“Año del Bicentenario, de la consolidación de nuestra Independencia, y de la conmemoración de las heroicas batallas de Junín y Ayacucho”

**FACULTAD DE INGENIERÍA**

**ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

Texto, Logotipo

Descripción generada automáticamente

**“Sistema de *Automatización de Carga de Datos y Modificación de Reportes en el Almacén de Datos*”**

**Integrantes:**

| **Nº** | **Código Universitario** | **Apellidos y Nombres** |
| --- | --- | --- |
| **1.-** | **2017057528** | **Ccalli Chata, Joel Robert** |
| **2.-** | **2020067145** | **Anahua Coaquira, Mayner Gonzalo** |
| **3.-** | **2020066924** | **Zeballos Purca, Justin Zinedine** |
| **4.-** | **2020069046** | **Salinas Condori, Erick Javier** |
| **5.-** | **2018000654** | **Paco Ramos, Aaron Pedro** |

**CURSO :**  **“Inteligencia de Negocios”**

**SECCIÓN**  **:**  **“A”**

**DOCENTE**  **:**  **Ing. “Patrick Jose Cuadros Quiroga”**

**Tacna - Perú**

**2024**

| CONTROL DE VERSIONES | | | | | |
| --- | --- | --- | --- | --- | --- |
| Versión | Hecha por | Revisada por | Aprobada por | Fecha | Motivo |
| 1.0 | Joel Ccalli  Erick Salinas  Aaron Paco  Justin Zeballos | Mayner Anahua | Mayner Anahua | 27/06/2024 | Versión Original |

Sistema de *Automatización de Carga de Datos y Modificación de Reportes en el Almacén de Datos*

Informe de SAD

FD-04

Versión *1.0*

INDICE GENERAL

**Contenido**

[***1.***](#_heading=h.gjdgxs) ***INTRODUCCIÓN 5***

[**1.1.**](#_heading=h.30j0zll) **Propósito (Diagrama 4+1) 5**

[**1.2.**](#_heading=h.1fob9te) **Alcance 5**

[**1.3.**](#_heading=h.3znysh7) **Definición, siglas y abreviaturas 5**

[**1.4.**](#_heading=h.2et92p0) **Organización del documento 5**

[***2.***](#_heading=h.tyjcwt) ***OBJETIVOS Y RESTRICCIONES ARQUITECTONICAS 5***

[2.1.1.](#_heading=h.3dy6vkm) Requerimientos Funcionales 5

[2.1.2.](#_heading=h.1t3h5sf) Requerimientos No Funcionales – Atributos de Calidad 5

[***3.***](#_heading=h.4d34og8) ***REPRESENTACIÓN DE LA ARQUITECTURA DEL SISTEMA 6***

[**3.1.**](#_heading=h.2s8eyo1) **Vista de Caso de uso 6**

[3.1.1.](#_heading=h.17dp8vu) Diagramas de Casos de uso 6

[**3.2.**](#_heading=h.3rdcrjn) **Vista Lógica 6**

[3.2.1.](#_heading=h.26in1rg) Diagrama de Subsistemas (paquetes) 7

[3.2.2.](#_heading=h.lnxbz9) Diagrama de Secuencia (vista de diseño) 7

[3.2.3.](#_heading=h.35nkun2) Diagrama de Colaboración (vista de diseño) 7

[3.2.4.](#_heading=h.1ksv4uv) Diagrama de Objetos 7

[3.2.5.](#_heading=h.44sinio) Diagrama de Clases 7

[3.2.6.](#_heading=h.2jxsxqh) Diagrama de Base de datos (relacional o no relacional) 7

[**3.3.**](#_heading=h.z337ya) **Vista de Implementación (vista de desarrollo) 7**

[3.3.1.](#_heading=h.3j2qqm3) Diagrama de arquitectura software (paquetes) 7

[3.3.2.](#_heading=h.1y810tw) Diagrama de arquitectura del sistema (Diagrama de componentes) 7

[**3.4.**](#_heading=h.4i7ojhp) **Vista de procesos 7**

[3.4.1.](#_heading=h.2xcytpi) Diagrama de Procesos del sistema (diagrama de actividad) 8

[**3.5.**](#_heading=h.1ci93xb) **Vista de Despliegue (vista física) 8**

[3.5.1.](#_heading=h.3whwml4) Diagrama de despliegue 8

[***4.***](#_heading=h.2bn6wsx) ***ATRIBUTOS DE CALIDAD DEL SOFTWARE 8***

[**Escenario de Funcionalidad 8**](#_heading=h.qsh70q)

[**Escenario de Usabilidad 8**](#_heading=h.3as4poj)

[**Escenario de confiabilidad 9**](#_heading=h.1pxezwc)

[**Escenario de rendimiento 9**](#_heading=h.49x2ik5)

[**Escenario de mantenibilidad 9**](#_heading=h.2p2csry)

[**Otros Escenarios 9**](#_heading=h.147n2zr)

## Introducción

En la era de la información, la capacidad de manejar y procesar grandes volúmenes de datos de manera eficiente y segura es crucial para cualquier organización. El proyecto "Sistema de Automatización de Carga de Datos y Modificación de Reportes en el Almacén de Datos" tiene como objetivo principal mejorar la gestión y accesibilidad de datos mediante la automatización de la carga y modificación de datos almacenados en archivos CSV. Para lograr esto, utilizaremos Amazon S3 (Simple Storage Service) para el almacenamiento seguro y escalable de archivos, y AWS IAM (Identity and Access Management) para gestionar el acceso y la seguridad de dichos datos.

## Propósito

El propósito de este proyecto es desarrollar un sistema automatizado que facilite la carga, almacenamiento y modificación de datos en archivos CSV, así como la generación y actualización de reportes en un almacén de datos. Al utilizar S3 para el almacenamiento y IAM para la gestión de permisos, buscamos garantizar la seguridad, integridad y disponibilidad de los datos, permitiendo a los usuarios acceder a la información de manera eficiente y con los niveles de permiso adecuados. Este sistema no solo reducirá la carga de trabajo manual y los errores asociados, sino que también mejorará la precisión y la relevancia de los reportes generados.

## Alcance

El alcance del Sistema de Automatización de Carga de Datos y Modificación de Reportes incluye:

* **Almacenamiento Seguro con S3**:
  + Configuración de buckets en Amazon S3 para el almacenamiento de archivos CSV.
  + Automatización de la carga de archivos CSV desde diversas fuentes al bucket de S3.
  + Implementación de estrategias de versionado y backups para asegurar la durabilidad y disponibilidad de los datos.

**Gestión de Accesos con IAM**:

* Creación de usuarios y roles en IAM con permisos específicos para acceder y manipular los archivos CSV en S3.
* Definición de políticas de permisos detalladas para controlar las acciones que cada usuario o aplicación puede realizar en los buckets de S3.
* Implementación de autenticación multifactor (MFA) y otras medidas de seguridad para proteger el acceso a los datos.

**Automatización y Actualización de Reportes**:

* Desarrollo de scripts o aplicaciones para automatizar la transformación y carga de datos desde los archivos CSV almacenados en S3 al almacén de datos.
* Generación y actualización automática de reportes basados en los datos más recientes disponibles en el almacén de datos.
* Implementación de alertas y notificaciones para monitorear el estado de los procesos de carga y actualización de datos.

**Monitoreo y Mantenimiento**:

* Establecimiento de procedimientos para la monitorización continua del sistema.
* Implementación de herramientas y procesos para la gestión de incidencias y el mantenimiento preventivo.

## Definición

| **Termino** | **Definición** |
| --- | --- |
| **Actor** | Usuario del sistema que puede participar De un  caso de uso. |
| **Arquitectura de software** | conjunto de elementos estáticos, propios del diseño intelectual del sistema, que definen y dan forma tanto al código fuente como al comportamiento del software en tiempo de ejecución.  Naturalmente este diseño arquitectónico ha de ajustarse a las necesidades y requisitos del  proyecto |
| **Caso de Uso** | Secuencia de acciones que el sistema realiza, la cual proporciona un resultado de valor  observable. |
| **Diseño** | Actividad creativa que proyecta objetos para  después fabricarlos. |
| **Escenario** | Especifica el comportamiento y limita el interés de un área específica del sistema para 1 o varios  stakeholders. |
| **Paquetes** | Agrupaciones de casos de uso y actores por  funcionalidad que proveen. |

1. **Análisis de Requerimientos**

## Requerimientos funcionales

| Código | Requerimiento Funcional Inicial |
| --- | --- |
| RFI 01 | El sistema debe permitir la carga automática de datos desde múltiples fuentes. |
| RFI 02 | Debe validar los datos durante el proceso de carga. |
| RFI 03 | El sistema debe integrar los datos en el almacén de datos de manera estructurada. |
| RFI 04 | Debe generar alertas en caso de errores durante la carga de datos. |
| RFI 05 | El sistema debe ser capaz de programar cargas de datos periódicas. |
| RFI 06 | Debe permitir la modificación de reportes para apuntar al nuevo almacén de datos. |

* 1. Cuadro de requerimientos No funcionales

| Código | Requerimiento no funcional |
| --- | --- |
| RNF 01 | El sistema debe ser escalable para manejar volúmenes crecientes de datos. |
| RNF 02 | Debe tener una alta disponibilidad (99.9% uptime). |
| RNF 03 | El tiempo de respuesta para la carga de datos no debe exceder los 10 minutos. |
| RNF 04 | El sistema debe cumplir con las normativas de seguridad de datos. |

* 1. Cuadro de requerimientos funcionales Final

| Código | Requerimiento Funcionales Finales |
| --- | --- |
| RF 01 | El sistema debe permitir la configuración de reglas de validación de datos personalizadas. |
| RF 02 | Debe soportar la carga de datos en tiempo real. |
| RF 03 | El sistema debe permitir la extracción y carga de datos mediante APIs. |
| RF 04 | Debe proporcionar opciones de transformación de datos durante la carga. |
| RF 05 | El sistema debe permitir la replicación de datos a otros almacenes de datos. |
| RF 06 | El sistema debe permitir la administración de permisos de usuario para la carga de datos. |
| RF 07 | Debe integrar una solución de backup y recuperación de datos automatizada. |

## Representación Arquitectónica

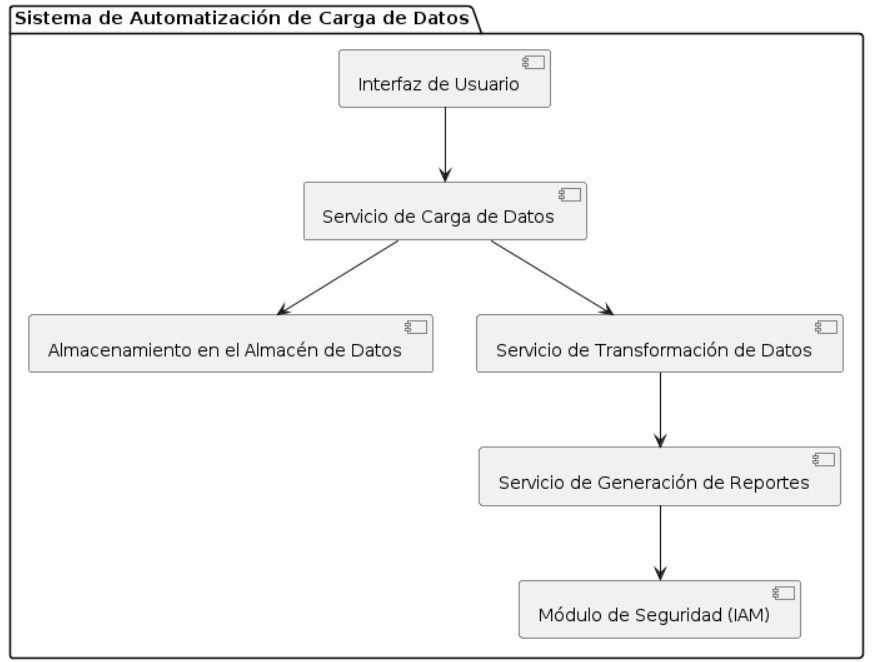
Para representar la arquitectura del "Sistema de Automatización de Carga de Datos y Modificación de Reportes en el Almacén de Datos" utilizando el modelo de 4+1 vistas, podemos dividirlo en cinco vistas diferentes: lógica, de desarrollo, de procesos, física y de escenarios. Aquí te presento cada una de estas vistas con una descripción y una representación arquitectónica.

### **1. Vista Lógica**

**Descripción**: La vista lógica describe la estructura de los módulos de software, sus relaciones y sus interacciones.

**Componentes principales**:

* **Interfaz de usuario**: Permite a los usuarios interactuar con el sistema.
* **Servicio de Carga de Datos**: Responsable de cargar y almacenar archivos CSV en S3.
* **Servicio de Transformación de Datos**: Procesa y transforma los datos según sea necesario.
* **Almacén de Datos**: Donde se almacenan los datos procesados.
* **Servicio de Generación de Reportes**: Genera y actualiza reportes basados en los datos almacenados.
* **Módulo de Seguridad (IAM)**: Gestiona el acceso y permisos a los datos y servicios.

****

### **2. Vista de Desarrollo**

**Descripción: La vista de desarrollo muestra la organización del software en módulos de código, bibliotecas y archivos.**

**Componentes principales:**

* **Repositorio de Código: Almacena el código fuente.**
* **Pipelines de CI/CD: Automatizan la construcción, prueba y despliegue del software.**
* **Entornos de Desarrollo y Pruebas: Donde los desarrolladores crean y prueban el código.**

**MODELO DE 4+1 VISTAS**

**Vista Lógica**: Describe los módulos principales del sistema y sus interacciones.

**Vista de Desarrollo**: Muestra la organización del software en módulos y componentes.

**Vista de Procesos**: Detalla los procesos concurrentes y cómo se comunican.

**Vista Física**: Describe la distribución física del sistema en hardware.

**Vista de Escenarios**: Proporciona ejemplos de cómo el sistema será utilizado en diferentes situaciones.

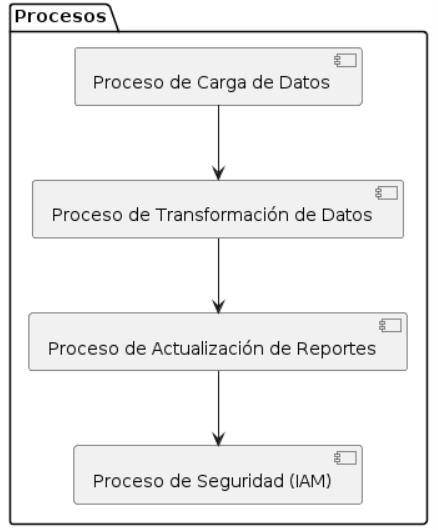
### 

### **3. Vista de Procesos**

**Descripción**: La vista de procesos describe los procesos del sistema y cómo se comunican e interactúan.

**Componentes principales**:

* **Carga de Datos**: Proceso que sube archivos CSV a S3.
* **Transformación de Datos**: Proceso que transforma los datos en CSV.
* **Actualización de Reportes**: Proceso que genera y actualiza reportes.



### **4. Vista Física**

**Descripción**: La vista física muestra la infraestructura de hardware y red en la que se ejecuta el sistema.

**Componentes principales**:

* **Instancia EC2**: Donde se ejecutan los servicios y aplicaciones.
* **Bucket S3**: Almacén de los archivos CSV.
* **Base de Datos**: Almacena los datos transformados.
* **Red VPC**: Proporciona la red privada virtual para la infraestructura.

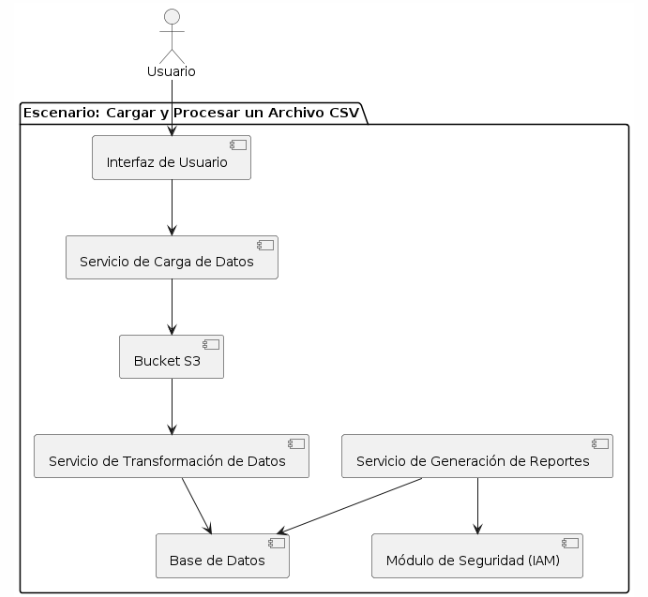
### **5. Vista de Escenarios**

**Descripción**: La vista de escenarios describe los casos de uso del sistema y cómo interactúan los componentes en situaciones específicas.

**Caso de uso principal**: "Cargar y procesar un archivo CSV"

**Escenario**:

1. **Usuario** sube un archivo CSV a través de la **Interfaz de Usuario**.
2. El **Servicio de Carga de Datos** recibe el archivo y lo almacena en un **Bucket S3**.
3. El **Servicio de Transformación de Datos** toma el archivo CSV del bucket y realiza las transformaciones necesarias.
4. Los datos transformados se almacenan en la **Base de Datos**.
5. El **Servicio de Generación de Reportes** accede a la base de datos y genera los reportes.
6. El **Módulo de Seguridad (IAM)** asegura que solo los usuarios autorizados puedan acceder y modificar los datos y reportes.



Diagrama

Descripción generada automáticamente

## Escenarios

* 1. **Vista Lógica**
* Diagrama de clases
* Diagrama de Secuencia

## Vista del Proceso

* Diagrama de Actividades

## Vista del Desarrollo

* Diagrama de Componentes
* Diagrama de Paquetes

## Vista Física

* Diagrama de Despliegue

## Estilos y Patrones Arquitectónicos

## Objetivos y limitaciones arquitectónicas

* 1. **Desempeño**
* Un usuario realiza una consulta al catálogo de productos en un momento normal de operación del sistema. El sistema muestra el resultado de la consulta en un tiempo no mayor a 3 segundos.

## Disponibilidad

* Una falla en un dispositivo I/O causa que un driver del sistema deje de funcionar en un momento normal de la operación. El sistema operativo deberá mostrar un mensaje de error retirando al dispositivo de la lista de dispositivos disponibles y continuar operando sin interrupción.

## Modificabilidad

* Un desarrollador agrega un caso de uso al esqueleto ejecutable de la arquitectura en tiempo de desarrollo. El código del caso uso es integrado y no requiere de modificaciones en más de 2 clases de la arquitectura.

## Rendimiento