****

**UNIVERSIDAD PRIVADA DE TACNA**

**FACULTAD DE INGENIERIA**

**Escuela Profesional de Ingeniería de Sistemas**

**Informe Final**

**Proyecto *Implementación de Almacén de Datos en AWS para Zofratacna***

Curso: *Inteligencia de Negocios*

Docente: *Ing. Patrick Jose Cuadros Quiroga*

Integrantes:

***Rivera Mendoza Jhonny***

***Cano Sucso Anthony Alexander***

***Jarro Cachi Jose Luis***

***Valverde Zamora Jean Pier Elias***

***Chambilla Zuñiga Josue***

**Tacna – Perú**

***2024***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| CONTROL DE VERSIONES | | | | | |
| Versión | Hecha por | Revisada por | Aprobada por | Fecha | Motivo |
| 1.0 | MPV | ELV | ARV | 10/10/2020 | Versión Original |

INDICE GENERAL

[1. Antecedentes 4](#_Toc171346668)

[2. Planteamiento del Problema 4](#_Toc171346669)

[a. Problema 4](#_Toc171346670)

[b. Justificación 4](#_Toc171346671)

[c. Alcance 4](#_Toc171346672)

[3. Objetivos 4](#_Toc171346673)

[4. Marco Teórico 4](#_Toc171346674)

[5. Desarrollo de la Solución 5](#_Toc171346675)

[a. Análisis de Factibilidad (técnico, económica, operativa, social, legal, ambiental) 5](#_Toc171346676)

[b. Tecnología de Desarrollo 6](#_Toc171346677)

[c. Metodología de implementación 6](#_Toc171346678)

[6. Cronograma 6](#_Toc171346679)

[7. Presupuesto 7](#_Toc171346680)

[8. Conclusiones 7](#_Toc171346681)

[Recomendaciones 7](#_Toc171346682)

[Bibliografía 7](#_Toc171346683)

[Anexos 7](#_Toc171346684)

[Anexo 01 Informe de Factiblidad 7](#_Toc171346685)

[Anex0 02 Documento de Visión 7](#_Toc171346686)

[Anexo 03 Documento SRS 7](#_Toc171346687)

[Anexo 04 Documento SAD 7](#_Toc171346688)

# Antecedentes

La gestión y análisis de datos en Zofratacna se realiza actualmente de manera manual y descentralizada, lo que genera ineficiencias y posibles errores en la interpretación de la información.

# Planteamiento del Problema

## Problema

La falta de un sistema centralizado y automatizado para la gestión de datos impide una eficiente toma de decisiones en Zofratacna.

## Justificación

Centralizar y automatizar el almacenamiento de datos permitirá mejorar la precisión y eficiencia en la gestión financiera y operativa, así como facilitar el acceso seguro a los datos para análisis futuros.

## Alcance

El proyecto abarca desde la recopilación de datos en archivos Excel hasta su integración en la infraestructura de AWS. Incluye la importación, transformación, almacenamiento seguro y actualización automática de los datos, así como la integración con los sistemas existentes de Zofratacna.

# Objetivos

El objetivo principal es optimizar la eficiencia operativa y la toma de decisiones mediante el análisis en tiempo real de datos financieros y operativos.

# Marco Teórico

* Amazon S3

Amazon S3 (Simple Storage Service) es un servicio de almacenamiento de objetos escalable y altamente duradero de Amazon Web Services (AWS). Permite almacenar y recuperar datos desde cualquier lugar de la web y ofrece una durabilidad del 99.999999999% para objetos almacenados y una disponibilidad del 99.99%. S3 se utiliza ampliamente como base para almacenar datos en la nube debido a su escalabilidad, seguridad y facilidad de uso.

* AWS Lake Formation

AWS Lake Formation es un servicio de AWS que simplifica y automatiza la creación, administración y seguridad de data lakes en AWS. Un data lake es un repositorio centralizado que permite almacenar todos los datos estructurados y no estructurados en su formato nativo hasta que se necesiten para análisis, procesamiento o visualización. Lake Formation proporciona herramientas para la ingestión de datos desde diversas fuentes, la limpieza, la transformación y la protección de datos mediante políticas de acceso.

* Glue Job

AWS Glue es un servicio de ETL (Extract, Transform, Load) que permite preparar y cargar datos para el análisis. Los Glue Jobs son unidades de trabajo en Glue que se utilizan para ejecutar tareas de ETL y procesamiento de datos. Un Glue Job define las transformaciones necesarias para convertir datos de un formato a otro, limpiar datos, filtrar datos y cargarlos en un destino como un data warehouse o un data lake. Glue Jobs se pueden programar para ejecutarse automáticamente en función de una programación específica o en respuesta a eventos específicos, como la llegada de nuevos datos.

Para construir un sistema de data lake en AWS utilizando S3, Lake Formation y Glue Job, el proceso típico involucra varios pasos:

* Almacenamiento en Amazon S3:

Los datos de diversas fuentes se almacenan en buckets de Amazon S3.

S3 proporciona durabilidad y escalabilidad para manejar grandes volúmenes de datos de forma eficiente.

* Ingestión de Datos con Lake Formation:

Lake Formation facilita la ingestión automatizada de datos desde múltiples fuentes como bases de datos, servicios en la nube y más.

Se pueden definir flujos de trabajo para la ingestión continua de datos en el data lake.

* Preparación y Seguridad de Datos:

Lake Formation permite definir políticas de seguridad y acceso para los datos almacenados en el data lake, asegurando que solo usuarios autorizados puedan acceder a ellos.

* Transformación de Datos con Glue Job:

Se crean Glue Jobs para transformar los datos crudos almacenados en S3 en formatos más útiles para análisis.

Las transformaciones pueden incluir limpieza de datos, conversión de formatos, enriquecimiento de datos, etc.

* Carga en Destinos Finales:

Los datos transformados se cargan en destinos finales como almacenes de datos (por ejemplo, Amazon Redshift) o se mantienen en el data lake para análisis adicional.

* Automatización y Monitoreo:

Los Glue Jobs y otros procesos pueden automatizarse y monitorearse utilizando servicios como AWS CloudWatch para asegurar el procesamiento eficiente y la integridad de los datos.

# Desarrollo de la Solución

## Análisis de Factibilidad (técnico, económica, operativa, social, legal, ambiental)

Se evaluaron los aspectos técnico, económico, operativo, social, legal y ambiental para asegurar la viabilidad del proyecto.

## Tecnología de Desarrollo

Se utilizará Amazon Web Services (AWS) para crear un almacén de datos robusto y escalable.

## Metodología de implementación

Se seguirá un enfoque estructurado que incluye la creación de documentos como el Documento de Visión, SRS (Especificación de Requerimientos de Software) y SAD (Documento de Arquitectura de Software).

# Cronograma

**Mes 1: Planificación y Preparación**

Semana 1-2: Planificación Inicial

* Reunión inicial para entender los requisitos de almacenamiento de datos.
* Definición del alcance del proyecto y objetivos específicos.
* Asignación de roles y responsabilidades del equipo.

Semana 3-4: Diseño de la Infraestructura AWS

* Diseño detallado de la arquitectura de almacenamiento en AWS utilizando servicios como AWS S3.
* Configuración de cuentas y permisos necesarios en AWS.
* Pruebas preliminares de configuración y seguridad de AWS S3.

**Mes 2: Implementación y Pruebas**

Semana 5-6: Implementación de Almacenamiento

* Creación de buckets en AWS S3 para el almacenamiento de datos.
* Desarrollo de scripts o procesos para la carga inicial de datos en formato Excel a AWS S3.
* Configuración de políticas de retención y seguridad de datos en AWS S3.

Semana 7-8: Pruebas y Ajustes

* Pruebas exhaustivas de carga y descarga de datos en AWS S3.
* Evaluación de la seguridad y accesibilidad de los datos almacenados.
* Ajustes y optimización de la configuración de AWS S3 según los resultados de las pruebas.

**Mes 3: Documentación**

Semana 9-10: Documentación

* Elaboración de documentación técnica detallada sobre la configuración de AWS S3.
* Preparación de material de capacitación para administradores de sistemas y usuarios finales sobre el uso de AWS S3.
* Realización de sesiones de capacitación y familiarización con AWS S3 para el equipo de Zofratacna.

Semana 11-12: Revisión Final

* Revisión final del proyecto con el cliente para asegurar que todos los requisitos se hayan cumplido.
* Preparación de informes de entrega y presentación final al cliente.
* Evaluación de lecciones aprendidas y preparación para la siguiente fase del proyecto.

# Presupuesto

La inversión inicial del proyecto podría ser aproximadamente $32,650, considerando los costos en el informe de factibilidad y el contexto proporcionado en el informe de factibilidad.

# Conclusiones

La implementación del almacén de datos en AWS para Zofratacna mejorará significativamente la gestión de datos y la toma de decisiones estratégicas.

# Recomendaciones

* Realizar una fase piloto para validar el sistema.
* Proveer capacitación continua a los usuarios.
* Implementar un plan de mantenimiento y soporte.

# Bibliografía

* Documentación de Amazon Web Services.
* Manuales internos de Zofratacna.
* Guías de mejores prácticas en Business Intelligence.

# Anexos

## Anexo 01 Informe de Factiblidad

## Anex0 02 Documento de Visión

## Anexo 03 Documento SRS

## Anexo 04 Documento SAD