

Proyecto de Integración y Automatización de Datos para IA para el proyecto: "Herramienta de monitoreo del consumo de recursos en AWS de la empresa Grupo ASD"

Daniela Rodriguez Fonseca¹, Diego Armando Baquero Gutiérrez²

¹⁻²Dpto. de Ingeniería

Universidad Central

Maestría en Analítica de Datos

Integración y Automatización de Datos para IA

Bogotá, Colombia

²drodriguez7@ucentral.edu.co, ²dbaquerog@ucentral.edu.co

October 7, 2023

Contents

1	Introducción (Max 250 Palabras) - (<i>Primera entrega</i>)	3
2	Características del proyecto de investigación que hace uso de Integración y Automatización de Datos para IA (Max 500 Palabras) - (<i>Primera entrega</i>)	3
2.1	Titulo del proyecto de investigación (Max 100 Palabras) - (<i>Primera entrega</i>)	3
2.2	Objetivo general (Max 100 Palabras) - (<i>Primera entrega</i>)	3
2.2.1	Objetivos específicos (Max 100 Palabras) - (<i>Primera entrega</i>)	3
2.3	Alcance (Max 200 Palabras) - (<i>Primera entrega</i>)	4
2.4	Pregunta de investigación (Max 100 Palabras) - (<i>Primera entrega</i>) .	4
2.5	Hipótesis (Max 100 Palabras) - (<i>Primera entrega</i>)	4
3	Reflexiones sobre el origen de datos e información (Max 400 Palabras) - (<i>Primera entrega</i>)	5
3.1	¿Cuál es el origen de los datos e información ? (Max 100 Palabras) - (<i>Primera entrega</i>)	5
3.2	¿Cuáles son las consideraciones legales o éticas del uso de la información? (Max 100 Palabras) - (<i>Primera entrega</i>)	5

3.3	¿Cuáles son los retos de la información y los datos que utilizara en Integración y Automatización de Datos para IA? (Max 100 Palabras) - (<i>Primera entrega</i>)	5
3.4	¿Qué espera de la utilización de Integración y Automatización de Datos para IA para su proyecto? (Max 100 Palabras) - (<i>Primera entrega</i>)	5
4	Diseño de integración y Automatización de Datos para IA (Diagrama) (<i>Primera entrega</i>)	6
5	Integración de datos (<i>Segunda entrega</i>)	6
6	Automatización de datos (<i>Segunda entrega</i>)	6
7	IA (<i>Segunda entrega</i>)	6
8	Proximos pasos (<i>Tercera entrega</i>)	7
9	Lecciones aprendidas (<i>Tercera entrega</i>)	8
10	Bibliografía	9

1 Introducción (Max 250 Palabras) - (*Primera entrega*)

La computación en nube se ha convertido en una solución popular para muchas empresas debido a sus beneficios de escalabilidad, flexibilidad y reducción de costos (Pragma USA Inc., s/f). Con la escalabilidad en la nube, las empresas pueden aumentar o disminuir el uso de sus recursos según sea necesario, lo que les permite ser más ágiles y eficientes en la gestión de sus operaciones. La flexibilidad es otra ventaja, ya que permite adaptarse a las fluctuaciones en la demanda de recursos informáticos de manera rápida y eficiente. Para el caso de Grupo ASD, tanto la escalabilidad como la flexibilidad son especialmente importantes ya que durante su operación se necesita ajustar los recursos de la mejor manera en función de la demanda de los proyectos.

Es importante destacar que, aunque el uso de Cloud Computing es muy beneficioso para las empresas, también puede ser complicado de controlar y monitorear, lo que a su vez puede generar costos innecesarios. AWS ofrece a las empresas la capacidad de acceder a una amplia gama de servicios y aplicaciones sin tener que preocuparse por el mantenimiento y la administración de la infraestructura (Amazon Web Services Inc., 2023).

Por lo que, desarrollar una herramienta que permita el monitoreo oportuno del consumo de recursos en la plataforma Amazon Web Services (AWS) proporcionará información detallada sobre el uso de los recursos en la nube, lo que permitirá a Grupo ASD tomar decisiones informadas sobre la asignación de recursos y ajustar el uso de estos.

2 Características del proyecto de investigación que hace uso de Integración y Automatización de Datos para IA (Max 500 Palabras) - (*Primera entrega*)

2.1 Título del proyecto de investigación (Max 100 Palabras) - (*Primera entrega*)

Desarrollo de una herramienta de análisis y monitoreo del consumo de recursos en AWS de la empresa Grupo ASD.

2.2 Objetivo general (Max 100 Palabras) - (*Primera entrega*)

Desarrollar una herramienta para el análisis y monitoreo de recursos de los diferentes proyectos implementados en la infraestructura en nube de Amazon Web Services (AWS) en la empresa Grupo ASD.

2.2.1 Objetivos específicos (Max 100 Palabras) - (*Primera entrega*)

- Realizar un proceso de recopilación y análisis de los datos de consumo de recursos de los diferentes proyectos de la empresa Grupo ASD a través de la infraestructura en nube AWS.

- Diseñar el modelo de datos con el fin de determinar la clasificación y categorías de los datos recopilados por medio de un documento de arquitectura de datos.
- Diseñar un modelo de análisis de datos para la identificación del uso y aprovisionamiento de los diferentes recursos utilizados, mediante una herramienta de visualización de datos.

2.3 Alcance (Max 200 Palabras) - (*Primera entrega*)

Se espera obtener una herramienta de visualización y un modelo de Machine Learning que permita a la empresa Grupo ASD monitorear de forma centralizada el consumo de recursos desplegados en AWS de cada una de sus cuentas; lo cual permitirá una mejor toma de decisiones y optimización de costos en la gestión de infraestructuras en la nube. Además, se espera que el trabajo realizado pueda servir como base para futuros proyectos similares.

2.4 Pregunta de investigación (Max 100 Palabras) - (*Primera entrega*)

¿Qué plataforma se puede implementar para el desarrollo de una herramienta efectiva de monitoreo centralizado del consumo de recursos en AWS que permita a la organización optimizar sus costos y recursos de manera proactiva?

2.5 Hipótesis (Max 100 Palabras) - (*Primera entrega*)

Se cree que mediante la implementación de un sistema de monitoreo que integre y automatice la recopilación de datos de consumo de recursos en AWS, se podrá identificar patrones de uso, predecir picos de demanda y generar alertas, lo que resultará en una reducción de costos y una optimización de recursos para la organización.

3 Reflexiones sobre el origen de datos e información

(Max 400 Palabras) - (*Primera entrega*)

3.1 ¿Cuál es el origen de los datos e información ? (Max 100 Palabras) - (*Primera entrega*)

Los datos serán obtenidos de la cuenta corporativa de Amazon Web Services (AWS) de la empresa Grupo ASD del servicio de Cloudwatch configurado para cada proyecto que despliega infraestructura en esta plataforma.

3.2 ¿Cuáles son las consideraciones legales o éticas del uso de la información? (Max 100 Palabras) - (*Primera entrega*)

Como requerimiento de la empresa Grupo ASD los datos de los proyectos como nombres, descripción o propósito no podrán ser incluidos en el documento. Si se requiere su uso deberán ser ofuscados.

3.3 ¿Cuáles son los retos de la información y los datos que utilizara en Integración y Automatización de Datos para IA? (Max 100 Palabras) - (*Primera entrega*)

Los datos de consumo de recursos de cada uno de los proyectos serán extraídos a través de la API de AWS, ya sea mediante una lambda para enviar los datos desde AWS o exportados de los registros almacenados en CloudWatch. Posteriormente, los datos serán almacenados en una base de datos y procesados mediante scripts desarrollados en algún lenguaje como JavaScript, Python o Go.

El reto se encuentra en automatizar la extracción de los datos para que esta sea lo más eficiente posible y lograr que se puedan obtener los datos en periodos cortos de tiempo. Así mismo, la integración de los datos, ya que estos provienen de diferentes cuentas y distintos recursos por cada cuenta.

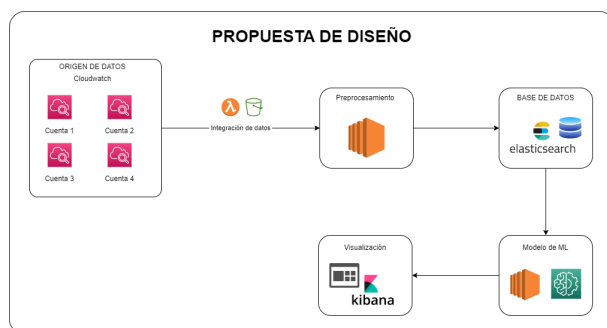
3.4 ¿Qué espera de la utilización de Integración y Automatización de Datos para IA para su proyecto? (Max 100 Palabras) - (*Primera entrega*)

Se espera obtener los conocimientos necesarios en diferentes herramientas para lograr automatizar los procesos de extracción e integración de los datos desde su origen, la implementación del pre-procesamiento y paso por el modelo de machine Learning y finalmente visualización de los datos.

Por otro lado, se desea realizar la automatización de la extracción de los datos desde el servicio de CloudWatch de AWS que permita recopilar datos en periodos cortos o tiempo real desde las múltiples cuentas, transformarlos y limpiarlos eficientemente implementando estos procesos de forma automatizada, para luego alimentar el algoritmo de Machine Learning para el monitoreo y análisis de consumo de recursos en AWS.

4 Diseño de integración y Automatización de Datos para IA (Diagrama) *(Primera entrega)*

A continuación, se presenta el diagrama del diseño de la propuesta de la automatización e integración de datos para el proyecto propuesto.



En primer lugar, se identificó el origen de los datos a partir del servicio de CloudWatch configurado en cada una de las cuentas de AWS de Grupo ASD. Este proceso se planea realizar utilizando los servicios de Amazon Lambda y S3, en donde se busca automatizar la extracción de los logs mediante una función lambda y posteriormente guardarlos en un bucket.

En segundo lugar, se realizará la integración y preprocesamiento de los datos, para después ser ingestados en la base de datos de ElasticSearch, posteriormente se implementará un modelo de Machine Learning.

Finalmente, se realizarán los diferentes tableros para el monitoreo de los consumos con la herramienta de visualización Kibana.

5 Integración de datos *(Segunda entrega)*

6 Automatización de datos *(Segunda entrega)*

7 IA *(Segunda entrega)*

8 Proximos pasos (*Tercera entrega*)

9 Lecciones aprendidas (*Tercera entrega*)

10 Bibliografía

- Amazon Web Services Inc. (2023). Información general sobre Amazon Web Services Documento técnico de AWS.
- Pragma USA Inc. (s/f). Computación en la nube: todo lo que un experto debe saber. Recuperado el 30 de abril de 2023, de <https://www.pragma.com.co/academia/conceptos/computacion-en-la-nube>