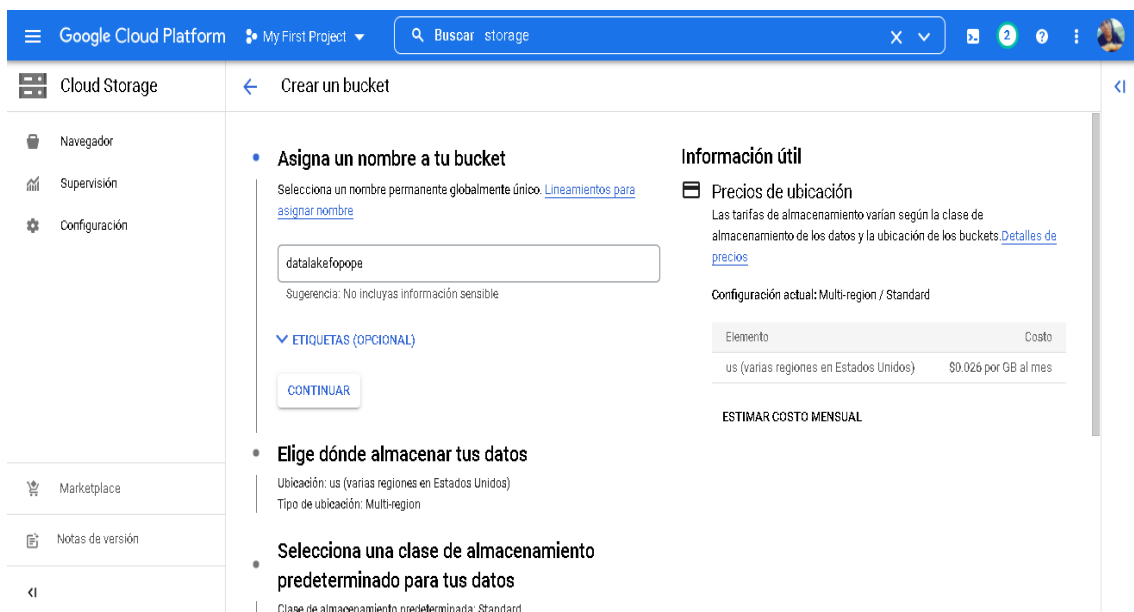
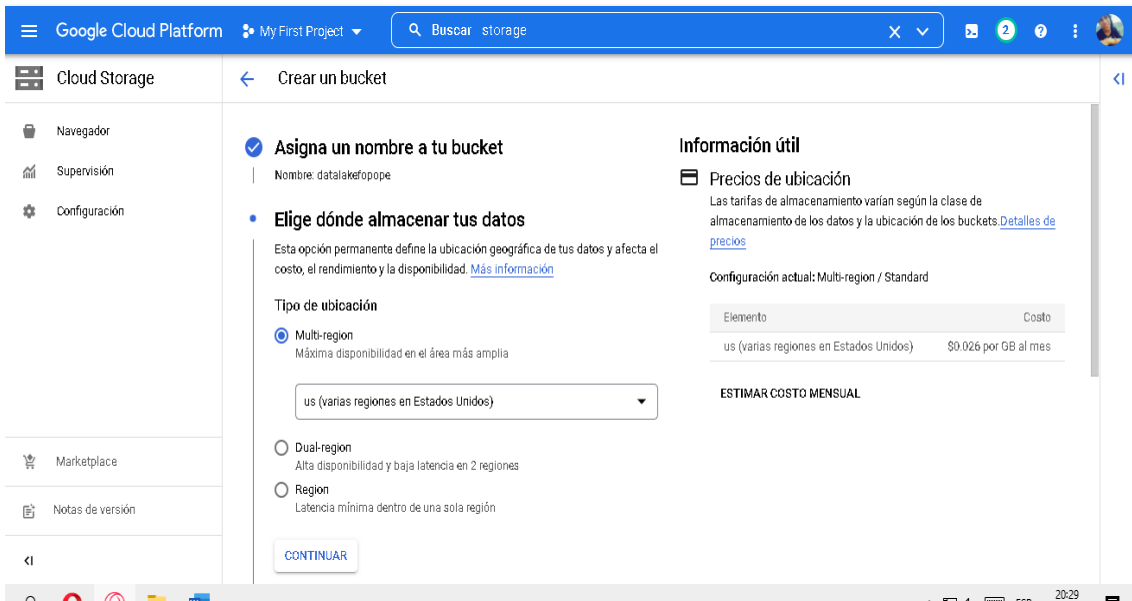
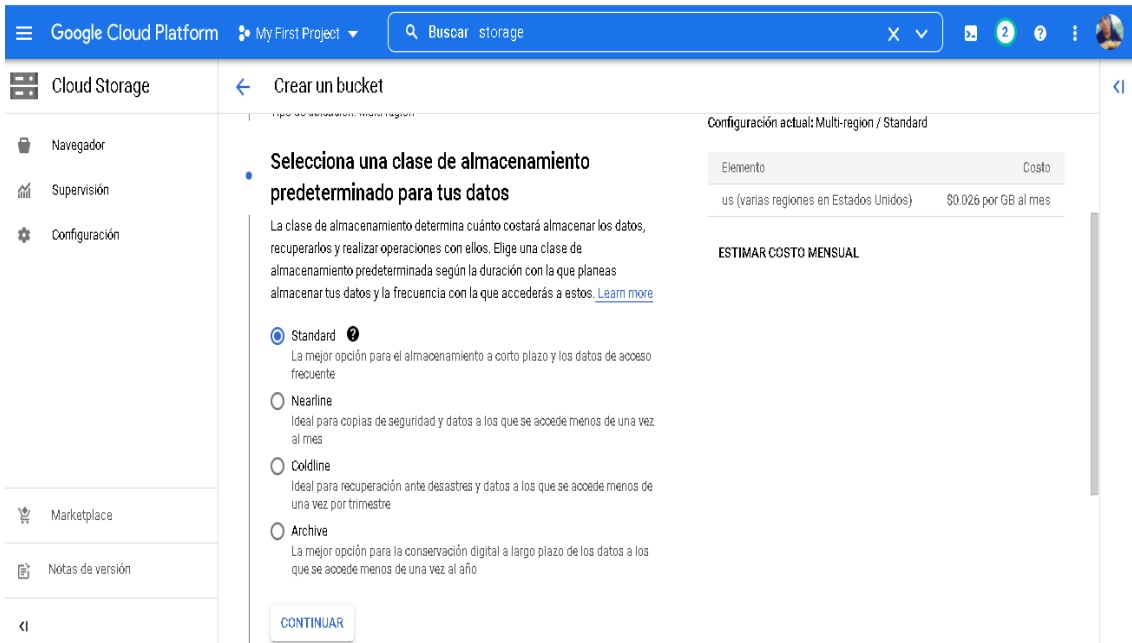


Figura 29: Configuración de depósito en Google Cloud Storage



Selección de una clase de almacenamiento Estándar.

Figura 30: Configuración de clase de almacenamiento en Google Cloud Storage



Control de acceso a los objetos

Figura 31: Configuración de control de acceso de objetos en Google Cloud Storage

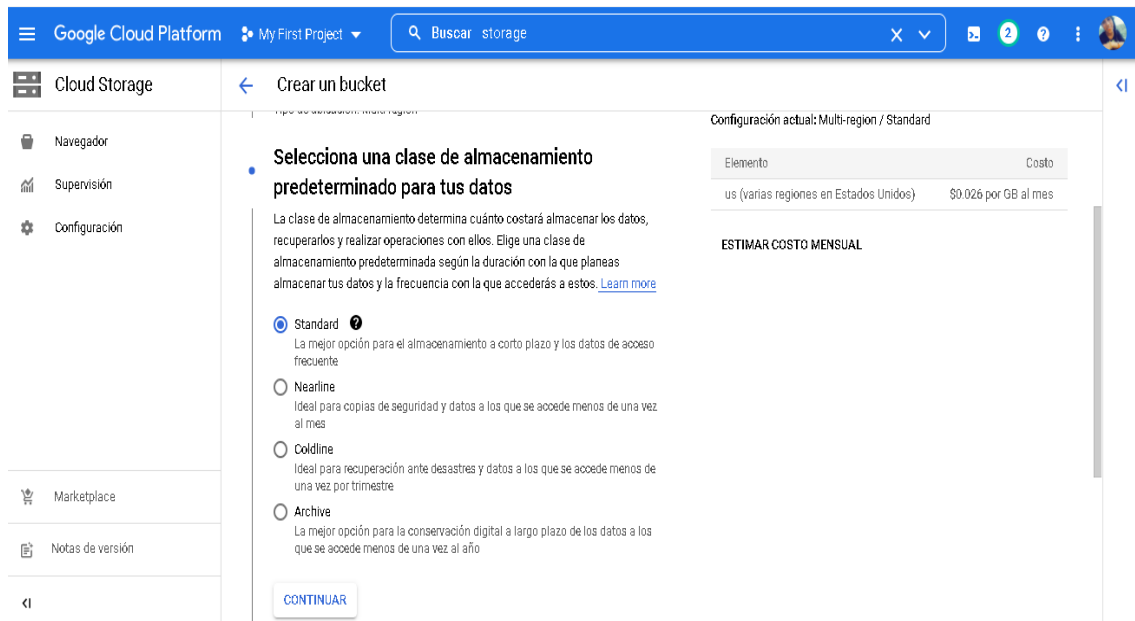
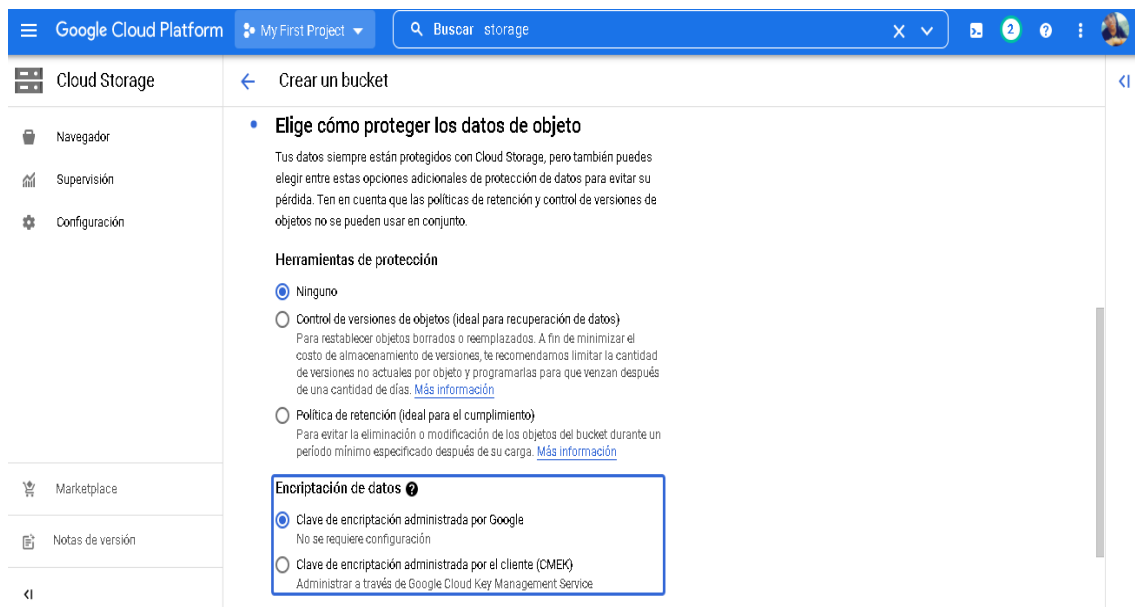
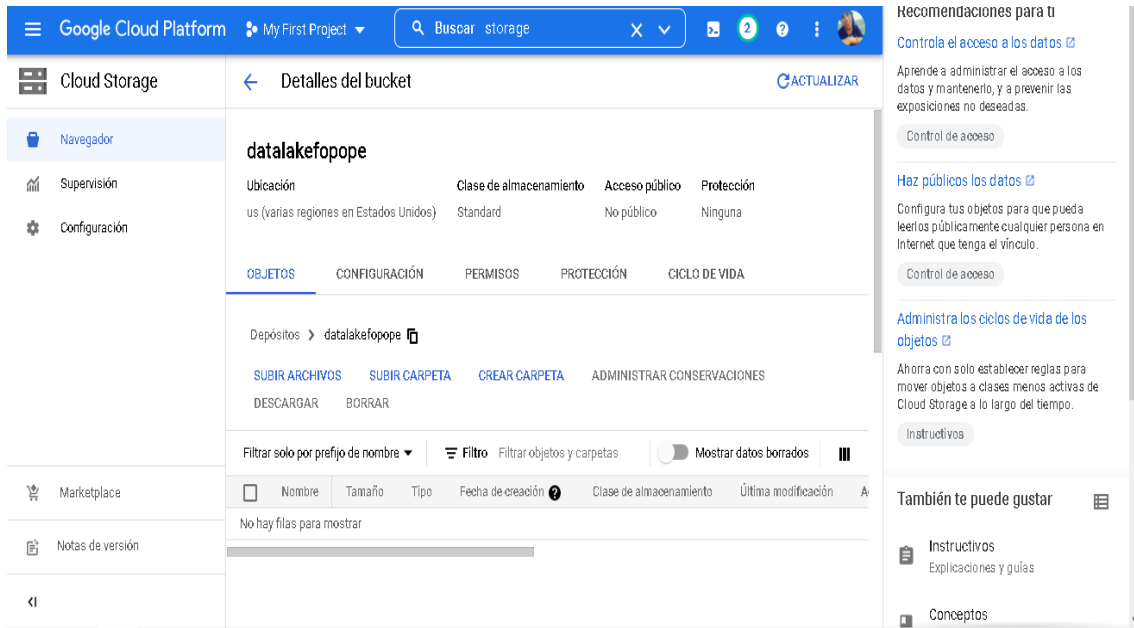


Figura 32: Configuración de encriptación en Google Cloud Storage



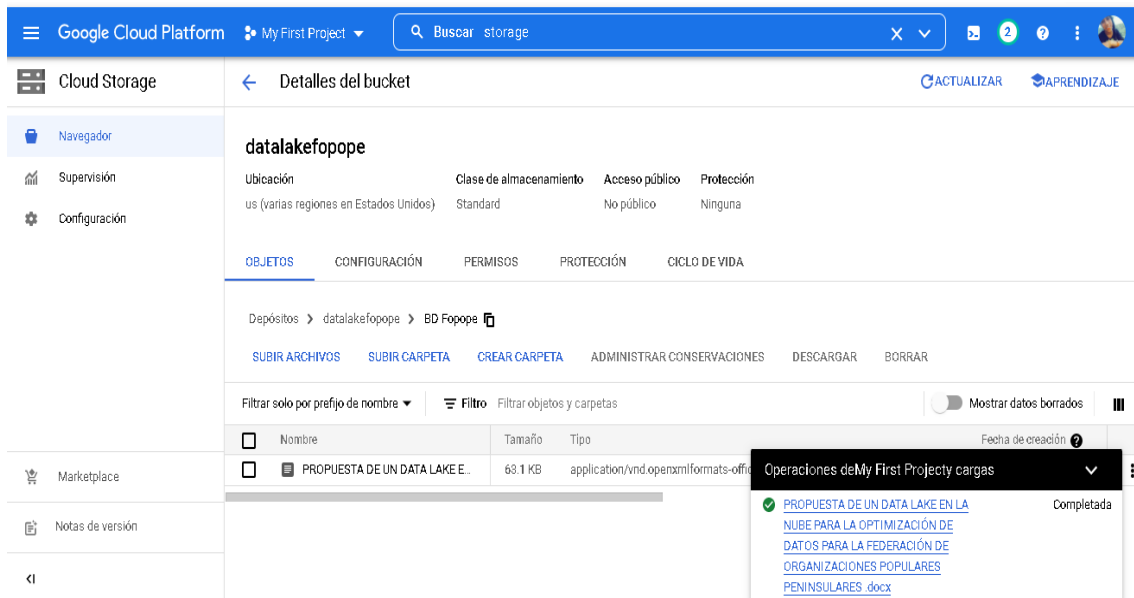
A partir de este paso funciona como un sistema de archivos donde se podrá almacenar manipular los datos de una manera ordenada.

Figura 33: Visualización del Bucket creado en Google Cloud Storage



Almacenamiento de los datos dentro del contenedor

Figura 34: Almacenamiento de datos dentro del Bucket creado en Google Cloud Storage



Configuración de seguridad para el Bucket creado

Figura 35: Configuración de permisos para el Bucket creado en Google Cloud Storage

The screenshot shows the 'Detalles del bucket' page in the Google Cloud Platform console. The left sidebar contains navigation options: 'Navegador', 'Supervisión', 'Configuración', 'Marketplace', and 'Notas de versión'. The main content area is titled 'Detalles del bucket' and includes a search bar with 'storage'. Below the search bar, there are two informational boxes: one about public sharing and another about uniform permissions. A 'QUITAR LA PREVENCIÓN DE ACCESO PÚBLICO' button is visible. The 'Permisos' section is active, showing a list of permissions for the bucket. The permissions are categorized by 'Función/Principal' and 'Nombre'. The table below shows the current permissions:

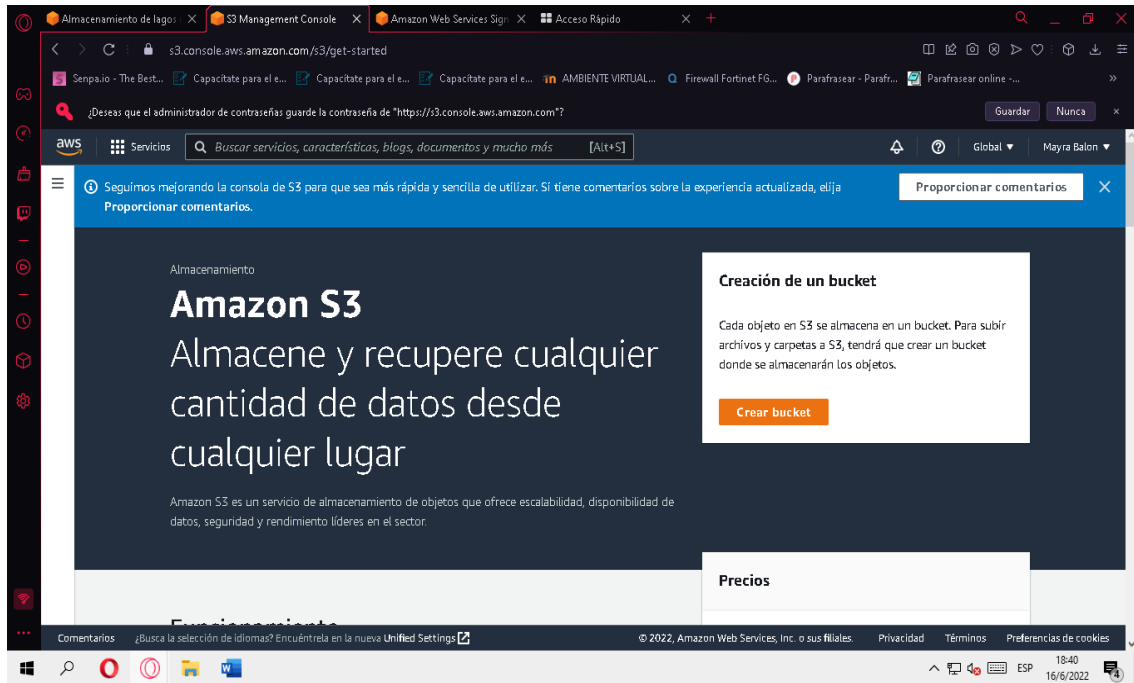
Función/Principal	Nombre	Herencia
<input type="checkbox"/>	Lector de buckets heredados de almacenamiento (1)	
<input checked="" type="checkbox"/>	Visualizadores del proyecto: focus-flight-352923	
<input type="checkbox"/>	Propietario de buckets heredados de almacenamiento (2)	
<input checked="" type="checkbox"/>	Editoras del proyecto: focus-flight-352923	
<input checked="" type="checkbox"/>	Propietarios del proyecto: focus-flight-352923	

Figura 36: Configuración de la protección para el Bucket creado en Google Cloud Storage

The screenshot shows the 'Detalles del bucket' page in the Google Cloud Platform console, specifically the 'PROTECCIÓN' tab. The left sidebar is the same as in Figure 35. The main content area is titled 'Detalles del bucket' and includes a search bar with 'storage'. Below the search bar, there are tabs for 'CONFIGURACIÓN', 'PERMISOS', 'PROTECCIÓN', and 'CICLO DE VIDA'. The 'PROTECCIÓN' tab is active, showing the 'Control de versiones de objetos (ideal para recuperación de datos)' section. This section is currently active, as indicated by a green checkmark and the text 'CONTROL DE VERSIONES DE OBJETOS ACTIVADO'. Below this, there is a warning message: 'No tienes reglas de ciclo de vida que se apliquen a versiones no actuales', with an 'ADMINISTRAR REGLAS' button. The 'Política de retención (ideal para el cumplimiento)' section is also visible, but its content is partially obscured by the browser's address bar.

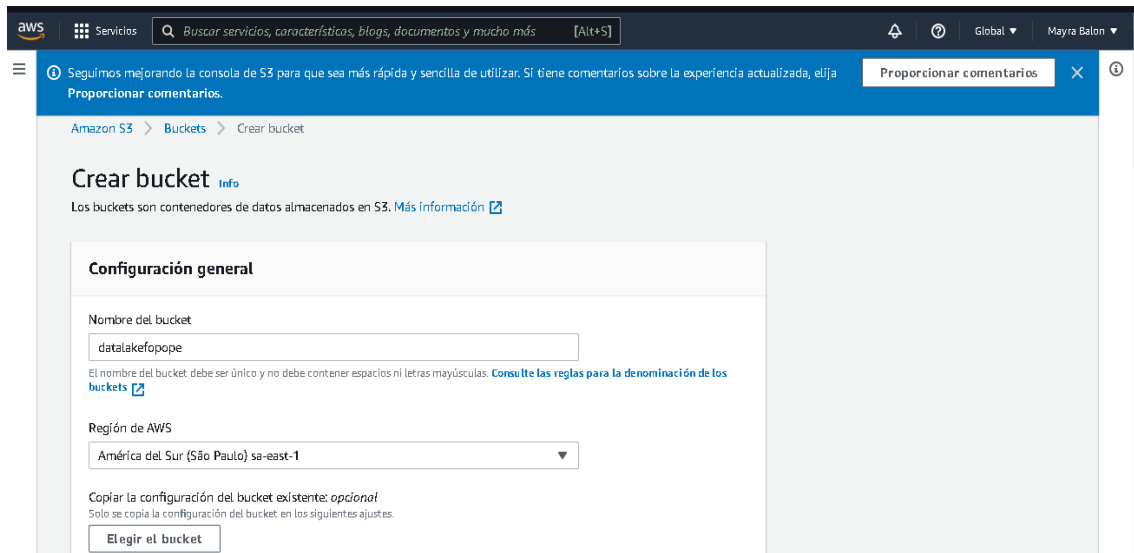
Vista de acceso de la consola de Amazon S3

Figura 37: Página Principal de AWS



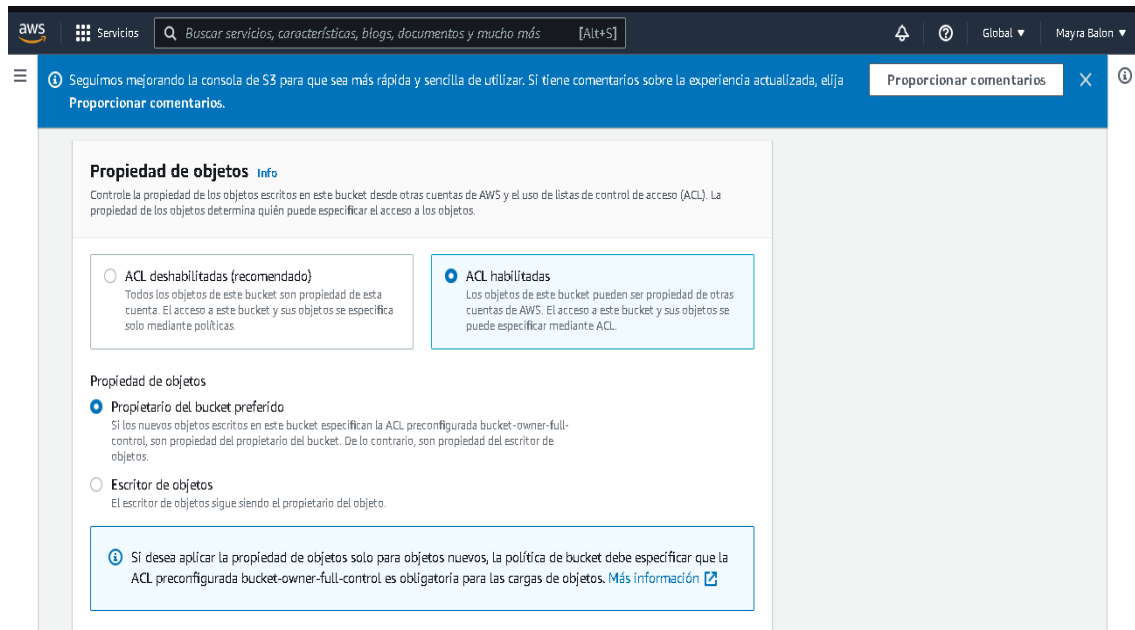
Creación de Bucket

Figura 38: Creación de Bucket dentro de la plataforma



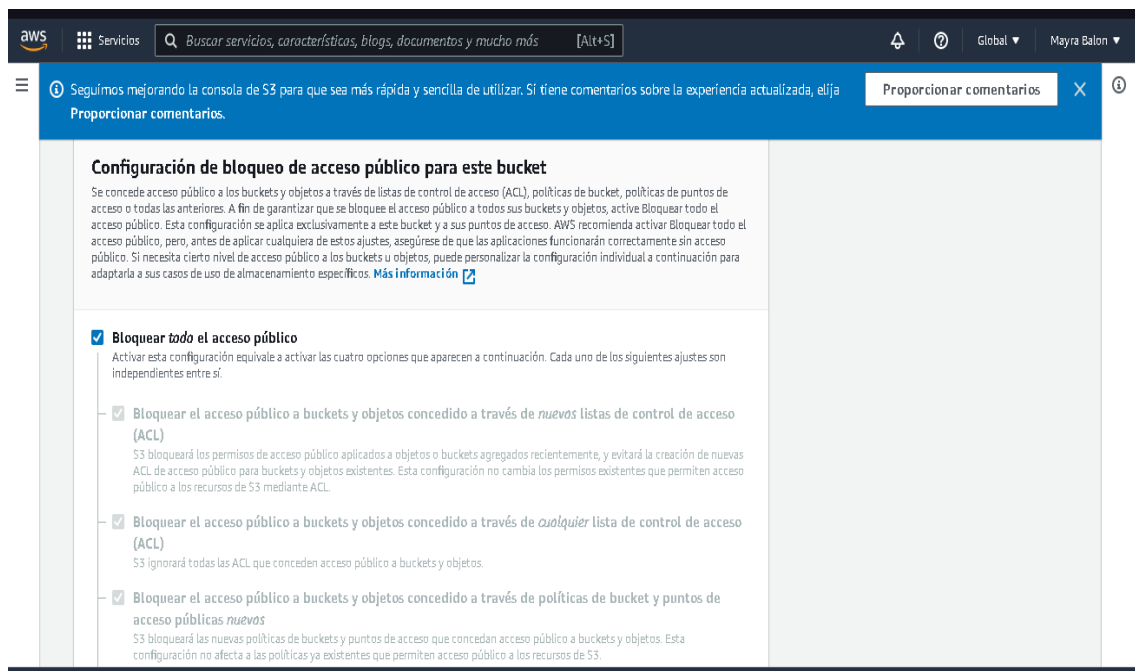
Configuración de las propiedades de objetos

Figura 39: Panel de configuración de objetos de AWS



Configuración de seguridad y bloqueo

Figura 40: Panel de configuración de bloqueo de acceso público para Bucket



Control de versiones

Figura 41: Panel de configuración y control de versiones de Bucket

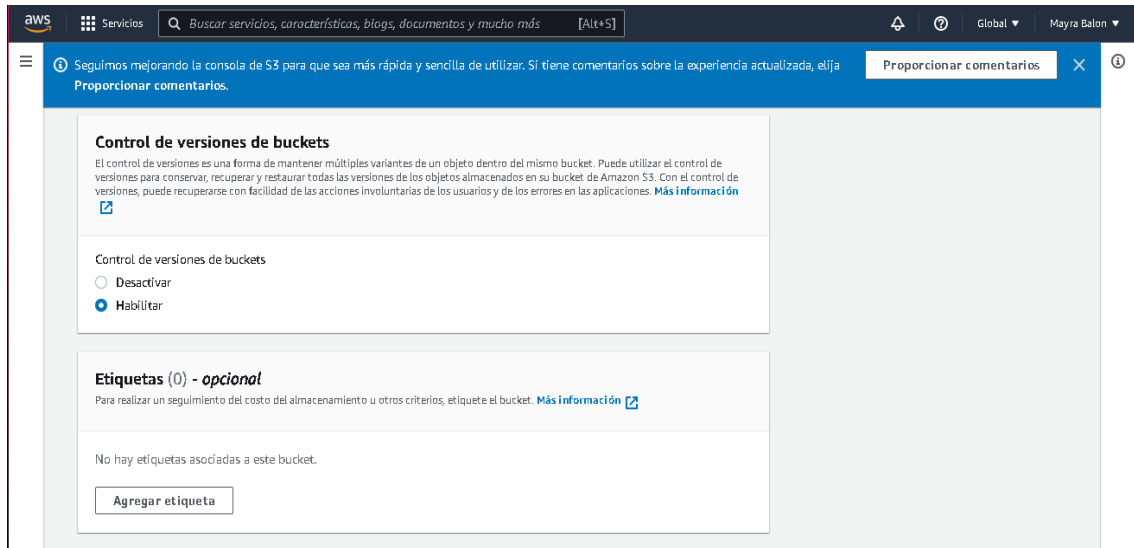
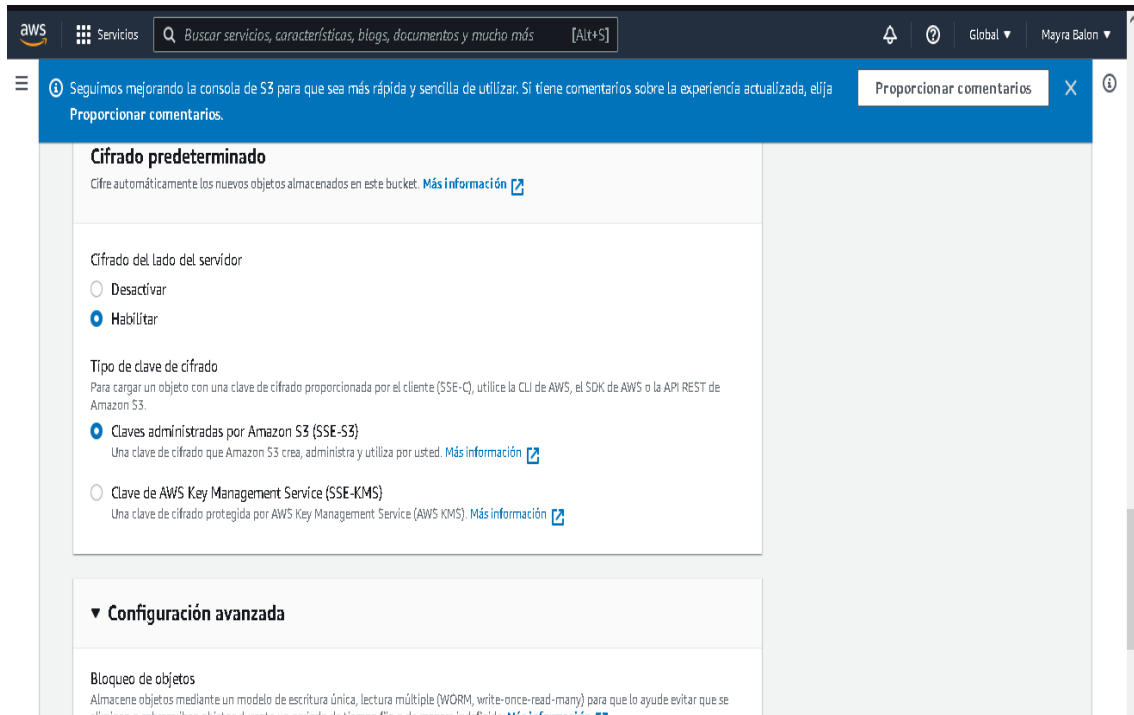
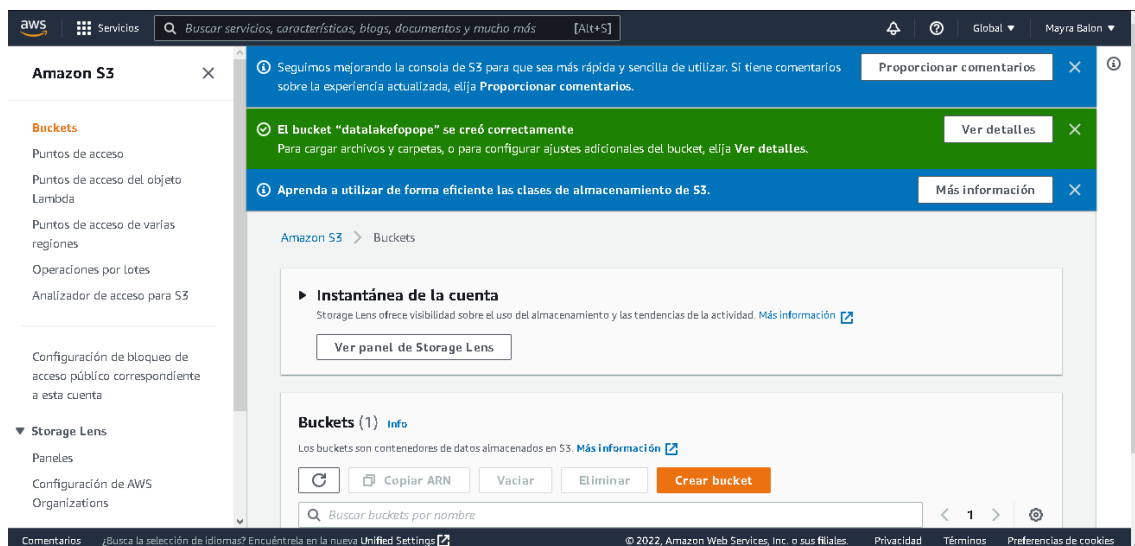


Figura 42: Configuración en el panel de cifrados predeterminados



Vista del Bucket creado

Figura 43: Visualización de la creación de Bucket



Dado que la cantidad de datos que guardan las empresas aumenta año tras año, no sorprende que las empresas se estén alejando de la subsistencia de los servidores locales tradicionales y estén invirtiendo en soluciones de almacenamiento en la nube. Con total flexibilidad y escalabilidad, sabiendo que el mantenimiento es responsabilidad del propietario de la nube, ofrece una solución simple adecuada para el futuro.

La competencia en la industria del almacenamiento en la nube en la última década ha llevado a que algunos grandes nombres ingresen al mercado, impulsando la innovación, asegurando la seguridad y manteniendo bajos los costos (lo cual es una buena noticia para los consumidores).

En este artículo, compararemos los mayores éxitos en el campo: Google Cloud Storage (GCS), Microsoft Azure Storage (Azure) y Amazon Web Services (AWS), en cuanto a costos, funciones y disponibilidad.

Tabla 35: Costo

Producto	Velocidad	Región	Costo por GB	Costo por año
Almacenamiento en la nube de Google	Estándar	Región de EE. UU.	\$0.026	~\$8
Almacenamiento de Microsoft Azure	Almacenamiento de blobs CALIENTES	Este de EE. UU. (ZRS)	\$0.023	~\$7
Servicios web de Amazon	Estándar de Amazon S3	Este de EE. UU. (Ohio)	\$0.023	~\$7

Como puede ver, Microsoft Azure Storage es un poco más barato, pero no hay una gran separación en comparación de los dos.

Los tres productos también tienen niveles COOL y COLD que le permiten archivar información en la que no es necesario acceder inmediatamente, pero puede incurrir en cargos por recuperar estos datos antes de tiempo.

Tabla 36: Accesibilidad

	Almacenamiento en la nube de Google	Almacenamiento de blobs de Azure	Amazonas S3
Regiones	24 regiones	56 regiones	24 regiones
API de acceso	Compatible con S3	Compatible con S3 (usando un proxy S3)	S3
SLA de disponibilidad	99,9%	99,95%	99,99%

Se puede analizar que Microsoft Azure tiene más centros de datos ubicados en todo el mundo, lo que ofrece más opciones, cumplimiento y resiliencia.

Tabla 37: Servicios de Google Cloud, Azure y AWS

Almacenamiento en la nube de Google	Almacenamiento de blobs de Azure	Amazonas S3
Administre los costos y el rendimiento del almacenamiento con OLM	Escalable, duradero y disponible	Los datos se almacenan como objetos dentro de recursos llamados "cubos"
Tipos de ubicación para diferentes necesidades de rendimiento y redundancia.	Autenticación con Azure Active Directory y control de acceso basado en roles (RBAC)	Las características de S3 incluyen capacidades para agregar etiquetas de metadatos a objetos a través de las clases.
Clases de almacenamiento para cualquier carga de trabajo.	Optimizado para lagos de datos. Compatibilidad con espacio de nombres de archivos y acceso multiprotocolo	Se puede acceder a los objetos a través de los puntos de acceso de S3 o directamente a través del nombre de host del depósito.

En esta tabla comparativa, se enumeran los productos de Google Cloud con disponibilidad general y se los asigna a ofertas similares en Amazon Web Services (AWS) y Microsoft Azure

Tabla 38: Comparación entre los servicios que ofrece Servicios de Google Cloud Storage, Data lake Azure y AWS

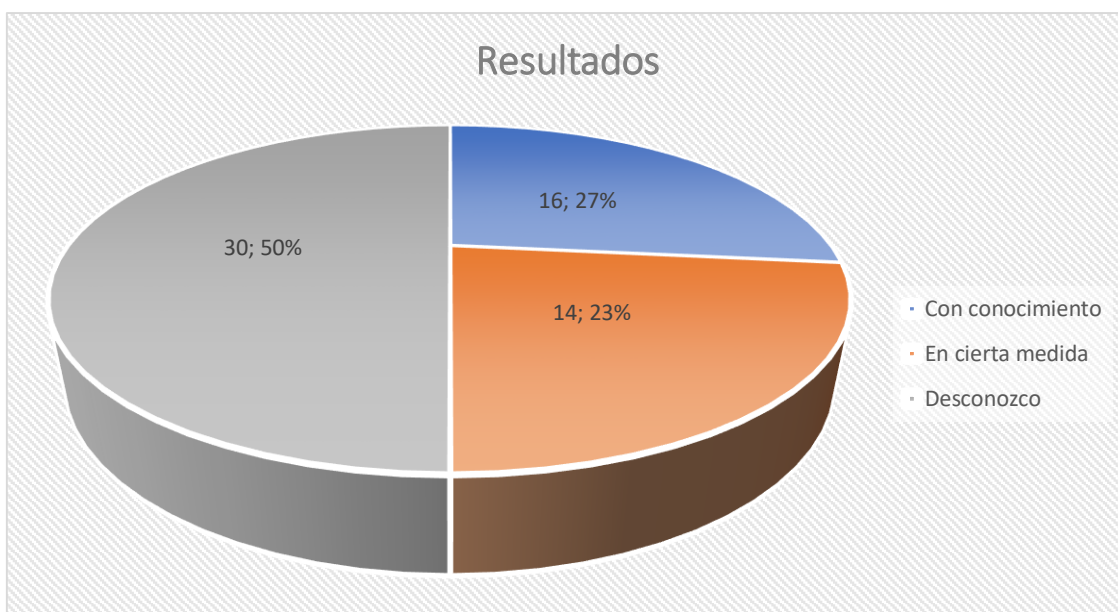
Beneficios cuentas gratuitas	Almacenamiento en la nube de Google	Almacenamiento de blobs de Azure	Amazonas S3
Cuenta	3 meses	12 meses	12 meses
Créditos	\$300 por 90 días	\$200 por 30 días \$100 por 12 meses	Pueden usar gratis durante un período de tiempo específico o se pueden usar una sola vez, según el servicio seleccionado
Productos en la plataforma	20 productos	40 productos	100 productos
Productos como servicio de data lake	4 productos	7 productos	18 productos
Servicios	<ul style="list-style-type: none"> • Google Cloud Storage. • Google Cloud DataProc. • Google Cloud Composer. 	<ul style="list-style-type: none"> • Azure Data Lake Store O Storage (ADLS) Gen1 / Gen2. • Azure Data Lake Data Factory. <ul style="list-style-type: none"> • Azure Databricks. • Power Bi. 	<ul style="list-style-type: none"> • Amazon Simple Storage Services (S3). • Amazon Data Pipeline. <ul style="list-style-type: none"> • Amazon Glue.

ANÁLISIS DE RESULTADOS DE LA ENCUESTA

La encuesta realizada fue dirigida a los Directivos de las Organizaciones de la Federación Populares Peninsulares, con el fin de identificar su nivel de conocimiento y manejo con respecto a la tecnología y su impacto sobre la adquisición.

PREGUNTA 1: ¿En qué medida usted conoce el Data Lake?

Figura 44: Conocimiento Del Data Lake

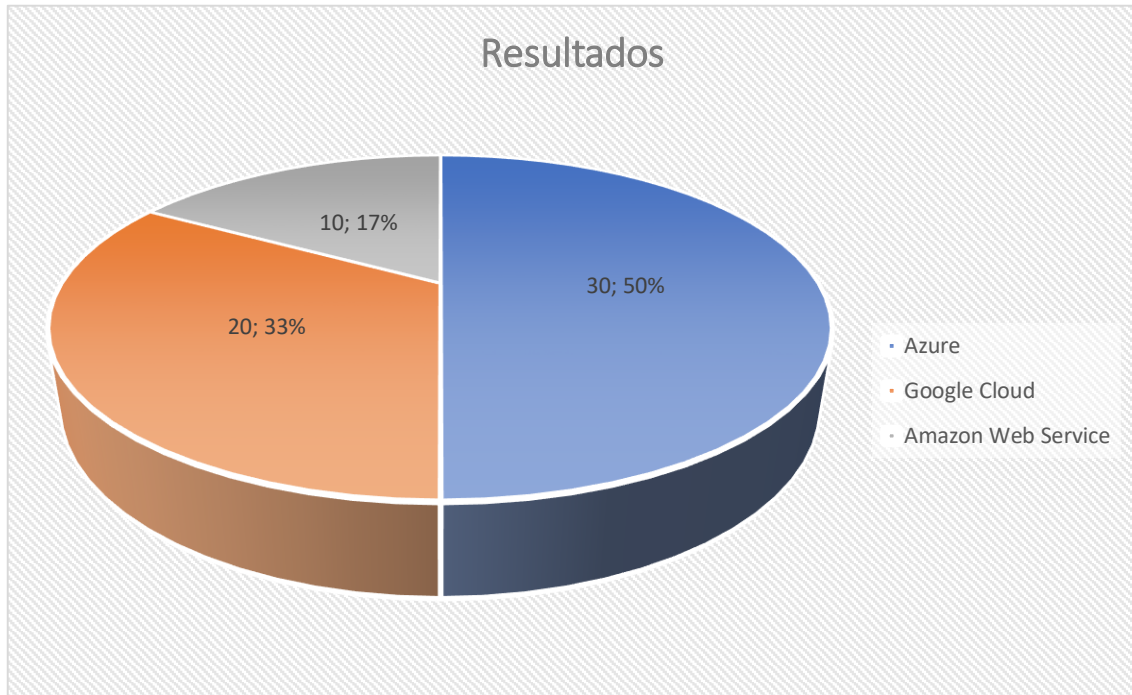


Interpretación: El 50 % de las personas encuestadas no tienen conocimiento de Lago de Datos, mientras que el 27% tiene conocimientos básicos y el 23% lo tiene en cierta medida.

Conclusión: Los resultados obtenidos nos demuestran que la mitad de los encuestados desconocen la tecnología de lago de datos, mientras que la unión de los dos porcentajes restantes 27% y 23% respectivamente conocen o tienen conocimientos básicos de esta tecnología, en el que podemos concluir en base a estos resultados que la mitad de los encuestados 50% no tienen conocimiento en lo absoluto y el 50 % restante lo tienen en cierta medida llegando a una mitad considerable.

PREGUNTA 2: ¿Qué servicios que ofrecen lago de datos de la nube conoce y piensa que sería más recomendable de usar?

Figura 45: Conocimiento de los servicios del Data Lake

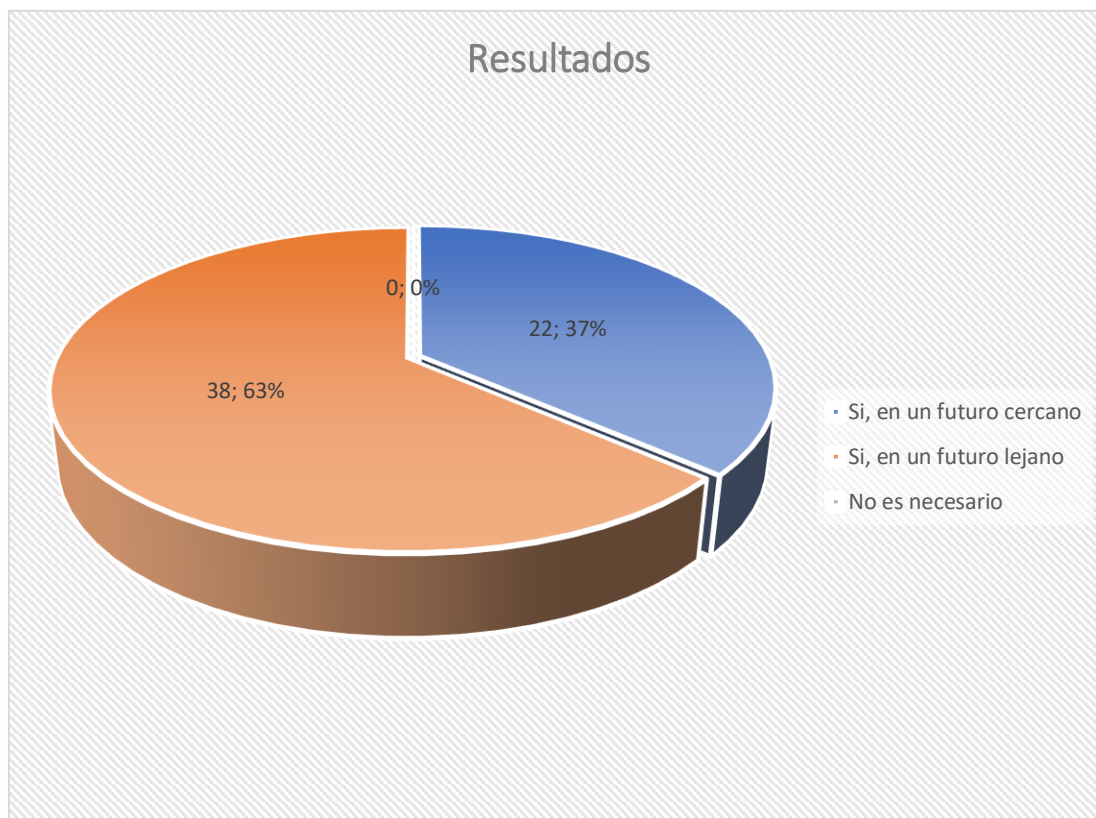


Interpretación: La mitad de los directivos a pesar de no conocer el concepto de Data Lake tienen nociones de las herramientas de las nubes entre las más conocidas esta con 50% Azure, mientras que Amazon conocen solo el 17 % y el restante conocen de Google Cloud con un 33%.

Conclusión: Los resultados obtenidos nos demuestran que la mitad de los encuestados conocen de Azure como servicio en la nube, como segundo servicio conocido esta Google Cloud y el menos conocido es Amazon Web Services, siendo los dos primeros los más conocidos y usados en la vida cotidiana, pero con otros servicios que ofrece estos mismos.

PREGUNTA 3: ¿Piensa usted que sería favorable para la organización la migración de su información a la nube y el análisis en un Data Lake?

Figura 46: Despliegue de un Data Lake en la Organización

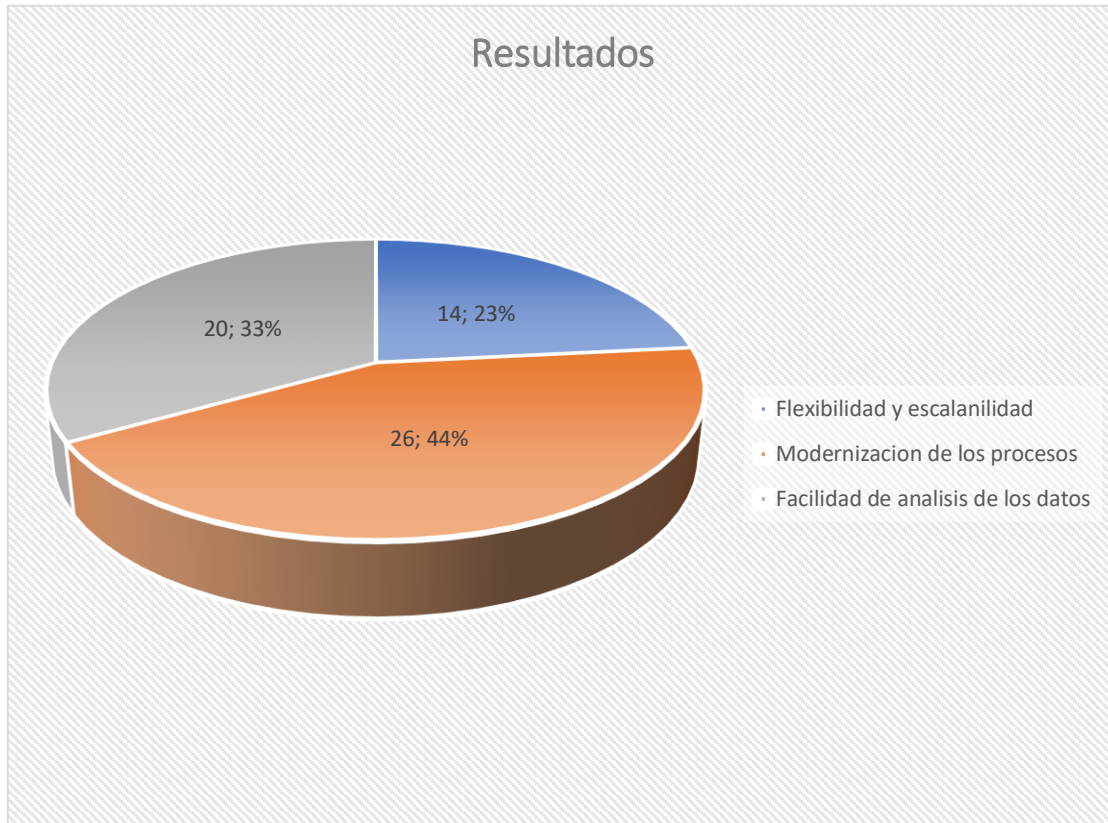


Interpretación: En su totalidad los encuestado piensan que si en necesario el despliegue de las herramientas en la nube para un Data Lake dentro de la organización, sea este a corto o largo plazo.

Conclusión: Los resultados obtenidos nos demuestran que en gran medida sean estos el 37% en un futuro cercano y el 63%, a un futuro lejano están de acuerdo con la migración y adopción de almacenamiento en la nube.

PREGUNTA 4: ¿Cuáles considera que sean las principales motivaciones para la adopción de un Data Lake?

Figura 47: Motivación a la adopción de un Data Lake

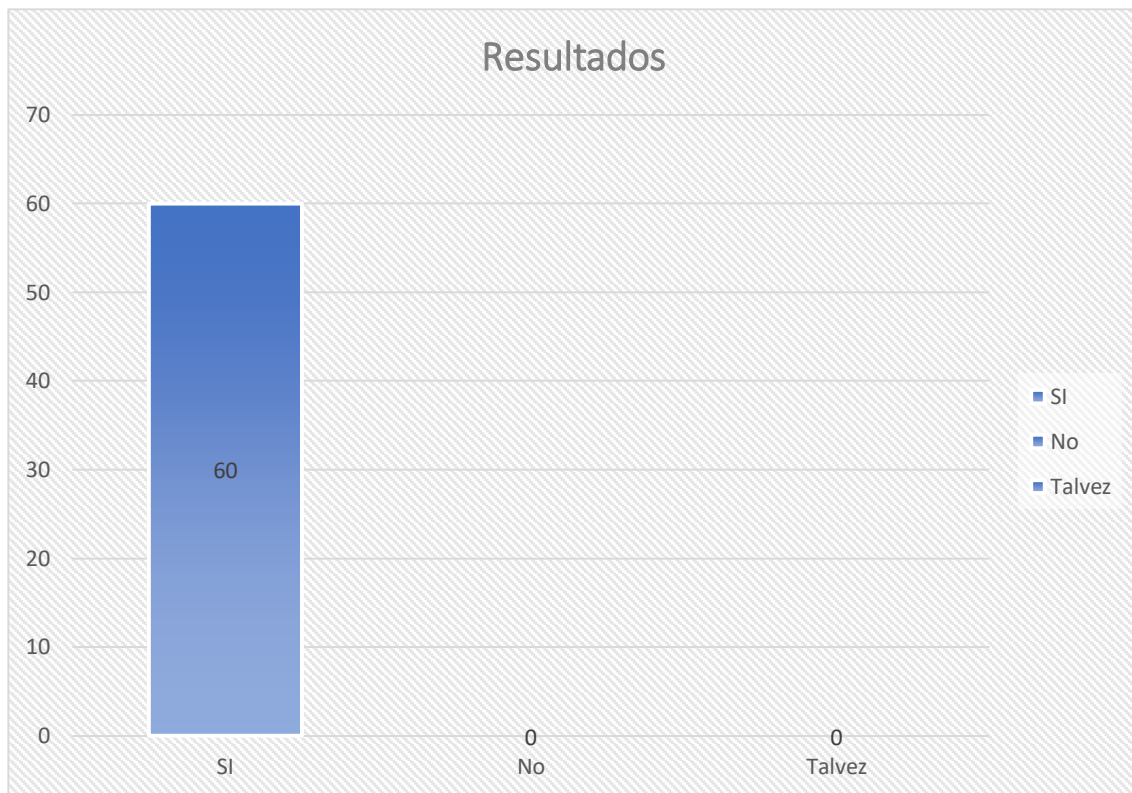


Interpretación: El 44 % de las personas encuestadas opinan que la principal motivación para la adopción de un Data Lake es la modernización de los procesos, mientras que el 33% es la facilidad de análisis de datos y el 23% piensan que es la flexibilidad y escalabilidad.

Conclusión: Los resultados obtenidos nos demuestran que hay diferentes opiniones en cuanto a la adopción de lago de datos en la federación en la que más destacan es la flexibilidad y escalabilidad, modernización de los procesos y la facilidad de análisis de datos.

PREGUNTA 5: ¿Considera que, al usar estas nuevas tecnologías como lago de datos, se puede ahorrar tiempo y dinero utilizando estos servicios?

Figura 48: Uso de nuevas tecnologías para ahorro de tiempo y dinero

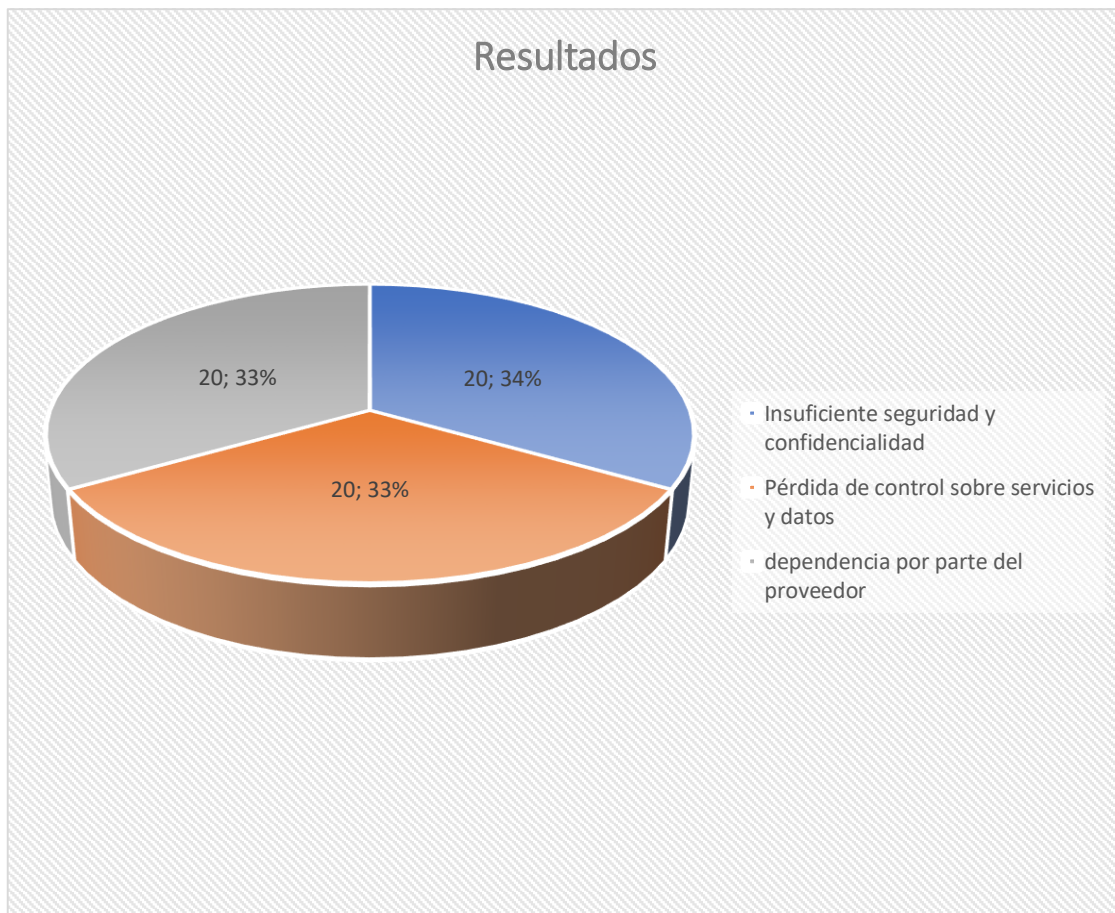


Interpretación: Los directivos en su totalidad 100% piensan que al usar estas tecnologías se puede ahorrar tiempo y dinero en los procesos.

Conclusión: Los resultados obtenidos nos demuestra que todos los encuestados piensan positivamente en la adopción de estas tecnologías benefician de alguna manera a las organizaciones.

PREGUNTA 6: ¿Cuál es la preocupación que usted tendría sobre la implementación servicios en la nube de un Data Lake?

Figura 49: Motivación a la adopción de un Data Lake

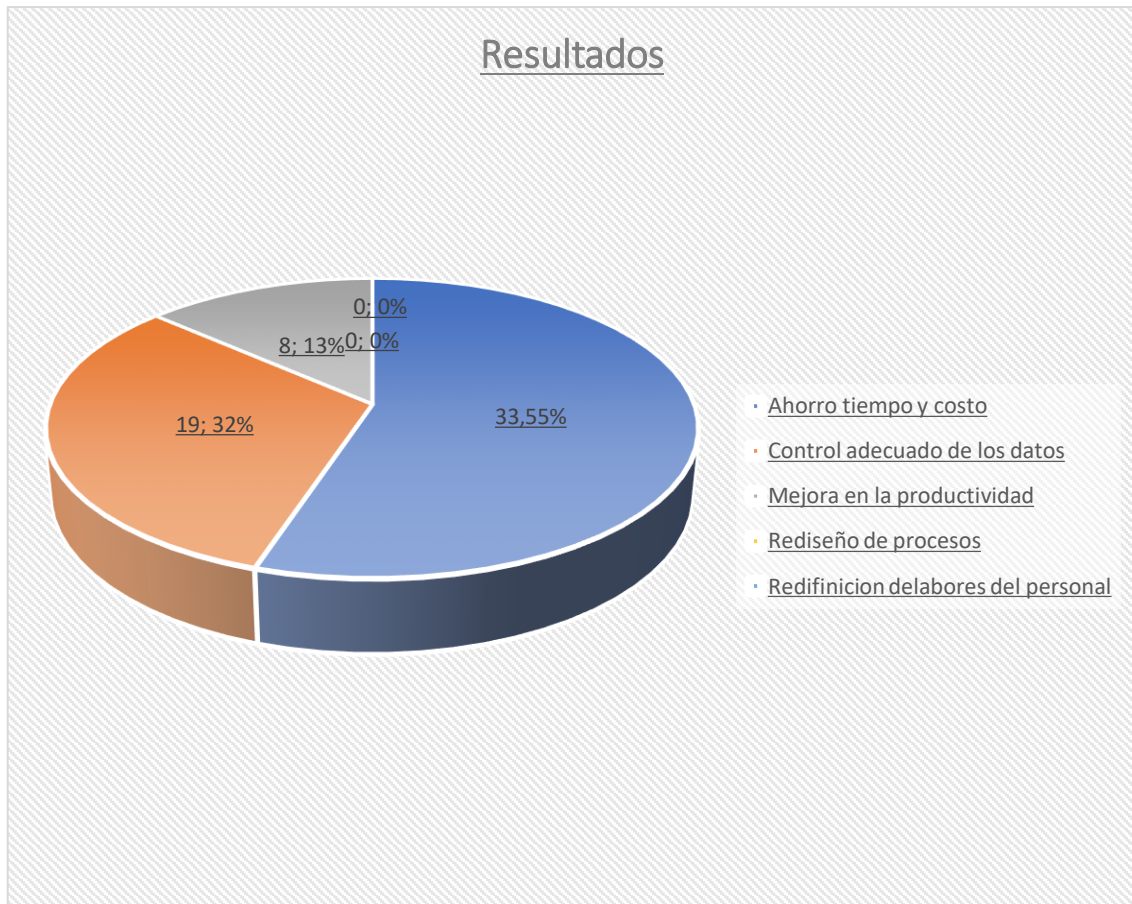


Interpretación: Los encuestados comparten el 33,3% de las opiniones de las preocupaciones al implementar este servicio en la organización.

Conclusión: Los resultados obtenidos nos demuestran que las opiniones más destacadas están, el 34% insuficiente seguridad y confidencialidad, con un 33% pérdida de control sobre servicios y datos y el 33% dependencia por parte del proveedor.

PREGUNTA 7: ¿Cuáles cree que sean los principales beneficios identificados al usar un Data Lake?

Figura 50: Beneficios del Data Lake

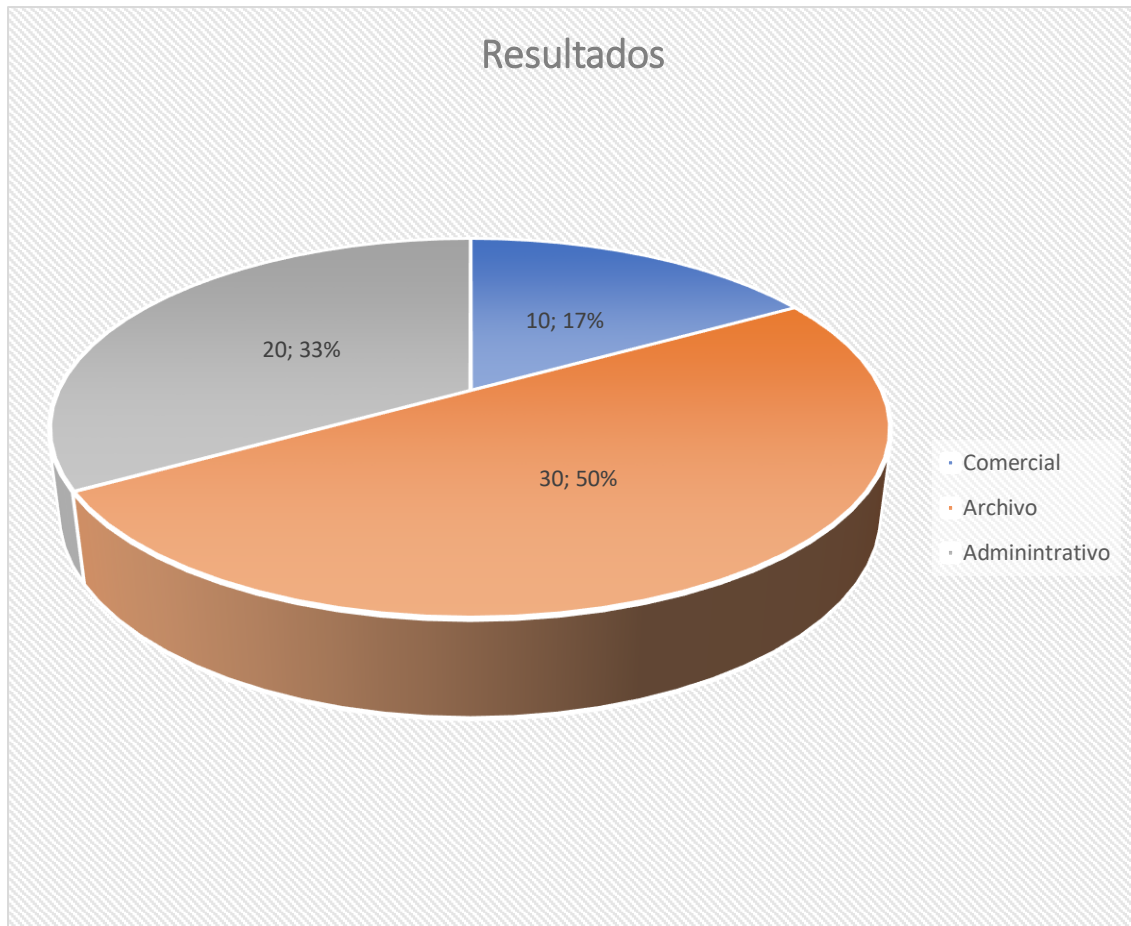


Interpretación: El 55% opinan que uno de los beneficios más importantes es el ahorro de tiempo y espacio, mientras que el 13% piensan que es la mejora de la productividad, y finalmente el 32% el del control adecuado de los datos.

Conclusión: Los resultados obtenidos nos demuestran que todos los encuestados piensan positivamente en que existen beneficios de las tecnologías de lago de datos, y que estas al adoptarlos dentro de la federación tendrán en gran medida varias ventajas en sus procesos de almacenamiento y análisis de datos.

PREGUNTA 8: ¿En qué áreas de la organización usted considera que se utilizan más las soluciones de análisis de datos?

Figura 51: Área de la organización a implementar estas herramientas.

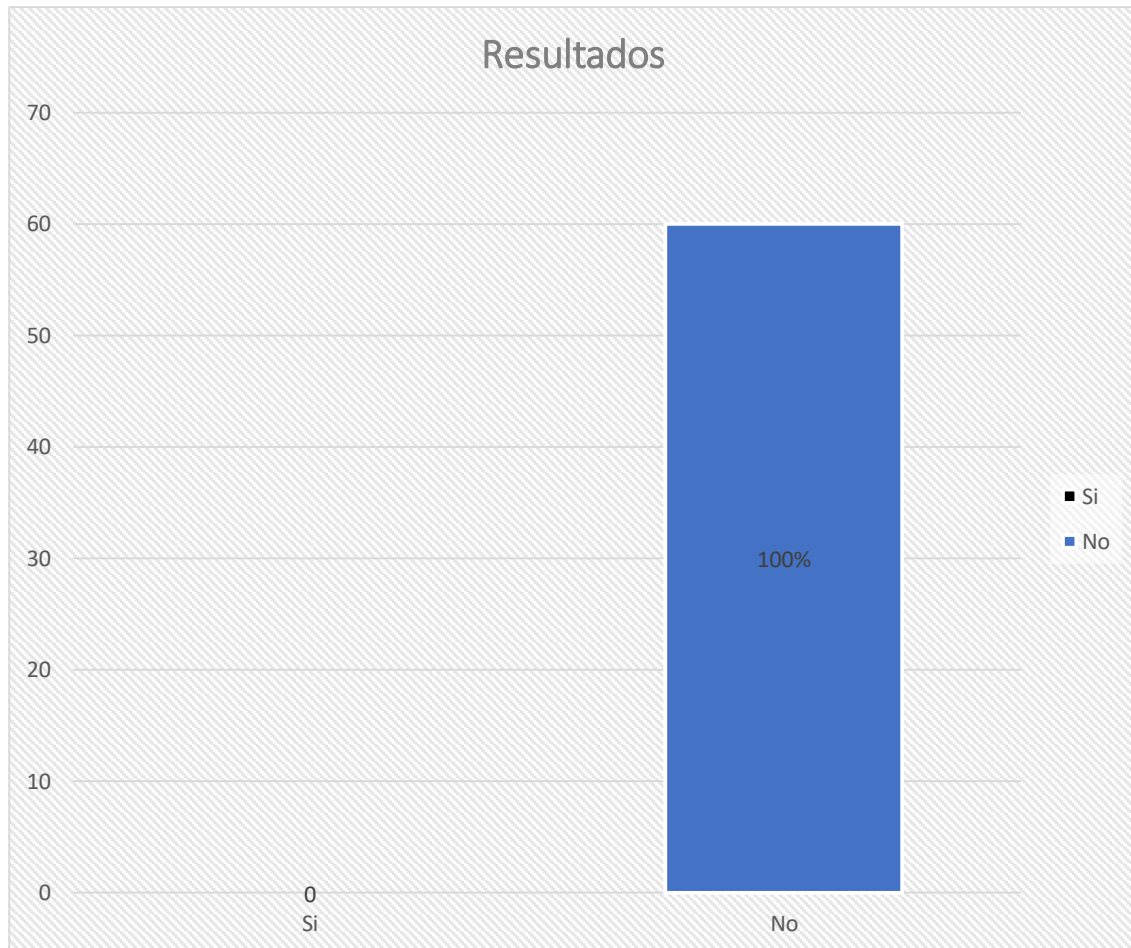


Interpretación: El 50% opinan que es útil y necesario tomar en cuenta al área de archivos para la adopción de estas herramientas, mientras que el 33% opinan que en el área administrativa sería lo más coherente y el 17% en el área comercial.

Conclusión: Los resultados obtenidos nos demuestran que las tres áreas más comunes para la implementación de un lago de datos son comercial, administrativa y archivística, siendo esta última la pronunciada por la mayoría por motivos de que esta área es la que más maneja datos y se considera mejor la adopción de estas tecnologías en esta área.

PREGUNTA 9: ¿Se realizan dentro de la organización capacitaciones sobre nuevas tecnologías?

Figura 52: Capacitación a las organizaciones.

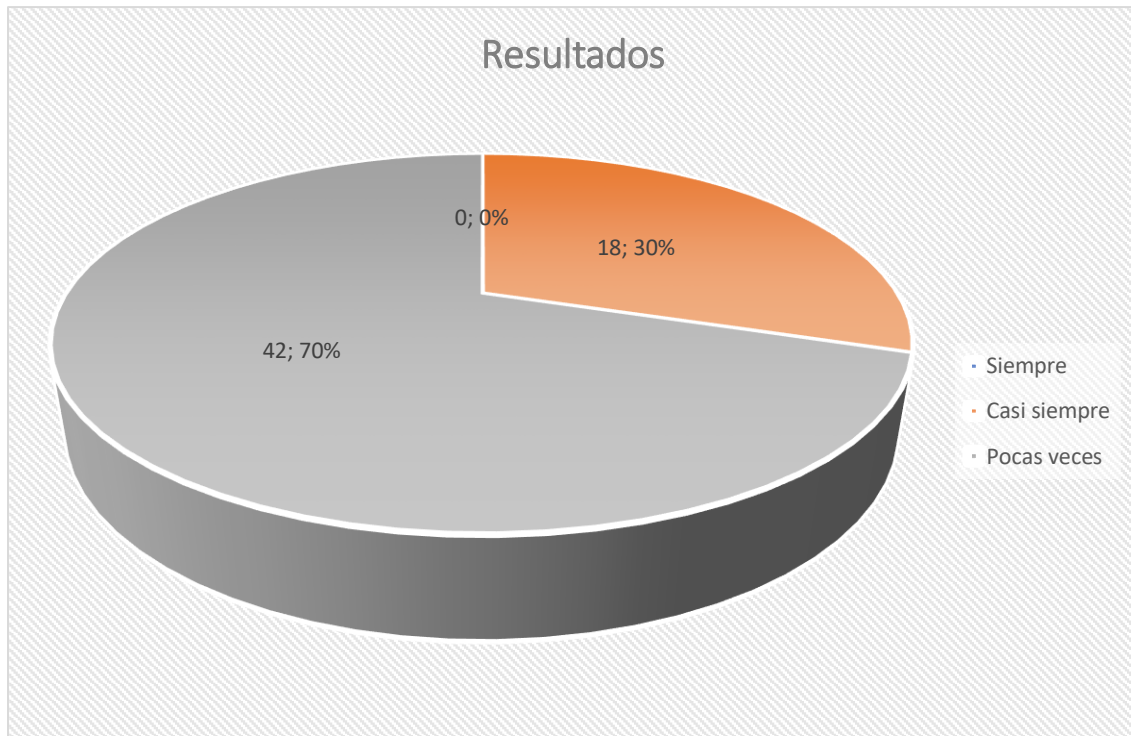


Interpretación: El 100% de los encuestados no poseen capacitaciones necesarias para el uso de estas nuevas tecnologías dentro de las organizaciones.

Conclusión: Los resultados obtenidos nos demuestran que en gran medida no cuentan con capacitaciones relevantes sobre nuevas tecnologías en la organización.

PREGUNTA 10: ¿Se realizan dentro de la organización capacitaciones sobre nuevas tecnologías?

Figura 53: Frecuencia de la Capacitación a las Organizaciones.



Interpretación: Un 70% afirman que pocas veces realizan capacitaciones sobre tecnologías en la organización, mientras que un 30% difiere de la misma afirmando que casi siempre se realizan.

Conclusión: Los resultados obtenidos nos demuestran que en gran medida que pocas veces cuentan con capacitaciones relevantes sobre nuevas tecnologías en la organización, por lo que difiere con un 30% en respuestas afirmativas. Aquellos que afirman que se efectúan casi siempre dichas capacitaciones lo realizan manera virtual como aprendizaje autónomo. Si existen capacitaciones son pocas las veces que se las realizan por parte de la institución. Por tal motivo no todo el personal cuenta con algún conocimiento amplio sobre tecnologías actuales.

CONCLUSIONES.

- Azure Storage es ideal para los casos en los que trabaja con diferentes formatos de datos y busca soluciones de almacenamiento rentables basadas en la frecuencia de acceso. Otra área de fortaleza es el cifrado de datos en reposo, y el cifrado puede administrarse por su cuenta. Sin embargo, puede que no sea adecuado para transferir grandes cantidades de datos con mucha rapidez.
- Las herramientas que se utilizaron para efectuar la guía están basadas con los criterios y conocimientos previos dentro de la fundación en tanto a la utilización de proveedores de servicios en la nube están Azure, Google Y Amazon, cumpliendo las fases establecidas en función de la Metodología NIST.
- Azure Blob Storage es ideal para los casos en los que trabaja con diferentes formatos de datos y busca soluciones de almacenamiento rentables basadas en la frecuencia de acceso realiza una mayor configuración y seguridad para la aplicación de sus servicios. Google Cloud Storage es excelente para almacenar y reproducir archivos de video de gran tamaño e incluso compartirlos de forma segura su plataforma es muy sencilla e intuitiva de realizar algún trabajo en sus servicios, aunque no poseen mayor seguridad para estos. S3, ofrece escalabilidad y disponibilidad y seguridad de los datos, aunque sus servicios se aplican por costos de uso.
- El data lake habilita espacios de exploración y experimentación sobre los datos, que, a través de diferentes técnicas y componentes, buscan encontrar su valor especial, para llegar a aportar información única para la toma de decisiones y la adaptación continua de la federación al entorno.
- Para lograr una mejor efectividad en cuanto a la práctica en la guía, se realiza de acuerdo con los criterios de la metodología NIST y a las experiencias que el personal responsable del departamento de TICS tiene como base en la utilización de estas herramientas en la nube con los proveedores más conocidos que ofrecen estas herramientas con servicio gratuito.
- La utilización de herramientas gratuitas facilita el ahorro de costos de hardware físico obteniendo una simulación de entorno real facilitan de la mejor manera posible el aprendizaje de nuevas tecnologías en la nube más segura y eficaz en comparación a las físicas.

RECOMENDACIONES.

- El personal que realice las funciones de monitorización, análisis, deberán ser el personal de TICS de la Federación de Organizaciones Populares Peninsulares (FOPOPE).
- El Servicio de data Lake no solo se usa como repositorio de archivos si no también puede ser base para la utilización de diferentes servicios aplicados al análisis de datos o big data como por ejemplo en Machine Learning o inteligencia artificial para minar los datos, los mismos son de pagos debido a la utilización de más componentes dentro de las herramientas ocupan costos por la utilización de estos servicios.
- Se recomienda la capacitación para el uso correcto herramientas ya que una configuración o mal uso de esta podría presentar problemas en el transcurso de su aplicación y presentar errores difíciles de comprender y resolver.
- En base a este proyecto, se recomienda el uso Microsoft Azure para la utilización en el procesamiento de datos, para su respectivo análisis para una posterior toma de decisiones en bien de la empresa.

BIBLIOGRAFÍA

- [1] James_Gabriel, «BIG DATA,» 22 Abril 2021. [En línea]. Available: <https://forum.huawei.com/enterprise/es/big-data/thread/731103-100261>. [Último acceso: Noviembre 2021].
- [2] A. M. C, «BIG DATA APLICADO EN EL SECTOR BANCARIO TRADICIONAL PARA,» Universidad de Palermo, Palermo, 2018.
- [3] J. R. FABIÁN GUERRERO L, «Big Data,» [En línea]. Available: <https://repository.ucatolica.edu.co/bitstream/10983/1320/1/DISE%C3%91O%20Y%20DESARROLLO%20DE%20UNA%20GU%C3%8DA%20PARA%20LA%20IMPLEMENTACI%C3%93N%20DE%20UN%20AMBIENTE%20BIG%20DATA%20EN%20LA%20UNIVERSIDAD%20CAT%C3%93LICA%20DE%20COLOMBIA.pdf>. [Último acceso: 2021].
- [4] E. G. Jiménez, «DEFINICIÓN DE UNA ARQUITECTURA DE REFERENCIA PARA,» Universidad de Medellin, Medellin, 2021.
- [5] G. d. C. Alberto, «Análisis real Time del mercado de valores con tecnologías Big Data,» Valladolid, 2019.
- [6] E. G. Garcia, «Enriquecimiento automático de un Data Lake con metadatos,» Universitat Politècnica de Catalunya. Departament d'Enginyeria de Serveis i Sistemes d'Informació, Catalunya, 2016.
- [7] M. Ojeda Orozco, «IMPLEMENTACIÓN DE DATA LAKE EN LA NUBE PARA ANALISIS DE MOVILIDAD EN LA CIUDAD DE MEXICO Y LA ZMG.,» Tlaquepaque, Jalisco, 2020.
- [8] «DISEÑO DE UN DATA LAKE PARA EMPRESARIOS DEL SECTOR DE SERVICIOS ECOTURISTICOS,» UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS, Bogota, 2020.
- [9] p. data, «Big Data: ¿En qué consiste? Su importancia, desafíos y gobernabilidad,» [En línea]. Available: powerdata.es/big-data. [Último acceso: Noviembre 2021].
- [10] FACSISTEL, «UPSE,» [En línea]. Available: <http://facsistel.upse.edu.ec/>. [Último acceso: 2021].
- [11] E. P. Nacional, «Líneas de Investigación,» [En línea]. Available: <https://fca.epn.edu.ec/index.php/fca/departamentos/depca/lneas-de-investigacion-depca>. [Último acceso: 2022].
- [12] M. S. a. D. M. B. Małysiak-Mrozek, Soft and Declarative Fishing of Information in Big Data Lake,, in IEEE Transactions on Fuzzy Systems vol. 26, no. 5, pp. 2732-2747, 2018.
- [13] C. Madera, «The next information architecture evolution: the data lake wave,» 2016.
- [14] G. BI, «Big Data, dispone de un Lago de Datos apropiado.,» [En línea]. Available: <https://gantabi.com/2022/02/14/tu-empresa-no-puede-aprovechar-el-big-data-si-no-dispone-de-un-lago-de-datos/>. [Último acceso: 2022].
- [15] «Plan De Creacion De Oportunidades,» 2017. [En línea]. Available: https://observatorioplanificacion.cepal.org/sites/default/files/plan/files/Plan-de-Creaci%C3%B3n-de-Oportunidades-2021-2025-Aprobado_compressed.pdf. [Último acceso: MAYO 2022].
- [16] M. Benassini, Introducción a la investigación de Mercados, México: PEARSON

EDUCACIÓN, 2001.

- [17] R. Llopis Goig, *El Grupo de Discusión Manual de Aplicación a la investigación social, comercial y comunicativa.*, Madrid: ESIC EDITORIAL, 2011.
- [18] N. I. o. S. a. Technology, «NIST: National Institute of Standards and Technology,» 2007. [En línea]. Available: <https://https://www.nist.gov/>. [Último acceso: Noviembre 2021].
- [19] amazon.com, «Cloud Services - Amazon Web Services (AWS).,» [En línea]. Available: <https://aws.amazon.com/es/>. [Último acceso: Noviembre 2021].
- [20] cloud.google.com, «Google Cloud - Cloud Computing.,» [En línea]. Available: https://cloud.google.com/compute?utm_source=google&utm_medium=cpc&utm_campaign=latam-LATAM-all-es-dr-BKWS-all-all-trial-e-dr-1009897-LUAC0009278&utm_content=text-ad-none-any-DEV_c-CRE_545476761277-ADGP_Hybrid%20%7C%20BKWS%20-%20EXA%20%7C%20Txt%20~%20Compu. [Último acceso: 2021Noviembre].
- [21] M. Azure, «Data Lake Analytics.,» [En línea]. Available: https://azure.microsoft.com/es-mx/services/data-lake-analytics/?&ef_id=Cj0KCQjwrJOMBhCZARIsAGEd4VEyA4ITYrsxy6iRSsxaeQo1Le9cQqB4NyyZlt7Ar-Nl4rxKmZXCw4IaAtUHEALw_wcB:G:s&OCID=AID2201052_SEM_Cj0KCQjwrJOMBhCZARIsAGEd4VEyA4ITYrsxy6iRSsxaeQo1Le9cQqB4NyyZlt7Ar-N. [Último acceso: Noviembre 2021].
- [22] M. Azure, «Data Lake Analytics,» 2021. [En línea]. Available: https://azure.microsoft.com/es-mx/services/data-lake-analytics/?&ef_id=CjwKCAjw7vuUBhBUEiwaEdu2pMstOk2h4ZoJ69D_9uJNyrUTLY0Tl13XKuCb5gxjFhNjVQf8FWZ8RoCLNcQAvD_BwE:G:s&OCID=AID2201052_SEM_CjwKCAjw7vuUBhBUEiwaEdu2pMstOk2h4ZoJ69D_9uJNyrUTLY0Tl13XKuCb5gxjFhN.
- [23] P. M. Mell y T. Grance, «NIST: The NIST Definition of Cloud Computing.,» 28 Septiembre 2011. [En línea]. Available: <https://www.nist.gov/publications/nist-definition-cloud-computing>. [Último acceso: 19 Julio 2021].
- [24] M. Build, «Introduction to Azure Data Lake Storage Gen2,» [En línea]. Available: <https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-introduction>.
- [25] A. e. AWS, «Almacenamiento de lago de datos en AWS,» [En línea]. Available: <https://aws.amazon.com/es/products/storage/data-lake-storage/>. [Último acceso: 2022].
- [26] G. Cloud, «Temas,» [En línea]. Available: <https://cloud.google.com/learn/what-is-a-data-lake?hl=es-419>. [Último acceso: 2022].
- [27] M. Azure, «¿Qué es Lago de datos?,» 2022. [En línea]. Available: <https://azure.microsoft.com/es-es/overview/what-is-a-data-lake/>.
- [28] J. M. Maqueira Marín y S. Bruque Cámara, «Agentes impulsores de la adopción de Cloud Computing en las empresas. ¿Quién mueve la nube,» [En línea]. Available: <https://journals.ucjc.edu/ubr/article/view/853>.
- [29] A. Cravero, I. Lefiguala, R. Tralma y S. Gonzalez, «Data Lake para la Dirección de Análisis Universitaria: Arquitectura y Metadata,» *Revista Ibérica de Sistemas e*

Tecnologías de Informação; Lousada, Temuco Chile, 2020.

- [30] J. C. Agudelo Patiño, «Data lakes: aplicaciones, herramientas y arquitecturas,» [En línea]. Available: <https://hdl.handle.net/11059/12186>. [Último acceso: 2020].
- [31] J. A. Jara Collahuazo, «Análisis de Factibilidad en la implementación de,» Mayo 2012. [En línea]. Available: <http://bibdigital.epn.edu.ec/handle/15000/4649> . [Último acceso: 01 2022].
- [32] W. Grijalva, «“Estudio de Factibilidad para la Migración a Cloud de la Plataforma Tecnológica de la Empresa Asistecooper S.A,”» Universidad de las Américas, Loja, 2017.
- [33] telecomunicaciones.gob.ec, «LEY DE COMERCIO ELECTRONICO, FIRMAS Y MENSAJES DE DATOS,» 2012. [En línea]. Available: <https://www.telecomunicaciones.gob.ec/wp-content/uploads/downloads/2012/11/Ley-de-Comercio-Electronico-Firmas-y-Mensajes-de-Datos.pdf>. [Último acceso: 2022].
- [34] L. O. d. Telecomunicaciones, «Artículo 24. Obligaciones de los prestadores de servicios de telecomunicaciones,» Asamblea Nacional, Quito, 2015.
- [35] A. I. GmbH, «Google Cloud Platform: What it is, how to use it, and how it compares,» Acronis International GmbH, [En línea]. Available: <https://www.acronis.com/es-mx/articles/google-cloud-platform/>. [Último acceso: Enero 2022].
- [36] A. Microsoft, «Azure Data Lake Store,» 2022. [En línea]. Available: <https://azure.microsoft.com/es-es/services/storage/data-lake-storage/#features>.

ANEXOS

Anexos 1: Entrevista al presidente de la Federación de Organizaciones Populares Peninsulares (FOPOPE).



UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA FACULTAD DE SISTEMAS Y TELECOMUNICACIONES ESCUELA DE INFORMÁTICA

FORMATO DE ENTREVISTA

“PROPUESTA DE UN DATA LAKE EN LA NUBE PARA LA OPTIMIZACIÓN DE DATOS PARA LA FEDERACIÓN DE ORGANIZACIONES POPULARES PENINSULARES (FOPOPE)”

Objetivo: Obtener información para establecer la situación actual del uso de tecnologías y percepción sobre el data lake en una nube en la Federación de Organizaciones Populares Peninsulares de la Provincia de Santa Elena.

Persona Entrevistada: Sr. Sergio Gonzalez Figuera - Presidente

Preguntas:

1. ¿Qué opina sobre el impacto de las tecnologías de la información en estos tiempos?
La tecnología cada día va avanzando, es de mucha importancia en la vida personal, laboral y profesional ya que nos ayuda a organizarnos mejor y aprender cosas nuevas.
2. ¿Cree que el avance de las tecnologías es positivo o contradictorio para las organizaciones?
La tecnología es positivo para las organizaciones ya que ayuda al desarrollo social y empresarial mas que todo a las organizaciones ayuda a su innovación con respecto a su sistema.
3. ¿Cuál es la importancia que usted considera sobre invertir en nuevas tecnologías para la institución?
Sería muy bueno aplicar las tecnologías ya que esto nos ayuda a simplificar el trabajo que anteriormente era muy demorado y mientras más avance la tecnología más se debe innovar y actualizar y por ende invertir en la misma.
4. ¿Considera usted que es necesario la adaptación de almacenamiento y procesamiento en la nube en la organización?
Si, tenemos personal capacitado ya que tienen experiencia en el manejo de almacenamiento en la nube.

5. ¿Qué tipo de gestiones se manejan dentro de la institución y como se lleva a cabo cada una?

En los últimos años se ha manejado base de datos en lo que comprende estatutos reglamentales, base de datos de sistema de gestión documental y sistema de archivo el cual se ha almacenado por 5 años hasta la fecha actual.

6. ¿Tiene alguna idea o conoce sobre la terminología de Computación en la nube?

Como presidente no, pero tenemos un departamento de TICS que tiene conocimiento y manejan en la actualidad almacenamiento en la nube.

7. ¿Tiene alguna idea o conoce sobre la terminología de Datos e información?

No, pero recalco el personal del departamento de TICS se encuentra preparado para este tipo de actualizaciones.

8. ¿Tiene alguna idea o conoce sobre la terminología de Lagos de Datos?

No, pero con capacitaciones lograríamos captar el tema.

9. ¿Tiene alguna idea o conoce sobre la terminología de Almacenamiento de Datos?

Si, por ejemplo resguardar la información con seguridad a posibles robos de datos.

10. ¿Cómo almacena los datos guardados de importancia de las organizaciones?

Nuestros datos en la organización están respaldados en la nube y en servidores que lo manejan los encargados de TICS.

11. ¿Estaría dispuesto a efectuar un cambio hacia la innovación dentro de la institución?

Si, porque toda empresa que no innova se queda en la vanguardia y no progresa y por lo consiguiente puede dejar de existir.

12. En cuanto a la gestión de cambios, ¿Qué tan preparados se encuentra la institución para sobrellevar una nueva modalidad de servicios de almacenamiento y análisis de datos?

Tenemos una línea base en la cual podemos realizar mejoras con respecto a la toma de decisiones a través de programas que analicen estos datos almacenados.

Anexos 2: Formato de encuesta - Federación de Organizaciones Populares peninsulares (FOPOPE).

UNIVERSIDAD ESTATAL PENÍNSULA DE SANTA ELENA FACULTAD DE SISTEMAS Y TELECOMUNICACIONES ESCUELA DE INFORMÁTICA



FORMATO DE ENCUESTA

“PROPUESTA DE UN DATA LAKE EN LA NUBE PARA LA OPTIMIZACIÓN DE DATOS PARA LA FEDERACIÓN DE ORGANIZACIONES POPULARES PENINSULARES (FOPOPE)”

Objetivo: Obtener información para establecer la situación actual del uso de tecnologías y percepción sobre el data lake en una nube en la Federación de Organizaciones Popular Peninsulares de la Provincia de Santa Elena.

DATOS INFORMATIVOS:

Nombre del encuestado: _____

Puesto que ocupa en la Institución: _____

PREGUNTAS:

1. ¿En qué medida usted conoce el Data Lake?

Con conocimiento

En cierta medida

Desconozco

Si la respuesta 1 es favorable seguir con los dos casos contrarios saltar a la 3

2. ¿Qué servicios que ofrecen lago de datos de la nube conoce y piensa que sería más recomendable de usar?

Azure

Google Cloud

Amazon Web Service

3. ¿Piensa usted que sería favorable para la organización la migración de su información a la nube y el análisis en un Data Lake?

Si, en un futuro cercano

Si, en un futuro lejano

No es necesario.

4. ¿Cuáles considera que sean las principales motivaciones para la adopción de un Data Lake?

- Flexibilidad y escalabilidad
- Modernización de los procesos
- Facilidad de análisis de los datos.

5. ¿Considera que, al usar estas nuevas tecnologías como lago de datos, se puede ahorrar tiempo y dinero utilizando estos servicios?

- Si
- No
- Tal vez

6. ¿Cuál es la preocupación que usted tendría sobre la implementación servicios en la nube de un Data Lake?

- Insuficiente seguridad y confidencialidad
- Pérdida de control sobre los servicios
- Dependencia por parte del proveedor

7. ¿Cuáles cree que sean los principales beneficios identificados al usar un Data Lake?

- Ahorro en tiempos y costo
- Control adecuado de los datos
- Mejora en la productividad
- Rediseño de Procesos
- Redefinición de labores de personal

8. ¿En qué áreas de la organización usted considera que se utilizan más las soluciones de análisis de datos?

- Comercial
- Archivo
- Administrativo

9. ¿Se realizan dentro de la organización capacitaciones sobre nuevas tecnologías?

- Sí
- No

10. ¿Con que frecuencia se realizan las capacitaciones?

- Siempre
- Casi Siempre
- Nunca

Figura 54: Entrevista al presidente de la Federación FOPOPE.



Figura 55: Encuesta con Dirigentes de las organizaciones.



Figura 56: Socialización del proyecto a los directivos y de la organización que conforman la Federación de Organizaciones Populares Provinciales FOPOPE



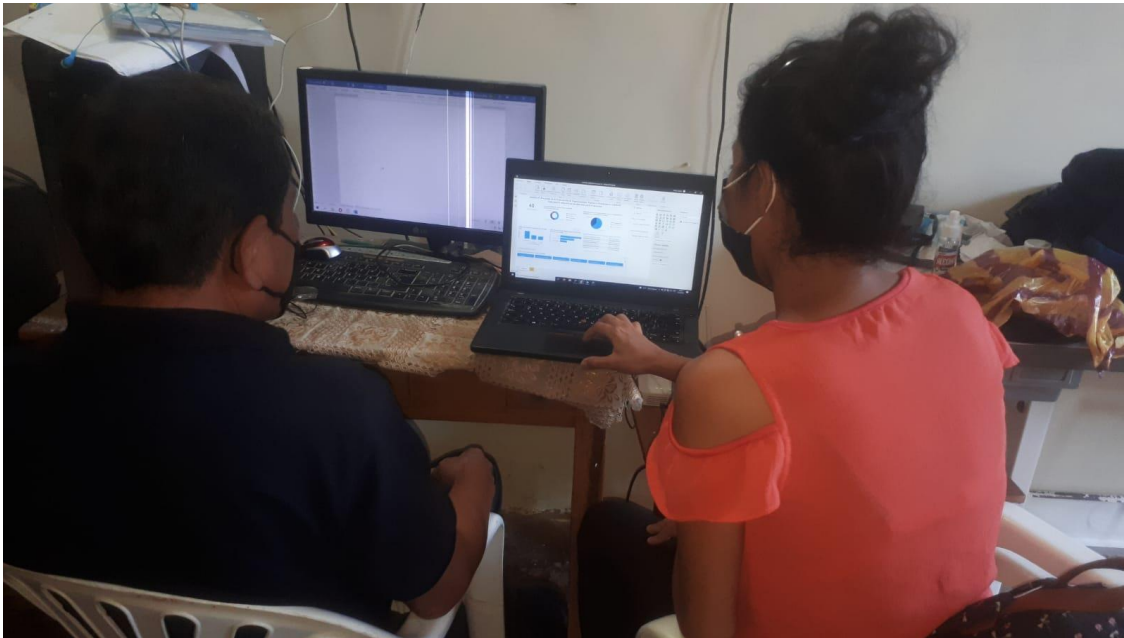
Figura 57: Encuesta a directivas de la Federación FOPOPE.



Figura 58: Presentación de la simulación en el Portal de Azure



Figura 59: Presentación de la Visualización el Power BI



La libertad, 07 de Octubre del 2022

CERTIFICADO ANTIPLAGIO

002-TUTOR MACS-2022

En calidad de tutor del trabajo de titulación denominado "**Propuesta de una guía de adopción de un data lake enmarcando los aspectos de importancia de cloud computing caso de estudio Federación de Organizaciones Populares Peninsulares (FOPOPE)**", elaborado por la estudiante Mayra Lilibeth Balón De La Cruz, egresado de la carrera de Tecnología de la Información, de la Facultad de Sistemas y Telecomunicaciones de la Universidad Estatal Península de Santa Elena, previo a la obtención del título de Ingeniero en TICs., me permito declarar que una vez analizado en el sistema antiplagio Urkund, luego de haber cumplido los requerimientos exigidos de valoración, el presente proyecto ejecutado, se encuentra con el 0% de valoración permitida, por consiguiente se procede a emitir el presente informe.

Adjunto reporte de similitud,



Document Information

Analyzed document	BALON DE LA CRUZ MAYRA LILIBETH COMPLETIVO.docx (D145635637)
Submitted	10/7/2022 3:58:00 PM
Submitted by	DANIEL NANI QUIRUMBAY YAGUAL
Submitter email	dquirumbay@upse.edu.ec
Similarity	0%
Analysis address	dquirumbay.upse@analysis.arkund.com

Atentamente,

Daniel Quirumbay Yagual.
C.I 0919659672
DOCENTE TUTOR