



Diseño e implementación de una propuesta arquitectura de ingeniería de datos en la nube para la construcción de un repositorio centralizado con capacidades (ETL) para la Unidad de Educación Presencial de la Universidad de las Fuerzas Armadas ESPE basado en la integración eficiente de datos en la nube siguiendo los parámetros de negocio y mejorando la toma de decisiones.

Rodríguez Preciado, Carlos Andrés

Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Trabajo de integración curricular, previo a la obtención del título de Ingeniero en

Tecnologías de la Información

Ing. López López, Andrea Margarita, MSc.

6 de marzo del 2024



Plagiarism and AI Content Detection Report

Trabajo Titulacion_Andres Rodriguez (...)

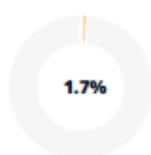
Scan details

Scan time:
March 1th, 2024 at 0:15 UTC

Total Pages:
65

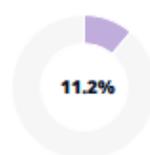
Total Words:
16156

Plagiarism Detection



Types of plagiarism		Words
Identical	0.2%	31
Minor Changes	0.1%	19
Paraphrased	1.3%	206
Omitted Words	7%	1133

AI Content Detection



Text coverage		Words
AI text	11.2%	1787
Human text	88.8%	13236

[Learn more](#)

🔍 Plagiarism Results: (11)

🌐 **Arquitectura de Datos en la Nube: Tendencias y Perspectivas Futuras** 0.6%

<https://es.linkedin.com/pulse/arquitectura-de-datos-en-la-nube-tendencias-y-futuras-jose-loza>

JOSE LOZA

Acepta...

🌐 **La sinergia de la arquitectura de datos en la nube y la inteligencia artificial** 0.5%

<https://es.linkedin.com/pulse/la-sinergia-de-arquitectura-datos-en-nube-y-artificial-jose-loza>

JOSE LOZA

Acepta...

🌐 **SCITEPRESS - SCIENCE AND TECHNOLOGY PUBLICATIONS** 0.2%

<https://www.scitepress.org/publishedpapers/2015/55297/>

An Overview of Cloud Services Adoption Challenges in Higher Education Institutions ...

.....

Ing. López López, Andrea Margarita, MSc.

C. C 1600687733



Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Certificación

Certifico que el trabajo de integración curricular: **“Diseño e implementación de una propuesta arquitectura de ingeniería de datos en la nube para la construcción de un repositorio centralizado con capacidades (ETL) para la Unidad de Educación Presencial de la Universidad de las Fuerzas Armadas ESPE basado en la integración eficiente de datos en la nube siguiendo los parámetros de negocio y mejorando la toma de decisiones.”** fue realizado por el señor **Rodríguez Preciado, Carlos Andrés**, el mismo que cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, además fue revisado y analizada en su totalidad por la herramienta de prevención y/o verificación de similitud de contenidos; razón por la cual me permito acreditar y autorizar para que se lo sustente públicamente.

Sangolquí, 6 de marzo del 2024

.....

Ing. López López, Andrea Margarita, MSc.

C. C 1600687733



Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Responsabilidad de Autoría

Yo, **Rodríguez Preciado, Carlos Andrés**, con cédula de ciudadanía N° **1718934381**, declaro que el contenido, ideas y criterios del trabajo de integración curricular: **“Diseño e implementación de una propuesta arquitectura de ingeniería de datos en la nube para la construcción de un repositorio centralizado con capacidades (ETL) para la Unidad de Educación Presencial de la Universidad de las Fuerzas Armadas ESPE basado en la integración eficiente de datos en la nube siguiendo los parámetros de negocio y mejorando la toma de decisiones.”** es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos, y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 6 de marzo del 2024

Rodríguez Preciado, Carlos Andrés

C.C.:1718934381



Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Autorización de Publicación

Yo **Rodríguez Preciado, Carlos Andrés**, con cédula de ciudadanía N° 1718934381, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de integración curricular:

Título: “Diseño e implementación de una propuesta arquitectura de ingeniería de datos en la nube para la construcción de un repositorio centralizado con capacidades (ETL) para la Unidad de Educación Presencial de la Universidad de las Fuerzas Armadas ESPE basado en la integración eficiente de datos en la nube siguiendo los parámetros de negocio y mejorando la toma de decisiones.” en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 6 de marzo del 2024

Rodríguez Preciado, Carlos Andrés

C.C.:1718934381

Dedicatoria

Dedico esta tesis a mis amados padres y querido hermano, pilares fundamentales en mi vida teniendo su inquebrantable apoyo y constante ánimo para mí han sido los que me han permitido mantenerme firme aún en todas las situaciones a través de este desafiante proceso y sobresalir por encima de todo. Gracias por iluminar mi camino con mucho de su amor y guía, con todo cariño.

Carlos Andrés Rodríguez Preciado

Agradecimientos

Con un corazón lleno de gratitud quisiera expresar mi sincero agradecimiento a las personas que hicieron posible esta etapa académica. A mi amada familia en particular a mi madre, padre y hermano que son mis mentores, guiándome a través de los desafíos, fortaleciéndome con su amor incondicional, apoyo constante y comprensión. Ustedes para mi han sido la fuente de mi motivación, sus sacrificios y confianza han infundido una fuerza incansable para superar los obstáculos.

A los maestros que han esculpido mi conocimiento y habilidades, gracias. Su grande experiencia y sabiduría han sido fundamentales para mi crecimiento académico. Una mención especial a la MSc. Andrea López quien fue mi primer acercamiento con los datos de igual forma con sus enseñanzas, comentarios, sugerencias críticas han ido enriqueciendo este trabajo y especialmente por haber confiado en mí. De igual forma al Ing. Omar Quimbita, Mgtr. por su constante apoyo y motivación en el proceso de la elaboración de este trabajo de titulación.

Agradezco a mis amigos de Banco que con su experiencia profesional han enriquecido mi comprensión y habilidad para manejar y modelar adecuadamente los datos. Agradezco a los ingenieros Bryan Ron, Diego DLT, Shinaryna C, Soledad G, Miguel A, Evelyn P, y todos quienes compartieron su conocimiento y apoyo.

A la Unidad de Educación Presencial (UEP), le debo un agradecimiento especial por darme la oportunidad de desarrollar la primera arquitectura de datos en la nube. También a la ingeniera Nathaly Ñacata, por unirse a este emocionante reto de automatización.

Finalmente, agradezco a todos los que me han acompañado en este recorrido académico, su impacto en mi vida es incalculable.

Carlos Andrés Rodríguez Preciado

Índice de Contenidos

Resumen.....	14
Abstract.....	15
Capítulo I.....	16
Introducción.....	16
Antecedentes	18
Implementación de Tecnología en Instituciones Educativas y Avances en Ingeniería de Datos.....	19
Planteamiento del Problema	20
Brecha en la Gestión de Datos.....	20
Revolución de los datos	20
Necesidad de mejora constante	20
Justificación.....	21
Objetivos	22
Objetivo General:	22
Objetivos Específicos:.....	22
Hipótesis.....	23
Capítulo II.....	25
Marco Metodológico.....	25
Diseño de la Investigación	25
Tipo de investigación	25
Recolección de Datos	26
Análisis de Datos.....	27
Planteamiento de la Revisión Sistemática.....	28

Revisión Sistemática.....	28
Propósito de la Revisión	28
Preguntas de la Investigación.....	28
Población y Muestra.....	30
Estado del arte.....	33
Arquitecturas de ingeniería de datos en la nube	33
Tendencias de la arquitectura de datos.....	34
Aplicación de las tecnologías en la nube en instituciones educativas....	34
Marco Teórico	35
Integración de Datos	35
Procesos ETL en la Nube	36
Avances de la Ingeniería de Datos en la Nube Aplicada a la Educación Superior y a la Innovación Educativa:	36
Alineación de Directrices Empresariales con Mejora Continua:	37
Método de trabajo	38
Tableros de visualización.....	40
Componentes Esenciales de una Arquitectura en la Nube	42
Capítulo III.....	44
Análisis de la Estructura de Datos en el Ámbito Académico	44
Requerimientos de Negocio para la Unidad de Educación Presencial (UEP) ...	44
Infraestructura Cloud	46
Componentes Clave de la Arquitectura de Ingeniería de Datos en la Nube.....	49
Resource group.....	50
Data factory (V2)	51
Virtual machine.....	52
SQL database	53

Logic App.....	54
Storage account	55
Key vault.....	56
Azure Databricks Service.....	57
Requisitos de instalación en Azure y versiones.....	58
Requisitos de Uso Funcional.....	60
Desarrollo Arquitectura para el Flujo de Información.....	61
Funcionamiento y aplicación práctica del proceso de implementación.....	63
Análisis de Costos y Gestión Eficiente de Recursos en la Adopción de Servicios en la Nube.....	63
Capítulo IV	67
Resultados	67
Resultados de Tableau	67
Pruebas Realizadas y Contribución al Logro de Objetivos	72
Capítulo V	77
Conclusiones.....	77
Recomendaciones	78
Bibliografía	80

Índice de Tablas

Tabla 1 <i>Requisitos de instalación en Azure</i>	58
Tabla 2 <i>Requisitos de Uso Funcional</i>	60
Tabla 3 <i>Costos de Implementación en Azure</i>	64

Índice de Figuras

Figura 1 <i>Estrategia de trabajo: Modelo en Cascada</i>	38
Figura 2 <i>Diagrama de Infraestructura Cloud</i>	47
Figura 3 <i>Detalle de la Factoría de Datos</i>	48
Figura 4 <i>Clúster</i>	49
Figura 5 <i>Resource group</i>	51
Figura 6 <i>Data factory (V2)</i>	52
Figura 7 <i>Virtual machine</i>	53
Figura 8 <i>SQL database</i>	54
Figura 9 <i>Logic App</i>	55
Figura 10 <i>Storage account</i>	56
Figura 11 <i>Key vault</i>	57
Figura 12 <i>Azure Databricks Service</i>	58
Figura 13 <i>Estructura de Carpetas en Data Lake Storage</i>	61
Figura 14 <i>Acceso y Copia de Archivos en la Máquina Virtual</i>	63
Figura 15 <i>Costos máquina virtual</i>	66
Figura 16 <i>Análisis de las Capacidades de las Aulas</i>	67
Figura 17 <i>Uso de aulas en la Universidad de las Fuerzas Armadas (ESPE)</i>	69
Figura 18 <i>Detalle del Uso de Aulas por Hora y Día</i>	70

Figura 19 <i>Análisis Avanzado: Modelo de Machine Learning con uso de Regresión Lineal de Estudiantes por Materia</i>	71
Figura 20 <i>Eficiencia del Aplicativo</i>	75
Figura 21 <i>Cumplimiento de Requisitos</i>	76

Resumen

Este trabajo de titulación se basa en una propuesta de diseño y desarrollo de una arquitectura de Ingeniería de Datos en la Nube para la Unidad de Educación Presencial (UEP) de la Universidad de las Fuerzas Armadas (ESPE), en lo personal es un progreso significativo en el manejo de datos académicos y también un progreso en el salto para realizar automatizaciones en la nube siendo una de las últimas tecnologías. El proyecto empieza con un análisis detallando sobre las necesidades y especificaciones que se obtiene de la Unidad de Educación Presencial (UEP) para luego guiar el diseño de una arquitectura de datos en la nube que cumple con los criterios más elevados de seguridad, escalabilidad y eficacia operativa. Realizamos la arquitectura, ensamblando componentes de alta tecnológicos como el Grupo de Recursos de Azure, Azure Data Factory, Máquinas Virtuales, Bases de Datos SQL, Servidores SQL, Logic Apps, Cuentas de Almacenamiento, Almacenes de Claves y el Servicio Azure Databricks, estos elementos seleccionados de suma importancia tienen como objetivo: automatizar los procesos de toda nuestra arquitectura, simplifican la integración y administración de datos que defienden la privacidad y disponibilidad de la información que son de suma importancia. La propuesta es una implementación efectiva de esta arquitectura y automatización que conduce a una administración más eficiente de la infraestructura académica, optimizando la utilización de recursos y potenciando las decisiones estratégicas a través de un análisis profundo de los datos históricos y la visualización, utilizando herramientas de última generación como Tableau.

Palabras clave: Arquitectura de Datos, Ingeniería de Datos, Automatización de Procesos, Integración de Datos.

Abstract

This dissertation presents a comprehensive proposal for the design and development of a Cloud Data Engineering architecture tailored for the Onsite Education Unit (UEP) of the Universidad de las Fuerzas Armadas (ESPE). This endeavor signifies a substantial advancement in academic data management and marks a pivotal transition towards cloud-based automation, harnessing leading-edge technologies. The project commences with a meticulous examination of the UEP's requirements and specifications, guiding the formulation of a cloud data architecture that adheres to the utmost standards of security, scalability, and operational efficiency. We proceed to construct the architecture, amalgamating sophisticated components such as Azure Resource Group, Azure Data Factory, Virtual Machines, SQL Databases, SQL Servers, Logic Apps, Storage Accounts, Key Vaults, and Azure Databricks Service. These meticulously chosen elements are purposed to streamline processes across our architecture, simplifying data integration and management, all the while safeguarding the privacy and availability of crucial information. The proposal epitomizes an efficacious implementation of this architecture and automation, culminating in the enhanced management of academic infrastructure, optimization of resource allocation, and augmentation of strategic decision-making through thorough analysis of historical data and visualization leveraging state-of-the-art tools such as Tableau.

Keywords: Data Architecture, Data Engineering, Process Automation, Data Integration.

Capítulo I

Introducción

La tecnología juega un rol muy importante en todas las áreas en el cual ayuda y permite como usuarios de la tecnología tener la libertad de mejorar en las tareas recurrentes gracias a la automatización. El mundo de las nuevas tecnologías que destaca la computación en la nube, que ofrece importantes un sin número de beneficios en el proceso de Extracción, Transformación y Carga (ETL). Esta sistemática no solo gestiona efectivamente grandes cantidades de datos, sino que también se puede modelar al gusto y se adapta a las necesidades del usuario en un entorno flexible y adaptable, lo cual es muy relevante para estas mejorar las decisiones empresariales.

Esta orientación tiene mayor impacto en la sociedad al considerar el manejo de datos con herramientas especializadas en entornos educativos, donde se estima que tenga un gran número de beneficios significativos. La infraestructura en la nube se vuelve un recurso clave, especialmente en escenarios donde se reciben planificaciones en formato Excel. Tener un plan en el que se pueda almacenar como un repositorio en la nube y un panel de control de fácil manejo debería mejorar marcadamente el proceso de toma de decisiones. La capacidad de acceder centralizadamente a los datos, junto con una visualización clara, debería sentar las bases para decisiones informadas y eficaces, potencialmente mejorando el rendimiento y la eficiencia en los entornos educativos. Además, dentro del ámbito de la innovación en la gestión universitaria, investigadores como (Khan et al., 2018) han explorado propuestas similares. Khan ha abordado la importancia de mantener un depósito de datos educativos como solución a desafíos semejantes en la gestión universitaria. En la investigación de (Khan et al., 2018) plantea un modelo de depósito de datos educativos como servicio (eDWaaS) que le sirvió para el manejo de datos históricos que son provenientes de varias universidades teniendo como

resultado un buen análisis, lo que se vuelve como punto de partida de información para el desarrollo del presente trabajo.

En el presente trabajo de titulación, se lleva a cabo un diseño de una arquitectura de datos de un ambiente en el que engloba el 100 % de la nube. El propósito es lograr que se cree un repositorio centralizado que pueda almacenar la información crítica en el proceso de asignación de aulas estos datos son provenientes de los especialistas que manejan el área de Unidad de Educación Presencial (UEP) en Excel.

También se discutirá sobre las capacidades de un ETL que son conocidos como Extracción, Transformación y Carga, en la cual se encargará en integrar en una infraestructura de nube. La mayoría de los datos, recopilados en formatos Excel, serán sometidos a un meticuloso proceso de extracción utilizando las funcionalidades avanzadas de los ambientes en la nube, para luego hacer de estos datos una transformación y reformateo siguiendo parámetros de lógica. Este conjunto de datos transformados se almacenará en un repositorio centralizado en el que su gestión se encontrará en la nube.

Tomando como referencias herramientas que nos facilitan el elaborar un ETL en la nube como Azure, la elección de implementar una infraestructura en la nube se convierte primeramente innovador al igual que una decisión estratégica y de alta calidad. Esta decisión de implementar no solo facilita el proceso de extracción de datos, sino también certifica que los datos transformados ya sean de gran valor y relevancia, para que al finalizar estén almacenados de manera segura siguiendo un estándar en un repositorio centralizado en la nube.

Este proyecto se trabajará en la nube, con las facilidades que nos da el portal Azure utilizando los mayores tipos de recursos para certificar una extracción eficiente y segura hacia los datos. De igual forma se establecerá una manera idónea de conectar Tableau para

visualizar y analizar los datos correctos que fueron extraídos de varios documentos de Excel propuestos hasta el almacenamiento en la base de datos. La herramienta de visualización de datos nos permite presentar la información de forma clara y segura para todos los aspectos esenciales que impulsan el crecimiento del entorno educativo como por ejemplo el uso de mejora en decisiones para la infraestructura de la universidad y eso con ayuda de dar un análisis con relación al uso de aulas y laboratorios. La información que también es proporcionada tiene el poder de abrir una puerta a la toma de decisiones informadas y al avance de la inteligencia de negocio educativo.

El valor añadido para este trabajo de titulación se radica en tener un resultado positivo sobre la evolución constante de la educación presencial especialmente en el uso de la infraestructura de la universidad, vinculando la toma de decisiones encaminadas mediante una práctica sustentable. En consecuencia, se tiene como finalidad potenciar el uso de estas tecnologías de gran valor, que son usadas en la actualidad por grandes empresas para que las nuevas generaciones comiencen a usar las tecnologías en la nube, creando un ambiente de automatizaciones y con ella una mejor toma de decisiones en entornos de carácter académico; finalmente esto contribuye con mejoras significativas en la Universidad de las Fuerzas Armadas (ESPE).

Antecedentes

En la última década, se ha podido tener la presencia de un crecimiento asombroso y una adopción de la tecnología en la nube en diferentes sectores, destacando tanto en las empresas bancarias como en el ámbito educativo. Empresas y organizaciones a escala global están en la era total de manejo de datos trasladando sus procesos a la nube. La razón de esto es muy sencilla porque la tecnología en la nube tiene un gran impacto al ser escalable, flexible y efectivo. Un ejemplo relevante de esto es el Vicerrectorado de Docencia, que se encuentra en

la búsqueda de una solución fuerte para la automatización de datos en la Universidad de las Fuerzas Armadas (ESPE).

Implementación de Tecnología en Instituciones Educativas y Avances en Ingeniería de Datos

En los últimos años, el modelo “Data Warehouse as a Services” (DWaaS) ha surgido con un crecimiento notable en el mercado. Sin embargo, los avances en cuestión de documentación significativos en este campo son muy escasos hasta el momento. El término DWaaS apareció por primera vez en 2012; a pesar de ello, los problemas de un buen rendimiento debido a la gran cantidad de datos y los procesos ETL que consumían bastante espacio y rendimiento continúan siendo un motivo de preocupación. Algunos estudios han intentado implementar los procesos ETL y OLAP en la plataforma de la nube, pero no han abordado claramente los problemas de escalabilidad hablando en el 2018. Se han propuesto diversas bodegas de datos “Data Marts” para el dominio educativo, pero no se puede identificar claramente en la literatura una implementación sólida de (DWaaS), en este dominio según lo redactado por (Khan et al., 2018).

Se plantea un sistema de repositorio de datos en la nube para la Unidad de Educación Presencial (UEP). Teniendo como solución una implementación de una arquitectura sólida y robusta de datos que otorgará diversas ventajas, y entre ellas mejora la eficiencia en la toma de decisiones a nivel: administrativo y académico. Esta arquitectura permitirá una gestión más eficiente de los datos, con mayor seguridad y escalabilidad.

La arquitectura de datos, al igual que nuestro mundo universitario tiene una evolución según (Khan et al., 2018) menciona que la metodología ETL (Extracción, Transformación y Carga) se ha convertido en un pilar para la gestión eficiente de grandes volúmenes de datos. Plataformas en la nube como AWS, Azure y Google Cloud son protagonistas en este gran

escenario digital por su destacada capacidad, donde la Universidad de las Fuerzas Armadas busca sobresalir y encontrar su lugar en este ámbito implementando prácticas avanzadas de ingeniería de datos.

Planteamiento del Problema

Brecha en la Gestión de Datos

La Universidad de las Fuerzas Armadas ESPE se enfrenta a retos actuales en la eficaz administración de información en la Unidad de Educación Presencial (UEP). La ausencia de una arquitectura de datos y automatizado obstaculiza la toma de decisiones bien fundamentadas, añadiendo una tarea práctica/operativa considerables a los responsables de manejar la información. La carencia de profesionales, tales como especialistas e ingenieros en tecnologías de la información, tanto en análisis como en diseño y herramientas, agudiza aún más la situación.

Revolución de los datos

El escaso estudio en la implementación de una arquitectura de información en la nube en el contexto educativo evidencia la poca familiaridad y la falta de práctica en esta área. Esta postura se sostiene en gran medida por el hecho de que estas implementaciones tienen un costo y, por ende, requieren una considerable inversión. La generación y adaptación de estas tecnologías en el entorno educativo proporcionan un escenario prominente para mejoras sustanciales, optimizando al personal en tareas repetitivas y excesivamente prácticas.

Necesidad de mejora constante

La falta de utilización de grandes datos numéricos como también de cualquier índole que logran una automatización completa. Iniciando desde el desembarco de datos crudos hasta el proceso de entregas de datos finalizados en visualizadores ayudando con los procesos

administrativos y académicos en la Unidad de Educación Presencial (UEP); usando la exploración de métodos específicos que permitan un manejo adecuado y eficaz con la información proporcionada de planificación por periodo en el cual esta data tiene campos de gran valor como el campus, departamento, asignatura, estado, edificio, aula, capacidad, hora de inicio, hora fin y entre otros, todos estos campos existen sin embargo es importante dar un uso efectivo para el análisis con la infraestructura.

Justificación

Con la necesidad de una correcta gestión de datos, utilizando arquitecturas para la Universidad de las Fuerzas Armadas (ESPE), se ha manifestado la necesidad de comenzar la administración de datos en la Unidad de Educación Presencial (UEP). Esto se debe principalmente a que en este sector se maneja un importante volumen de información, cuyo aprovechamiento es crucial para el avance de la Institución.

Implementar un modelo de arquitectura de ingeniería de datos basado en la nube con el objetivo de automatizar de forma eficiente la información para la toma de decisiones que nos sirve al dar reportes con respecto a un mejor uso de aulas y laboratorios, y además reducir la carga laboral de las personas encargadas en la gestión de datos de la Unidad de Educación Presencial (UEP). Esta propuesta presenta múltiples beneficios notables, que incluyen un avance en la calidad y celeridad de las decisiones, gracias a la unificación e inmediata disponibilidad de la información relevante. Además, se pretende optimizar el tiempo dedicado a tareas operativas y administrativas, permitiendo que el personal de la Unidad de Educación Presencial (UEP) tenga un enfoque en actividades estratégicas y de mayor impacto.

Este proceso es innovador que destaca a la universidad de las Fuerzas Armadas (ESPE), dado que no se han encontrado precedentes del uso de este prototipo en investigaciones y referencias previas, demostrando que esta implementación será un símbolo

de transformación hacia el futuro de la gestión de la información. Desde el punto de vista práctico y teórico, los resultados podrían generar una basta mejora en la Unidad de Educación Presencial en la ESPE.

Objetivos

Objetivo General:

Automatizar procesos de planificación académica semestral referente a la asignación de infraestructura de la Universidad (aulas y laboratorios) mediante el análisis de datos históricos y proyecciones, creación de un repositorio centralizado y el uso de herramientas de integración de datos en la nube y de visualización y generación de reportes para la Unidad de Educación Presencial (UEP).

Objetivos Específicos:

Analizar la data histórica y proyección de datos de planificación académica por semestre sobre procesos de asignación de infraestructura a los departamentos en la Unidad de Educación Presencial (UEP).

Diseñar el flujo de datos y los procesos de extracción, transformación y carga (ETL) y su configuración en la nube.

Crear y desplegar tableros interactivos en Tableau para la visualización de datos.

Alcance

Se propone la construcción de un tablero en Tableau para monitorear la disponibilidad y distribución de aulas:

- Se presentará un tablero interactivo que contendrá una regresión lineal para examinar el crecimiento de la demanda de aulas en base a la cantidad de estudiantes.

- Permitirá una visualización de la capacidad de aulas en su uso para así poder detectar inconsistencias en su distribución, como aulas y laboratorios subutilizados.
- Identificar aulas y laboratorios de la universidad que no hayan sido ocupadas durante el semestre para optimizar su asignación.

Se propone una implementación y operación de un repositorio en la nube:

- Se establecerá un repositorio centralizado en Microsoft Azure para poder almacenar y administrar una variedad de datos vinculados a la disponibilidad de aulas y la carga académica en Excel. Este repositorio también podrá alojar otros datos, como metadatos, parámetros e información adicional.
- Esta arquitectura incluirá una habilidad para implementar una solución integral de extracción, transformación y carga de datos en la nube, posterior por el despliegue del modelo desarrollado en Tableau con las medidas de seguridad correspondientes.
- La arquitectura propuesta facilitará una completa integración de datos en la nube, desde su extracción, transformación en Databricks hasta su carga en Tableau. Donde permitirá un acceso rápido y seguro a todos los datos necesarios.

Hipótesis

Hipótesis General: El núcleo de esta tesis es que la integración de datos en la nube, mediante la creación de un repositorio centralizado con capacidades ETL, mejorará considerablemente el proceso de toma de decisiones en la Unidad de Educación Presencial (UEP).

Hipótesis Específica 1: Un análisis exhaustivo de los requerimientos de datos de la UEP permitirá comprender a fondo las necesidades específicas de información académica y

administrativa, resultando en una gestión más eficaz de la información para la toma de decisiones estratégicas.

Hipótesis Específica 2: La elaboración detallada del diseño del flujo de datos y la implementación de procesos de extracción, transformación y carga (ETL) facilitará la conversión efectiva de datos brutos en información coherente y valiosa, mejorando la capacidad de planeación y asignación de infraestructura académica.

Hipótesis Específica 3: Un meticuloso ajuste y preparación del entorno en la nube utilizando plataformas avanzadas como Azure asegurará una extracción segura y eficiente de datos, garantizando la disponibilidad e integridad necesarias para una toma de decisiones informada y precisa.

Hipótesis Específica 4: La efectiva implementación de procesos ETL y su integración con herramientas de visualización como Tableau facilitará un acceso ágil y seguro a datos fiables, mejorando la capacidad de análisis y la eficiencia en la gestión de la infraestructura académica.

Hipótesis Específica 5: La exitosa implementación de un tablero de control para visualizar la disponibilidad y carga de aulas; el adecuado funcionamiento del repositorio en la nube conducirá a mejoras significativas en la toma de decisiones y eficiencia operativa en la UEP al brindar una visión detallada de la ocupación de aulas en relación con su capacidad y uso.

Capítulo II

Marco Metodológico

Diseño de la Investigación

Se opta por el diseño cuasiexperimental por su habilidad de ponderar el efecto de una intervención en un ambiente educativo controlado; tal es el caso de la adopción de una estructura de ingeniería de datos en la nube para la Unidad de Educación Presencial (UEP) de la Universidad de las Fuerzas Armadas ESPE. Al contrario, a los discursos experimentales comunes, donde se dispone aleatoriamente a los sujetos en grupos de control y tratamiento, dicha asignación aleatoria puede no ser factible o ética en un diseño cuasiexperimental. Por tanto, este diseño será útil para evaluar rigurosamente el impacto de la ingeniería de datos en la nube para la toma de decisiones y eficiencia operativa dentro de la Unidad de Educación Presencial (UEP), esta decisión metodológica se basa en la investigación de (Chavarría Romero, 2021), quien empleó un enfoque similar para valorar la precisión de la implementación de un modelo de Business Intelligence en el área de Trade Marketing de la marca Puma.

Tipo de investigación

En la Universidad de las Fuerzas Armadas ESPE, se realizará una línea de investigación exhaustiva para lo cual este proyecto aprecia un enfoque metodológico mixto, abordando aspectos tanto cualitativos como cuantitativos. Para este tipo de investigación lo que se realizará en primer lugar será seleccionar con la intención de obtener una comprensión total, sobre la implementación y el impacto que va a dar a la arquitectura de ingeniería de datos en la nube en la Unidad de Educación Presencial (UEP) como proceso de mejora en la toma de decisiones.

La investigación cualitativa tiene como objetivo el profundizar en las experiencias, perspectivas, y necesidades del personal administrativo de la Unidad de Educación Presencial

(UEP) con respecto a los datos y la toma de decisiones por ese motivo los relatos mediante los funcionales y análisis de contenido se trabaja en entender los retos y oportunidades que rodean la implementación de esta tecnología mencionado por (Alfaro, 2023).

Paralelo en la investigación cuantitativa tendrá como objetivo el hecho de la recopilación y análisis de datos relacionados con la eficacia y el impacto de la arquitectura de datos en la nube, en otras palabras examinar los datos para validar su coherencia y analizar los datos recogidos siendo capaces de identificar patrones, tendencias y conexiones significativas creando un diccionario de datos como lo menciona (Alfaro, 2023).

La unión de estos dos enfoques metodológicos sirve para el presente trabajo de titulación donde tiene un gran impacto debido a que nos proporciona una visión más completa y rigurosa sobre la implementación de la arquitectura de datos en la nube en la Unidad de Educación Presencial (UEP) y su efecto en la toma de decisiones. Esta edificación de datos cualitativos y cuantitativos fortalece la validez y confiabilidad de los resultados, estableciendo un fundamento sólido para las conclusiones y recomendaciones del presente estudio.

Recolección de Datos

Evaluación de Requerimientos:

En cuanto a la recolección de datos para entender a base las necesidades y las demandas específicas relacionadas con los datos, se realizará reuniones con el equipo directivo de la Unidad de Educación Presencial (UEP). Durante este proceso, también se llevará a cabo un análisis absoluto sobre los documentos propuestos de los procedimientos ya existentes en la Unidad de Educación Presencial (UEP), con la intención de identificar posibles mejoras que puedan abordarse a través de la implementación de la arquitectura de ingeniería de datos en la nube.

Diseño del Flujos de Datos y Procesos ETL

En la fase a este diseño del flujo de datos y procesamiento ETL, se animará la colaboración estrecha entre ingenieros de datos externos de alta cualificación en este campo y el personal directivo de la Unidad de Educación Presencial (UEP). De igual forma se seleccionará una plataforma en la nube (Azure) por el nivel de experiencia y uso, que será personalizada para adaptarse a las demandas específicas del proyecto, asegurando así una integración positiva en los datos.

Creación y Configuración de Procesos ETL

A lo largo de esta grandiosa etapa se implementarán procesos de Extracción, Transformación y Carga (ETL), con un énfasis total en la calidad y eficiencia en la manipulación de los datos. Se realizarán pruebas exhaustivas para asegurar que los procesos ETL se desarrollen adecuadamente, garantizando una extracción, transformación y cargas óptimas de los datos en un repositorio centralizado en la nube. De forma paralela, se desarrollarán tableros de control utilizando Tableau para visualizar la disponibilidad y carga de los datos de manera nítida y accesible, brindando herramientas eficaces para la toma de decisiones.

Análisis de Datos

El análisis de la información se llevará a cabo empleando métodos estadísticos avanzados, incluyendo análisis descriptivo y pruebas de hipótesis, para comparar la información previa y posterior a la incorporación de la estructura de ingeniería de datos en la nube en la Unidad de Educación Presencial (UEP). Este proceso tiene como finalidad valorar el impacto que dicha intervención ha tenido en la mejora del proceso de toma de decisiones y el rendimiento operativo dentro de la entidad.

Aspectos Éticos

Se procurará el consentimiento debidamente informado del personal administrativo implicado. Asimismo, se asegurará la privacidad de la información recolectada. Este estudio se desarrollará siguiendo los principios éticos dispuestos por la universidad y acatando todas las normativas pertinentes.

Planteamiento de la Revisión Sistemática

La revisión sistemática desempeña un papel fundamental en la investigación, ofreciendo una visión integral de la literatura existente relacionada con la implementación de arquitecturas de ingeniería de datos en la nube y los procesos de Extracción, Transformación y Carga (ETL) en entornos educativos. Su objetivo principal es explorar y analizar a fondo este campo de estudio para informar de manera precisa y detallada el desarrollo de la propuesta de arquitectura de datos para la Universidad de las Fuerzas Armadas ESPE.

Revisión Sistemática

Propósito de la Revisión

Con el fin de obtener una visión detallada sobre las corrientes actuales, estrategias y obstáculos en la adopción de arquitecturas de ingeniería de datos en la nube y procesos ETL en entidades educativas, se plantea esta revisión.

Preguntas de la Investigación

Las cuestiones fundamentales que se abordarán en esta revisión incluyen:

1. ¿Cuáles son las tendencias sobresalientes en la adopción de arquitecturas de ingeniería de datos en la nube en contextos educativos?
2. ¿Cómo se han enfrentado los procesos de ETL en instituciones educativas semejantes?

3. ¿Cuáles son los desafíos recurrentes y las mejores prácticas evidenciadas en la bibliografía sobre la administración de datos en la nube en contextos académicos?
4. ¿Existen estudios específicos que aborden la inclusión de tecnologías en la nube y procesos de ETL en instituciones de educación superior semejantes a la Universidad de las Fuerzas Armadas ESPE?

Criterios de Inclusión y Exclusión

Se definen criterios precisos para la selección de estudios, asegurando la relevancia y calidad de la información recopilada.

Criterios de Inclusión:

- Estudios publicados durante los últimos 4 años.
- Investigaciones centradas en la adopción de arquitecturas de ingeniería de datos en la nube y procesos de ETL en instituciones educativas.
- Estudios que abordan desafíos y las mejores prácticas en la administración de datos en contextos académicos.

Criterios de Exclusión:

- Estudios no relacionados con el uso de tecnologías en la nube.
- Investigaciones no enfocadas en procesos de ETL y arquitecturas de ingeniería de datos.
- Trabajos que no proporcionen información relevante para la adopción en instituciones de educación superior.

Estrategia de Búsqueda

La estrategia de búsqueda se realizará en bases de datos académicas como lo que es Google Scholar, PubMed, IEEE Xplore, Scopus y otras relevantes, utilizando palabras clave específicas. También se incluirán en la búsqueda trabajos de conferencias y revistas inclusive LinkedIn por la información que existe especializadas en el área.

Análisis de Datos

Se analizará detalladamente sobre los estudios que han obtenido incluidos, identificando patrones, desafíos comunes y mejores prácticas. La información recopilada será vital para respaldar y enriquecer la propuesta de trabajo de titulación en la Universidad de las Fuerzas Armadas ESPE.

Conclusión del enfoque de la Revisión Sistemática

Esta revisión sistemática será punto inicial para la elaboración de nuestra arquitectura porque no solo suministrará una base sólida para entender el estado actual de la investigación en el área, sino que también permitirá identificar brechas en la bibliografía que mi tesis abordará de manera innovadora. Este proceso es riguroso porque se estableces los fundamentos críticos para el diseño y ejecución de la propuesta de arquitectura de ingeniería de datos en el contexto educativo específico de la Universidad de las Fuerzas Armadas ESPE.

Población y Muestra

El estudio se enfocará en la población de la Unidad de Educación Presencial (UEP) de la Universidad de las Fuerzas Armadas ESPE, incluyendo tanto al personal docente como al administrativo. Sin embargo, debido a consideraciones logísticas y a la complejidad del proyecto, la participación directa de los profesores en la recolección de datos puede no ser posible. Por consiguiente, la muestra se centrará en el personal administrativo. Sin embargo, se

garantizará una representación equilibrada de diversas áreas y roles dentro de la Unidad de Educación Presencial (UEP) mediante una muestra estratificada cuidadosamente diseñada. Es importante resaltar que la existencia de archivos de datos preexistentes del cuerpo docente facilitará este proceso de muestreo, garantizando así la representatividad y validez de los resultados obtenidos.

Reseña y resultados de trabajos similares

1. "Data Warehousing and Data Mining in the Education System"

Este estudio de (Priyadarshini & Rath, 2020) se enfoca en la adopción de almacenes de datos y minería de datos en el ámbito educativo. Resalta la importancia de los sistemas de apoyo a decisiones para gestionar instituciones educativas de manera proactiva, especialmente en contextos donde la financiación gubernamental es limitada. La normalización de los datos a través de almacenes de datos es crucial para ofrecer una perspectiva unificada de la información en entornos educativos que son multifacéticos y competitivos.

El proyecto de almacén de datos desarrollado en el estudio resultó ser una herramienta muy útil al satisfacer las necesidades de los usuarios, permitiendo una administración más eficaz de los recursos humanos, estudiantiles e infraestructurales de la institución educativa. La adopción del almacén de datos facilitó la extracción de información relevante, como el crecimiento profesional y la tasa de rotación del personal, así como el rendimiento académico y las actividades extracurriculares de los estudiantes. Además, el análisis exhaustivo de los datos almacenados permitió a los responsables de la toma de decisiones formular estrategias y políticas para mejorar tanto la experiencia de los estudiantes como el ambiente laboral del personal. En general, se observó que el 80-85% de las decisiones se basaron en los informes generados por el sistema propuesto, lo que indica una alta tasa de adopción y uso por parte de

los usuarios, lo que contribuyó significativamente a la eficiencia y efectividad de la administración institucional (Priyadarshini & Rath, 2020).

2. "Extract Transform Load (ETL) Process in Distributed Database Academic Data Warehouse."

Este trabajo (Yulianto, 2019) se enfoca en la adopción de procesos ETL en almacenes de datos académicos con fuentes distribuidas. Para abordar la complejidad de esta tarea, se realiza un análisis detallado de las fuentes de datos y se utiliza Pentaho Data Integration v6.1. Las pruebas realizadas incluyen integridad, duplicación y metadatos para garantizar la precisión y consistencia de los datos.

Para sistemas de bases de datos distribuidas, se han llevado a cabo actividades específicas en cada base de datos, como fusiones y creación de dimensiones de conformidad. El uso de Pentaho Data Integration como herramienta de código abierto ha demostrado ser efectivo en todas estas etapas del proceso ETL, lo que sugiere que este enfoque puede ser ampliamente aplicable en otras industrias para la adopción de almacenes de datos (Yulianto, 2019)

3. "La inteligencia de negocio como apoyo a la toma de decisiones en el ámbito académico"

Se propone un sistema basado en inteligencia de negocios para administrar eficientemente los datos académicos (Dixson & Maturel, 2015). Este sistema, que utiliza datos reales de una facultad universitaria, incluye un modelo multidimensional, reportes configurables, análisis OLAP y cuadros de mando integral. Los resultados obtenidos indican mejoras significativas en la toma de decisiones y la administración de información en el ámbito académico.

Además (Dixson & Maturel, 2015) resalta la importancia del Business Intelligence (BI) en la administración empresarial, donde su adopción genera impactos significativos en la optimización de costos, toma de decisiones financieras, administración de la cadena de suministros y la relación con los clientes. La integración del BI con sistemas empresariales como Enterprise Resource Planning (ERP) es esencial para la formulación de estrategias de análisis y desarrollo empresarial. Finalmente, el BI juega un papel crucial en el Business Performance Management, facilitando la extracción y análisis de datos para la mejora continua del desempeño organizacional.

4. "Significance of data integration and ETL in business intelligence framework for higher education"

En el estudio de (Rodzi et al., 2015) nos comentan que la trascendencia de la integración de datos y del proceso ETL en el contexto de la inteligencia de negocios aplicada a la educación superior. Este análisis académico hace un enfoque detallado en cómo se efectúa esta fusión de datos provenientes de múltiples fuentes, y para ello toma como modelo de investigación a la Universiti Teknologi Malaysia (UTM).

Estado del arte

Arquitecturas de ingeniería de datos en la nube

Hemos observado un incremento significativo en la implementación de las arquitecturas de ingeniería de datos en la nube dentro de un amplio abanico de sectores. Investigaciones recientes, entre las que destaca el estudio realizado por (Guerrero & Londoño, 2016) han destacado el potencial de estas arquitecturas para manejar grandes volúmenes de datos, fortalecer la eficiencia en las operaciones y facilitar la toma de decisiones basada en datos en el marco de las tecnologías más avanzadas.

Este enfoque está respaldado por el meticuloso estudio llevado a cabo por (Guerrero & Londoño, 2016) junto a sus colaboradores con el fin de apoyar los procesos de investigación y desarrollo de aplicaciones en la nube sujetos a unos estándares de calidad en el software aceptables. Sus descubrimientos señalan la relevancia y el impacto favorable de las arquitecturas de ingeniería de datos en la nube en la sociedad contemporánea como se analiza en el trabajo de (Dixson & Maturel, 2015).

Tendencias de la arquitectura de datos

Atendiendo a las tendencias más recientes, las arquitecturas de datos en la nube están disfrutando de un notable auge tal como lo señala (Loza, 2023) este progreso se caracteriza por sus rigurosos mecanismos de seguridad en el acceso a los datos y por su mayor flexibilidad en el manejo de la información de igual forma la implementación de arquitecturas de datos en la nube se presenta como una solución idónea para afrontar los retos actuales, permitiendo una gestión de los datos más dinámica, escalable y segura. Estas corrientes evidencian la relevancia de la nube en la evolución de las arquitecturas de datos para satisfacer las exigencias del entorno tecnológico contemporáneo.

Aplicación de las tecnologías en la nube en instituciones educativas

La implementación de tecnologías en la nube en el sector educativo ha demostrado su notable impacto en la enseñanza, investigación y administración educativa (Juma & Tjahyanto, 2019).

El modelo de adopción de la computación en la nube en la educación superior presenta desafíos y beneficios particulares así mismo como diferentes investigaciones han resaltado que la tecnología en la nube favorece el aprendizaje, la enseñanza y la realización de actividades educativas clave (Alshdadi et al., 2015).

La tendencia actual apunta a un aumento en la gestión de datos educativos a través de la nube, con un enfoque en la gestión eficaz de grandes cantidades de información (Rafika Helaimia, 2023). La adaptabilidad y escalabilidad de las arquitecturas de datos basadas en la nube están cambiando la manera en que las instituciones educativas administran su información (Ghosh, 2022).

Marco Teórico

Para el diseño y desarrollo de una Infraestructura de Ingeniería de Datos en la Nube para la Universidad de las Fuerzas Armadas (ESPE) se tendrá la construcción y la evolución de esta infraestructura de ingeniería de datos en la nube destinada a la Unidad de Educación Presencial (UEP) de la Universidad de las Fuerzas Armadas (ESPE) que se fundamenta en una robusta estructura teórica y esto es fundamental para el uso óptimo y adecuado en el que incluye los siguientes elementos:

Integración de Datos

La administración eficiente de la información se posiciona como un pilar indispensable para la generación de decisiones estratégicas en el entorno académico. Como apuntan (Yahuarcani & Devia, 2022), resulta vital la implementación de sistemas aplicativos integrados que simplifiquen el proceso de Extracción, Transformación y Carga (ETL), especialmente en la recogida y tratado de datos provenientes de fuentes variadas, como los sensores de equipos inteligentes. Esta perspectiva no solo favorece la gestión eficiente de datos, sino que adicionalmente ofrece herramientas útiles para los usuarios, como la visualización y descarga de datos en formatos de fácil acceso. La adopción de metodologías ágiles como SCRUM asegura una implementación exitosa del proyecto y abre puertas a futuras investigaciones en ámbitos relacionados con la salud y la medicina.

Procesos ETL en la Nube

La transición de los procesos ETL hacia la nube se presenta a menudo como una solución óptima para las compañías que enfrentan desafíos en términos de escalabilidad, rendimiento e integración con sus infraestructuras existentes, como se resalta en el trabajo de (Gutierrez Arana & Asca Arevalo, 2023). Esta evolución hacia la nube permite a las empresas sacar provecho de la flexibilidad que ofrecen los recursos computacionales, asegurando un rendimiento superior incluso durante los periodos de mayor demanda. Adicionalmente, confiere un control más riguroso y posibilidades de personalización avanzadas en los scripts de los procesos ETL, además de garantizar acceso a una diversidad de servicios de integración y herramientas avanzadas para el análisis. Para resumir, la migración a la nube representa un conjunto de ventajas sustanciales que pueden mejorar de manera significativa la eficiencia y efectividad de los procesos de gestión de datos a nivel empresarial.

Avances de la Ingeniería de Datos en la Nube Aplicada a la Educación Superior y a la Innovación Educativa:

Profundizando en los avances de la Ingeniería de Datos en la Nube y su aplicación en la Educación Superior y la Innovación Educativa, es indiscutible el papel que desempeña la tecnología en la nube. Plataformas de renombre como AWS, Azure y Google Cloud, entre otras, han redefinido la manera en que se gestionan los datos. Estas plataformas brindan un entorno robusto y seguro para el desarrollo de arquitecturas de datos eficientes, ofreciendo, a su vez, una optimización significativa de los procesos administrativos y un enfoque que apoya en los datos para la toma de decisiones en el sector educativo (Paredes-Parada, 2021).

Siguiendo a (Paredes-Parada, 2021) la implantación de estas tecnologías en las Instituciones de Educación Superior (IES) se requiere tener un liderazgo y una visión clara al igual que un compromiso sólido con la mejora con el avance de la educación. Es importante

equipar a los docentes con métodos de enseñanza y mantener una actitud de constante renovación que fomente la generación de conocimiento en las Instituciones de Educación Superior (IES).

Como sabemos la evolución de la tecnología es impresionante y más si hablamos de la nube en cuestión de transformación la gestión de datos. Estos levantamientos dan a las instituciones educativas la oportunidad de optimizar sus procesos internos administrativos para la mejorar de toma de decisiones. Las mencionadas plataformas de tecnología en la nube proporcionan un entorno seguro y de igual forma solido para crear arquitecturas de datos eficientes, capaces de manejar grandes volúmenes de datos de manera eficaz y hecho de ofrecer la escalabilidad y capacidad de procesamiento necesarias para adaptarse a las demandas crecientes del panorama educativo actual (Paredes-Parada, 2021).

Esta propuesta de arquitectura de datos en la nube para mi va más allá de una necesidad operativa que representa una innovación en el sector de la educación superior. Tener un repositorio centralización de datos con la implementación de procesos ETL y la adopción de una infraestructura en la nube colocan a la Unidad de Educación Presencial (UEP) de la Universidad de las Fuerzas Armadas (ESPE) estar en el auge de la innovación educativa mejorando la optimización de los procesos internos de la institución, sino que además facilita el camino hacia la mejora continua y la excelencia académica. La capacidad increíble para analizar datos de forma integral abre totalmente un mundo de posibilidades para identificar patrones, tendencias y áreas de mejora que impulsa la calidad de la educación y con ello contribuye a los estudiantes tener una experiencia actual.

Alineación de Directrices Empresariales con Mejora Continua:

En la Universidad de las Fuerzas Armadas ESPE está comprometida con la alineación estratégica y mejora continua, considerándola imprescindible para el éxito. Se reconoce la

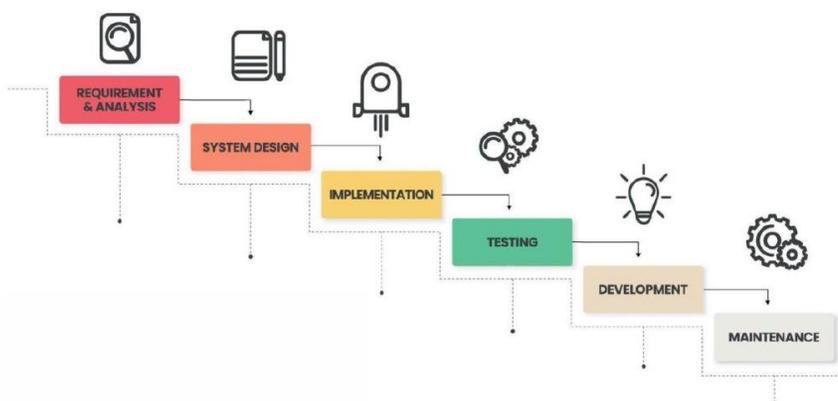
necesidad de adaptar competentemente los datos en la nube solventando las demandas actuales y futuras que su objetivo es ir asegurando que la arquitectura sea flexible y escalable. De igual forma se entiende la importancia de responder proactivamente a los cambios en el panorama educativo, lo que esto tiene como resulta un éxito a largo plazo de la institución. El adecuado uso de los objetivos institucionales ayuda a mantener la relevancia y sostenibilidad de la arquitectura propuesta.

Método de trabajo

Esta metodología para este trabajo de titulación es idónea porque se enfoca en el éxito. Uno de los enfoques más reconocidos y utilizados en el dominio del desarrollo de software como de datos es el modelo de cascada. El modelo presente también conocido como el modelo de ciclo de vida clásico, se distingue por su estructura secuencial y lineal donde cada fase del ciclo debe ser completada antes de proceder a la siguiente. El modelo de cascada que se proporciona es tener una guía clara y precisa para el desarrollo del trabajo. Facilitando la organización, la implementación y la supervisión de las actividades de manera eficaz. En este procedimiento, se siguen una serie de etapas ordenadas, que se extienden desde la recopilación de los requisitos iniciales hasta la implementación final del sistema.

Figura 1

Estrategia de trabajo: Modelo en Cascada



Nota. La ilustración presentada representa la estrategia de trabajo adoptada para este proyecto. Esta estrategia sigue la metodología del modelo de cascada, que se distingue por su enfoque lineal y estricto durante las etapas de desarrollo del software tomado de (Whale Design, 2023).

Análisis y Requerimientos

La primera etapa tiene como objetivo comenzar con una rigurosa evaluación y exactitud en la determinación de las necesidades del proyecto. El propósito es comprender de manera profunda las exigencias del proyecto e identificar sus metas fundamentales. Esta información vital que abarca desde los diccionarios de datos hasta la experiencia del usuario, establece un sólido punto de referencia para cada etapa que le sigue (Prieto, 2023).

Diseño del Sistema

En esta etapa es la que se convierte la visión conceptual en realidad por que se realiza un levantamiento meticuloso y el diseño de la arquitectura del proyecto, cimentando las bases técnicas que permitirán un flujo eficiente, procesamiento de datos, modelado y visualización, tal como indica (Prieto, 2023).

Implementación

Esta etapa es esencial observamos que se implementa la visión de las etapas anteriores a través del desarrollo de algoritmos predictivos. Los requerimientos y la arquitectura diseñada se convierten en un código funcional y eficaz. El algoritmos se basa en cómo se implementa mejor a los objetivos del proyecto (Prieto, 2023).

Pruebas

Éste es el momento se enfoca la configuración de la arquitectura como un punto crucial para la validación y mejora de la capacidad del sistema para hacer predicciones. Se

seleccionan los conjuntos de datos con sumo cuidado para asegurar que el modelo representa y generaliza adecuadamente situaciones con el mundo real (Prieto, 2023).

Desarrollo

En esta etapa, se implementa el sistema en un entorno de desarrollo. Todo el trabajo realizado hasta el momento se pone en práctica en un entorno controlado, permitiendo al o los desarrolladores y equipos de pruebas evaluar el desempeño del sistema bajo condiciones simuladas (Prieto, 2023).

Mantenimiento

Esta etapa es importante como sabemos el tener una arquitectura en la nube no nos asegura que siempre va a funcionar automáticamente es importante que un experto la maneje como por ejemplo un arquitecto de datos por que los secretos, las suscripciones, los parámetros internos de seguridad suelen cambiar y esto nos sirve para asegurar su relevancia y efectividad constantes en el tiempo de igual forma es importante implementar mecanismos de seguimiento proactivo para detectar posibles desviaciones en el rendimiento del sistema y garantizar su funcionamiento óptimo constante (Prieto, 2023).

Tableros de visualización

Los paneles o tableros de visualización como experiencia propia que se vive en la vida laboral estos adquieren una importancia cuando se saben utilizar bien los datos a veces son creados con el fin de dar una historia de seguir una secuencia, de mostrar resultados para gerencia, mejorar el análisis de información. Estas herramientas también se las conoce como "dashboards" que son esenciales para ofrecer una representación visual de datos e información relevante y así poder aportar en gran medida a las decisiones que se basan en pruebas contundentes utilizando los datos teniendo como propósito estos paneles es permitir a los usuarios revisar y comprender la información dada de manera poderoso y rápida.

Dentro de tantas soluciones disponibles, Tableau y Power Bi se destacan como una de las opciones más favoritas en la industria, gracias a su habilidad de crear ilustraciones interactivas que surgen a partir de diversas fuentes de información. Al utilizar Tableau nuestro objetivo es que las entidades educativas puedan tener una capacidad de elaborar paneles a medida para supervisar aspectos esenciales como la disponibilidad de aulas y laboratorios por semestre optando por una mejor distribución de esta.

La conexión de Tableau con otras herramientas de análisis de datos como Azure y dentro de ella bases de datos tiene un potencial de abrir una gran cantidad de oportunidades para poder trabajar amablemente analizando y dando a lugar la detección de patrones, la creación de estrategias y la evaluación que pobremente se tomen en base a los resultados de los datos. Con la meta de perfeccionar la administración educativa mediante el uso eficaz de la información existente se analiza a fondo los principios de las técnicas y las mejores relacionadas con la creación e implementación de paneles informativos usando Tableau, tal como lo indica el estudio de (Benavente Toro, 2023).

Los estudios de (Benavente Toro, 2023) también nos comentan sobre la creación de tableros de visualización de datos para el uso de registros en este caso el de los médicos se fortaleció la toma de decisiones en el ámbito de medicina, se diseñaron tableros con una arquitectura escalable, capaz de incorporar nuevos datos en el futuro llevándonos a nuestro trabajo también elaborar una mirada hacia el futuro. De esta forma se pudieron emplear tanto los datos estructurados como los no estructurados de los registros médicos y con el procesamiento previamente el texto libre para su almacenamiento en un almacén de datos.

La mayoría de estos repositorios y tableros proporcionan varios elementos visuales, como la distribución por edades de los pacientes, la serie de tiempo de las consultas médicas y los hábitos más comunes de los pacientes, entre otros. Y para concluir nos presentan una creación de perfiles para pacientes por especialidad médica y correlacionar factores con condiciones médicas específicas.

Componentes Esenciales de una Arquitectura en la Nube

El Conjunto de Recursos

El Conjunto de Recursos como un contenedor lógico, que nos es útil para almacenar y unir todos los recursos que nos dispone en el portal Azure. Proporciona un método efectivo para administrar, controlar el acceso y organizar los recursos de manera coherente.

La Fábrica de Datos

La Fábrica de Datos de Azure se nos presenta como una herramienta ETL que está a mi experiencia completamente administrada que nos ayuda con simplifica de la creación, programación y orquestación de flujos de trabajo de datos efectivos. Podemos observar que juega un papel significativo en el procesamiento de los datos y la integración de sistemas en la nube, como destacan (Proaño et al., 2018).

La Máquina Virtual

Las Máquinas Virtuales nos brindan una infraestructura más estable y seguro al igual que una flexibilidad para poder ser escalable y ejecutar aplicaciones y servicios. Reflejan ser más esenciales para albergar aplicaciones específicas que requieran un entorno de sistema operativo Windows e información de gran valor.

La Base de Datos SQL y Servidor SQL

En Azure estos dos van de la mano por que para tener una base de datos necesitamos el servidor SQL que se establecen como elementos fundamentales para el almacenamiento y la gestión de datos estructurados proporcionado una solución de base de datos segura en cada uno de los ip donde estará el manejo de datos y escalable para aplicaciones empresariales.

La Lógica de Aplicación

La Lógica de Aplicación se utiliza para automatizar procesos empresariales y conectar servicios en la nube y locales. Es vital para orquestar flujos de trabajo y automatizar tareas comunes.

La Cuenta de Almacenamiento

Una Cuenta de Almacenamiento proporciona un mejor espacio organizado donde se pueden almacenar datos no estructurados y también los datos estructurados de una manera tan eficiente y dependiendo el uso su económica esta parte varia porque existe también almacenamientos de millones de millones porque tenerlos almacenados y seguros son costosos. Se convierte en un recurso indispensable para almacenar y gestionar grandes volúmenes de datos en la nube.

El Depósito de Claves

El Depósito de Claves como importante en los requisitos propuestos sirven demasiado porque sin ellos el uso de algunas herramientas dejan de ser funcionales y se emplea para almacenar y manejar de manera segura secretos, claves y certificados. Es crucial para la gestión de identidades, la autenticación y la seguridad de los datos en la nube.

El Servicio Azure Databricks

El Servicio Azure Databricks presenta una plataforma de análisis y procesamiento de datos unificada y plenamente gestionada. Este se vuelve esencial para el análisis de datos en tiempo real, ciencia de datos y aprendizaje automático en la nube.

El Servicio Azure Databricks presenta una plataforma de análisis y procesamiento de datos unificada y plenamente gestionada. Este se vuelve esencial para el análisis de datos en tiempo real, ciencia de datos y aprendizaje automático en la nube.

Capítulo III

Análisis de la Estructura de Datos en el Ámbito Académico

El ámbito académico se caracteriza por la necesidad de analizar y comprender los datos para optimizar los procesos educativos. En la institución en cuestión, la Unidad de Educación Presencial (UEP) ha emergido como un recurso crucial de información valiosa. Este amplio depósito de datos no solo documenta la actividad académica diaria, sino que también nos ofrece una perspectiva sin igual sobre la eficacia y rendimiento de nuestros programas educativos.

El tener una estructura de datos de la Unidad de Educación Presencial (UEP) es introducirse en un océano de posibilidades para la investigación. Cada uno de estos elementos desde la identificación del personal docente hasta los detalles de la programación de las aulas, representa una oportunidad sin precedentes para conocer y entender los procesos internos de nuestra institución de una mejor manera.

Con este estudio de la intención de explorar la profundidad, la complejidad y la riqueza de la estructura de datos de la Unidad de Educación Presencial (UEP) nos adentramos a un análisis meticuloso y riguroso que busca descubrir patrones, identificar tendencias y en última instancia lograr un conocimiento significativo que impulse la mejora continua en el ámbito educativo.

Requerimientos de Negocio para la Unidad de Educación Presencial (UEP)

En nuestro trabajo de titulación es importante nombrar todos los requerimientos por que será como nuestro objetivo a cumplir detallando que el negocio necesita una visión precisa, clara de las necesidades y expectativas de la Unidad de Educación Presencial (UEP) en lo que respecta en el uso de datos. Estos requisitos nos permitirán tener a nuestros como una guía fundamental durante el proceso de diseño e implementación de la arquitectura de ingeniería de

datos en la nube para poder garantizar que la solución propuesta cumpla de manera eficaz y eficiente con las demandas operativas y estratégicas de la Unidad de Educación Presencial (UEP).

Integración de Datos: La Unidad de Educación Presencial (UEP) requiere una solución que logre integrar varios datos sin importar la cantidad incluyendo información administrativa, académica para que sean almacenados en un repositorio central. Siendo esta una integración en el cual debe cumplir con asegurar la relación, eficiencia y fiabilidad de los datos.

Acceso a la Información: El acceso a la información como la mayoría de los datos esto se debe manejar con cautela y con un estricto acceso tanto para la modificación de nuevos archivos como los almacenamientos y finalmente proporciona un acceso rápido y sencillo a la información relevante para la Unidad de Educación Presencial (UEP).

Seguridad de los Datos: La seguridad de los datos es prioridad para la Unidad de Educación Presencial (UEP) y también va de la mano con el tema atención con respecto al acceso de la información en el que se busca tener una propuesta donde se deberá contar con robustos mecanismos de seguridad para salvaguardar la confidencialidad, integridad y disponibilidad de la información.

Análisis Avanzados: La Unidad de Educación Presencial (UEP) espera que la solución permita realizar análisis avanzados como machine learning en la implementación de tableros con los datos almacenados incluyendo el uso de modelos predictivos como regresiones lineales y técnicas de aprendizaje automático para generar conocimientos significativos.

Automatización de Procesos: Un reto es la automatización porque lo que quieren es economizar tiempo en el análisis de datos y la solución de esta debe tener la capacidad de automatizar procesos esenciales, como la extracción, transformación y carga (ETC) de datos para mejorar la eficiencia operativa y disminuir la carga de trabajo manual.

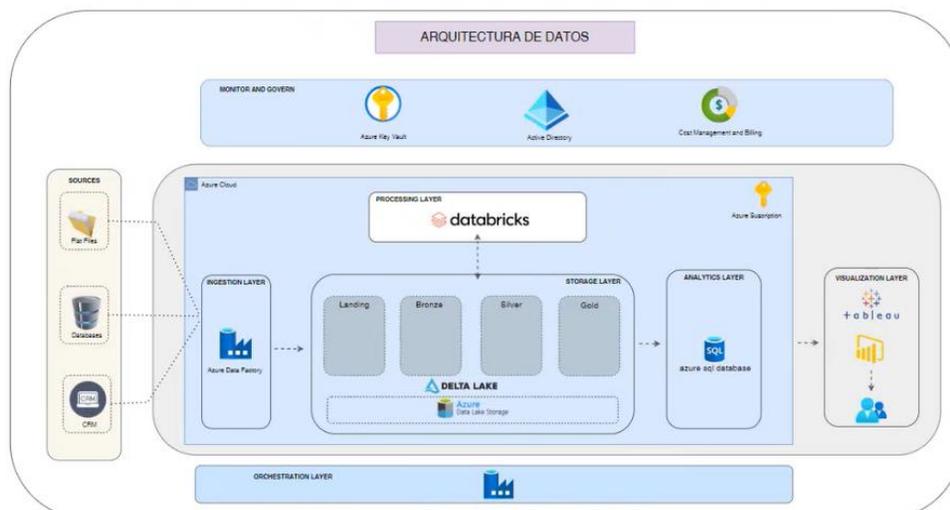
Escalabilidad y Flexibilidad: Ser escalable en la actualidad es importante como sabemos la tecnología va mejorando y se estima en un futuro nuevas implementaciones de mejora en la Unidad de Educación Presencial (UEP) y unos de los posibles objetivos es que sea capaz de manejar volúmenes grandes de datos sin comprometer el rendimiento.

Infraestructura Cloud

En la etapa de diseño de la Infraestructura Cloud, se vuelve importante en este trabajo de titulación porque es el corazón en el cual se exploraron meticulosamente las demandas particulares de los proyectos similares. Cumpliendo con el objetivo de identificar los servicios de Microsoft Azure que mejor se alinearan con las necesidades de implementación previstas, llevando a cabo una evaluación detallada de cada uno de los elementos y profunda de las funciones esenciales y adicionales requeridas para el proyecto. Esta evaluación incluyó factores vitales como la escalabilidad, la seguridad y la integración con otras herramientas ya existentes en el sistema.

Basándose en estos requisitos y en la experiencia propia se escogieron varios servicios de Azure que se alineaban de manera óptima con los objetivos particulares del proyecto. Entre estos servicios se encontraban Factoría de Datos y Azure Databricks que están siendo usadas en diferentes empresas bancarias seleccionadas por su capacidad para satisfacer los criterios establecidos para el procesamiento, transformación y carga de datos, así como por el entorno de desarrollo.

La planificación arquitectónica también se analizó consideraciones sobre la distribución y el despliegue de los recursos en la nube, diseñando una estructura que permitiera aprovechar al máximo las capacidades de escalabilidad de Azure, asegurando un rendimiento eficiente incluso frente a posibles aumentos en la carga de trabajo. Detallando a continuación las características de esta arquitectura propuesta.

Figura 2*Diagrama de Infraestructura Cloud*

Nota. Este diagrama engloba toda una arquitectura en la nube implementada en el proyecto, mostrando la disposición y la conexión de los diferentes recursos utilizados para soportar el servicio en la nube. Fuente: Autoría Propia.

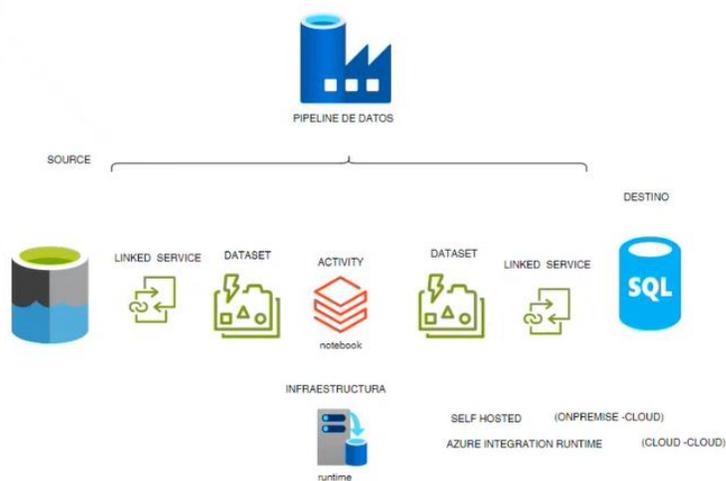
Factoría de Datos: Aquí nos centramos porque sirve como el núcleo central para la creación de flujos de trabajo especializados en el procesamiento y la ingestión de datos. Dentro de este entorno se logra establecer dos directorios principales en el área de Ingestas, cada uno con roles específicos. Uno de ellos es el Directorio "DATA", que alberga un pipeline especializado en transformar datos desde su estado original hasta una forma refinada y preparada para su análisis. Este proceso incluye actividades como la limpieza de datos, la estandarización de formatos y la aplicación de reglas de negocio para asegurar la calidad de la información.

En cuanto al funcionamiento Azure Data Factory (ADF), se basa en la construcción de pipelines de datos, que funcionan como tuberías conectadas. Para esto, se requiere una fuente de datos, como un Data Lake, y un destino, como una base de datos SQL. Todo el ecosistema

del Data Factory depende de servicios vinculados “*Linked Services*”, que actúan como puntos de conexión tanto en el origen como en el destino. Estos servicios se configuran a través de dataset, que representan los metadatos de los datos entrantes y salientes. Una vez configurados estos elementos, podemos utilizar nuestras actividades, aprovechando la integración con Azure Databricks, para llevar a cabo las operaciones necesarias. Además, para la infraestructura de ejecución, se requieren recursos como “*Self-Hosted*” para la conexión de local a la nube, y Azure “*Integration Runtime*” para operar dentro de la misma nube. Para cargar archivos Excel, se necesita el Self-Hosted, mientras que para el procesamiento de datos se requiere la comunicación de nube a nube.

Figura 3

Detalle de la Factoría de Datos



Nota. Esta figura proporciona una visión detallada de las funciones ejecutadas por la Data Factory en el proyecto, destacando cada paso del proceso de manejo de datos, desde su adquisición hasta su entrega final. Fuente: Autoría Propia.

Azure Databricks: Se estableció como el núcleo central para el procesamiento de datos y modelos del proyecto, con una configuración personalizada para garantizar un rendimiento eficiente y una gestión efectiva de los recursos. Se implementó un único clúster central con las siguientes características clave:

Tipo de Clúster: Se configuró un solo clúster central para simplificar la administración y optimizar el uso de recursos.

Figura 4

Clúster

The screenshot shows the Azure Databricks configuration page for a cluster named 'Cluster_espe'. The interface is in 'Preview' mode. The configuration is set to 'Single node' with 'No isolation shared' access mode. The performance settings include 'Databricks Runtime Version 10.4 LTS (includes Apache Spark 3.2.1, Scala 2.12)' and 'Standard_DS3_v2' node type with '14 GB Memory, 4 Cores'. The cluster is configured to 'Terminate after 16 minutes of inactivity'. The summary box indicates 1 Driver, 14 GB Memory, 4 Cores, and a cost of 0.75 DBU/h for Standard_DS3_v2 nodes.

Nota. Esta figura presenta una instantánea visual del clúster utilizado en el proyecto, proporcionando información sobre su configuración y rendimiento en un momento específico.

Fuente: Autoría Propia.

Componentes Clave de la Arquitectura de Ingeniería de Datos en la Nube

Los componentes clave que conforman la arquitectura de ingeniería de datos en la nube diseñada para la Unidad de Educación Presencial (Unidad de Educación Presencial (UEP)) de la Universidad de las Fuerzas Armadas ESPE. Cada componente desempeña un papel

fundamental en la ingestión, transformación, almacenamiento y análisis de datos, contribuyendo así a la mejora de la gestión y toma de decisiones en el ámbito académico. Desde el Grupo de Recursos que proporciona el entorno de infraestructura hasta el Servicio Azure Databricks que permite el análisis avanzado de datos, cada elemento se integra para crear una solución integral que cumple con los requisitos específicos de la Unidad de Educación Presencial (UEP).

El grupo de recursos en Azure es un contenedor lógico que organiza y administra todos los recursos relacionados en una solución en la nube. Proporciona coherencia y control de acceso, permitiendo a los usuarios administrar y supervisar eficazmente todos los recursos de Azure utilizados en un proyecto específico.

Resource group

El Grupo de Recursos desempeña un papel fundamental al servir como un contenedor organizativo para todos los recursos asociados con el proyecto en la plataforma de Azure. Su contribución principal reside en proporcionar coherencia y control de acceso, lo que facilita una gestión eficiente de los recursos en la nube. Es comparable a una maleta donde se almacenan todos los componentes necesarios para construir una arquitectura estandarizada y profesional. Al mantener todos los recursos del proyecto organizados dentro de un único Grupo de Recursos, se promueve la consistencia y se simplifica la administración, lo que resulta en una implementación más fluida y una gestión más efectiva del entorno en la nube.

Figura 5

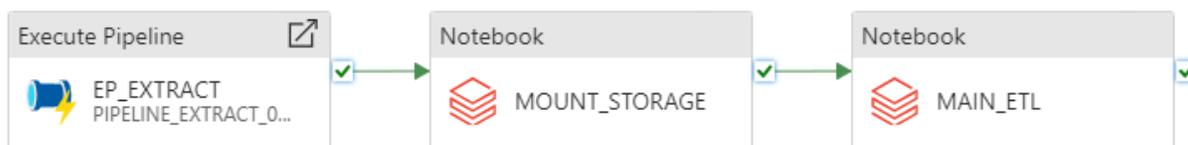
Resource group

<input type="checkbox"/>	 adf-etl-espe-001	Data factory (V2)
<input type="checkbox"/>	 azurevm	API Connection
<input type="checkbox"/>	 db-espe-001 (server-espe-001/db-espe-001)	SQL database
<input type="checkbox"/>	 dbw-etl-001	Azure Databricks Service
<input type="checkbox"/>	 dlsespe001	Storage account
<input type="checkbox"/>	 kv-espe-001	Key vault
<input type="checkbox"/>	 logic-espe-001	Logic app
<input type="checkbox"/>	 server-espe-001	SQL server

Nota. Visualizamos una muestra de nuestros recursos en uso

Data factory (V2)

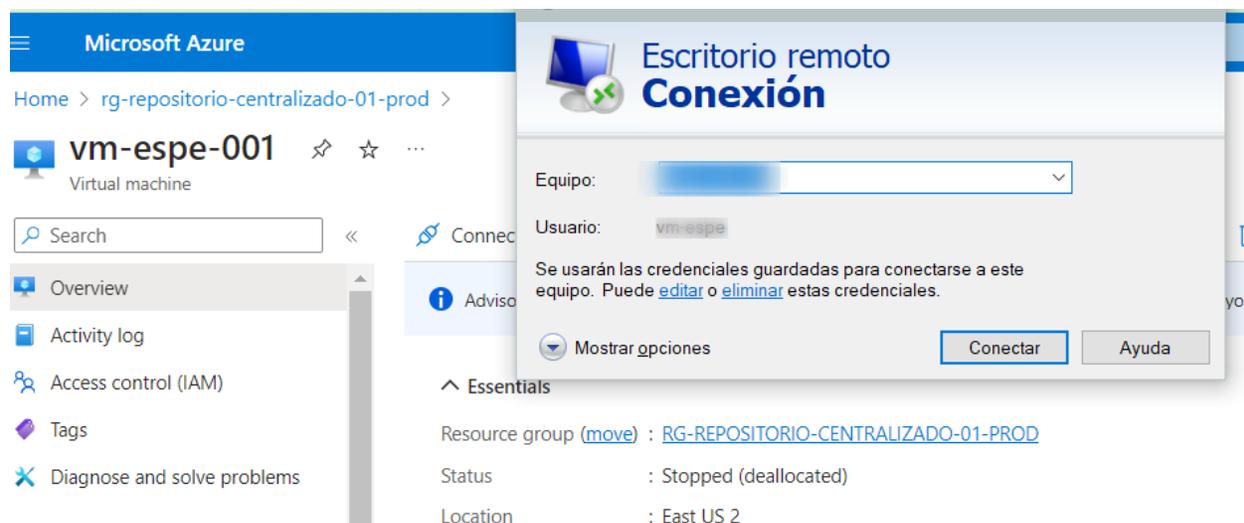
Como lo hemos mencionado es un componente esencial para esta arquitectura al poseer una herramienta integra y completamente administrada para la creación, programación y orquestación de flujos de trabajo de datos de manera eficiente. Teniendo como contribución principal la automatización de los procesos de Extracción, Transformación y Carga (ETL) lo que nos asegura la composición fluida de datos provenientes de diversas de varios archivos en un repositorio centralizado. De igual forma esto funciona como motor de nuestra automatización o conocida como orquestación de datos del proyecto, Azure Data Factory permite configurar y ejecutar flujos de trabajo complejos, definiendo las transformaciones necesarias para garantizar la coherencia y la calidad de los datos. Al mismo tiempo la capacidad de programar las ejecuciones de estos flujos de trabajos en momentos específicos con el uso de triggers que vienen siendo como lanzamiento de actividades ejemplo a las 9 am lanza el triggers o acción de ejecución del pipeline de datos asegurando que los procesos se ejecuten de manera oportuna y eficaz, lo que contribuye a la sincronización y actualización regular de la información en el repositorio centralizado.

Figura 6*Data factory (V2)*

Nota. Demostración del pipeline en producción

Virtual machine

Desempeña un papel esencial en la arquitectura al proporcionar un entorno flexible y escalable en la nube para ejecutar aplicaciones y servicios críticos del proyecto. Su contribución se centra en garantizar la eficiencia y el rendimiento óptimo de las operaciones, al tiempo que ofrece un espacio seguro para el procesamiento de datos. Aquí es donde se llevan a cabo tareas clave, como el análisis, modelado y visualización de datos, antes de su integración en el repositorio centralizado. La configuración de la máquina virtual se enfoca en la seguridad de los datos, implementando medidas de protección y control de acceso para salvaguardar la información sensible. Además, su capacidad de escalabilidad permite adaptarse fácilmente a las demandas cambiantes del proyecto. La Máquina Virtual proporciona un entorno de trabajo seguro y flexible que impulsa la eficiencia y efectividad del proyecto en su conjunto.

Figura 7*Virtual machine*

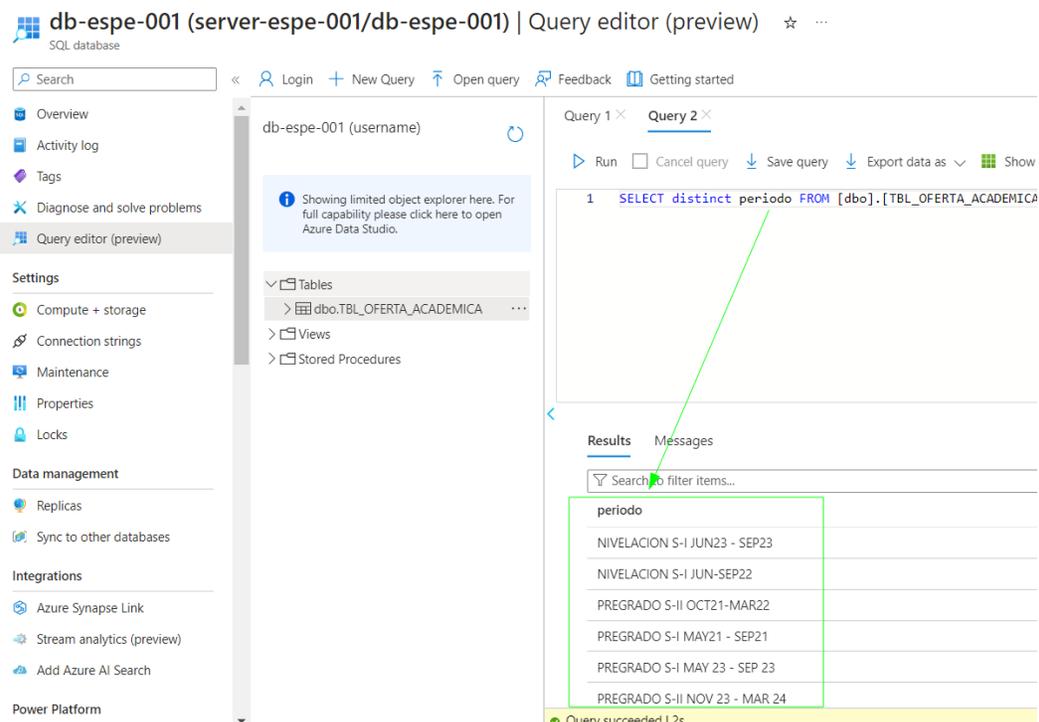
Nota. Máquina virtual y método de conexión a la maquina

SQL database

Proporciona un entorno robusto y seguro para almacenar y gestionar los datos estructurados de la Unidad de Educación Presencial (UEP). Su contribución principal radica en garantizar la integridad, disponibilidad y seguridad de la información académica y administrativa, proporcionando así una base sólida para la toma de decisiones informadas. Además de servir como repositorio centralizado de datos, la Base de Datos SQL también cumple un papel importante en la conectividad con otras herramientas y plataformas, como Tableau, facilitando así el análisis y la visualización de los datos almacenados.

Figura 8

SQL database



Nota. Bases de datos demostrando mediante una sentencia de SQL los resultados de almacenamiento y los tipos de periodo que se obtienen.

Logic App

Su capacidad para integrar una lógica en el servicio y automatizar flujos de trabajo ha permitido optimizar diversas tareas fundamentalmente el apagado automático de la máquina virtual. Esta es una automatización que no solo contribuye a la eficiencia energética y la reducción de costos, sino que también asegura una gestión más eficientemente a los recursos en la nube sido un fragmento de gran importancia para la mejora de la operatividad del proyecto, permitiendo una gestión más eficiente y económica de los recursos tecnológicos disponibles y uso adecuado de cada servicio propuesto.

Figura 9*Logic App*

The image shows two panels from an Azure Logic App configuration. The top panel, titled 'Recurrence', is used to define how often the action runs. It includes the following settings:

- Interval:** 1
- Frequency:** Week
- Time zone:** (UTC-05:00) Bogota, Lima, Quito, Rio Branco
- On these days:** Monday, Tuesday, Wednesday, Thursday, Friday
- At these hours:** 17
- At these minutes:** 30

A preview box below these settings states: 'Runs at 17:30 on Monday, Tuesday, Wednesday, Thursday, Friday every week'. Below the preview is a button labeled 'Add new parameter'.

An arrow points down to the second panel, titled 'Start virtual machine'. This panel contains the following configuration:

- Subscription Id:** Azure subscription 1
- Resource Group:** rg-repositorio-centralizado-01-prod
- Virtual Machine:** vm-espe-001

At the bottom of this panel, it says 'Connected to carodriguez19@espe.edu.ec. [Change connection.](#)'

Nota. Visualizamos la recurrencia y el manejo de que se enciende la máquina virtual

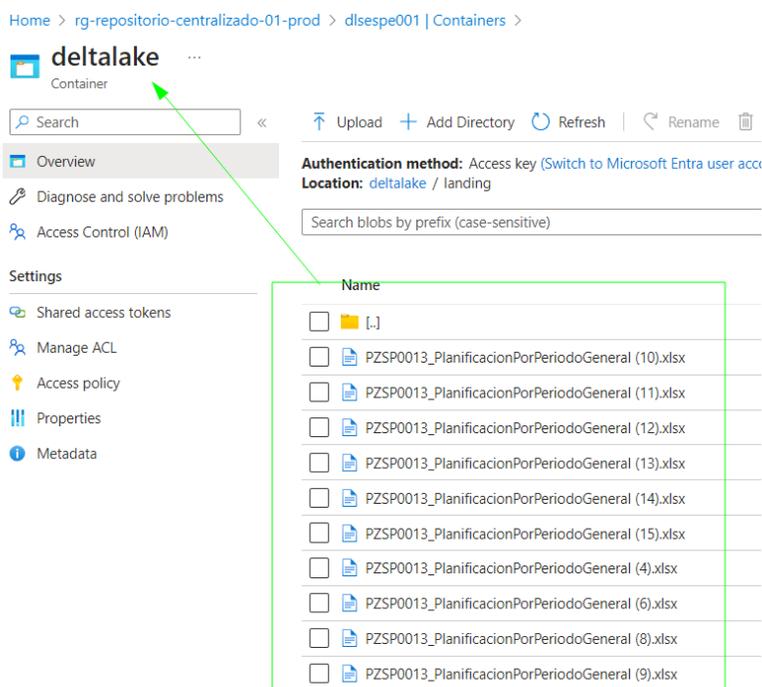
Storage account

Este recurso se al instalar es importante configúralo para que nos sirva como data lake que nos ayudara para proporcionar un almacenamiento eficiente y económico de datos no estructurados chocando la necesidad de manejar grandes volúmenes de información de manera segura y accesible. La implementación de esta cuenta garantiza un repositorio centralizado y controlado para los datos que nos facilita su procesamiento y análisis posterior. Este enfoque estratégico no solo asegura la integridad y disponibilidad de los datos también

sienta las bases para una toma de decisiones informada y ágil basado en información precisa y confiable.

Figura 10

Storage account



Nota. Muestra el contenedor que está dentro desde nuestro storage account

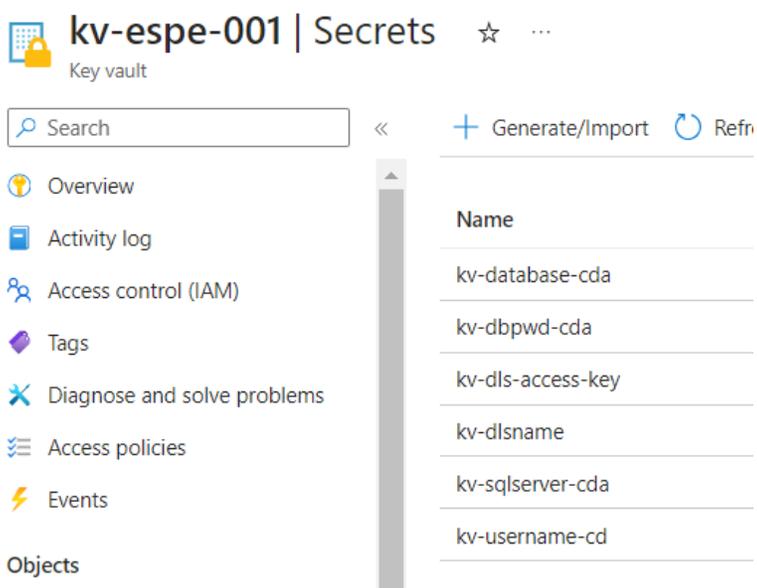
Key vault

Representa un segmento donde se almacenarán nuestros secretos que nos sirven para comunicar de un servicio a otro teniendo mayor seguridad que se ha implementado para salvaguardar la información confidencial y las credenciales de acceso. Su diseño se enfoca en la gestión segura de secretos, claves que garantizan así la protección de identidades y datos en el extenso mar de información digital que es la nube. La valiosa contribución de este Almacén de Claves ha sido crucial en la consolidación de la integridad y confidencialidad de la

información que se resguarda. Ello a su vez permite una manipulación de datos segura, mediante herramientas como Databricks y la conexión con diversas bases de datos.

Figura 11

Key vault



Nota. Demostración de los secretos creados

Azure Databricks Service

Una plataforma robusta y administrada para el análisis avanzado y el procesamiento de datos en la nube. Su contribución fue esencial para llevar a cabo tareas críticas como la limpieza y transformación de datos utilizando herramientas como PySpark y versiones de Python. Esta plataforma permitió realizar análisis detallados y aplicar técnicas de machine learning para obtener insights valiosos que respaldaron la toma de decisiones informada. Además, facilitó la integración con otras herramientas y servicios, permitiendo el flujo eficiente de datos desde su procesamiento hasta su almacenamiento en SQL para su posterior análisis y uso.

Figura 12

Azure Databricks Service

```

# Crear una lista para almacenar los DataFrames
dfs = []

# Leer cada archivo Excel y cargarlo en un DataFrame
for excel_file in excel_files:

    dffile = spark.read.format("com.creatytics.spark.excel") \
        .option("header", "true") \
        .option("inferSchema", "true") \
        .load(excel_file)

    datoFechaCierre = dffile\
        .take(3)[1][0].split("-")[2]\
        .replace("SEPT", "SEP")\
        .replace(" ", "")

    datoPeriodo = dffile\
        .take(3)[1][0].split(":")[1]

    df = spark.read.format("com.creatytics.spark.excel") \
        .option("header", "true") \
        .option("inferSchema", "true") \
        .option("dataAddress", "Sheet0!A5") \
        .load(excel_file)\
        .filter("DOCENTE IS NOT NULL")\
        .select(F.col("*"), F.to_date(F.lit(datoFechaCierre), "MMWyy").alias("FECHA_CIERRE"), F.lit(datoPeriodo).alias("PERIODO"))

```

Nota. Visualizamos un fragmento esencial del código donde hacemos una parte del ETL

Requisitos de instalación en Azure y versiones

Tabla 1

Requisitos de instalación en Azure

Componente	Prioridad	Ubicación	Versión/Configuración
Grupo de recursos	Alta	East US 2	Estándar
Azure Data Factory	Alta	East US 2	Data Factory (V2)

Componente	Prioridad	Ubicación	Versión/Configuración
Máquina Virtual	Alta	East US 2	Windows, Tamaño: Estándar DS1 v2 (1 vcpu, 3.5 GiB de memoria)
Base de Datos SQL	Alta	East US 2	Nivel de Precios: General Purpose - Serverless: Gen5, 1 vCore
Servidor SQL	Alta	East US 2	Nombre del Servidor: server-espe- 001.database.windows.net
Logic App	Alta	East US 2	Estándar
Cuenta de Almacenamiento	Alta	East US 2	Rendimiento: Estándar, Replicación: Almacenamiento redundante local (LRS), Tipo de Cuenta: StorageV2 (propósito general v2), Espacio de nombres Jerárquico: Habilitado
Almacén de Claves	Alta	East US 2	SKU (Nivel de Precios): Estándar
Servicio Azure Databricks	Alta	East US 2	Clúster de Cómputo, Versión de Tiempo de Ejecución Databricks 10.4 LTS (incluye Apache Spark 3.2.1, Scala 2.12), Tipo de Nodo: Standard_DS3_v2, 1 Controlador, 14 GB de Memoria, 4 Núcleos

Nota. Lista de componentes necesarios para la instalación y configuración del entorno en Azure, incluyendo servicios y recursos específicos requeridos para el funcionamiento del sistema. Fuente: Autoría propia

Requisitos de Uso Funcional

Tabla 2

Requisitos de Uso Funcional

Requisitos	Descripción
Conexión a Internet	Acceso a Internet para utilizar los servicios y recursos en la nube de Azure.
Cuenta de Azure activa	Suscripción activa en Azure con el nivel de suscripción estándar para acceder y utilizar los servicios y recursos necesarios.
Tableau Desktop	Software de visualización de datos para crear informes interactivos y sirven para tener un mejor control en los paneles.
Credenciales de la ESPE	Acceso a las credenciales de la Universidad de las Fuerzas Armadas ESPE para autenticación y autorización en los servicios y plataformas utilizados.

Nota. Conjunto de requisitos necesarios para utilizar funcionalidades específicas del sistema, incluyendo la necesidad de contar con una cuenta de Azure activa y suscripción estándar, así como la disponibilidad de Tableau con las credenciales de la ESPE.

Desarrollo Arquitectura para el Flujo de Información

Se implemento una Arquitectura de Datos por capas. A continuación, se proporciona una descripción más detallada de cada una de las capas y zonas utilizadas:

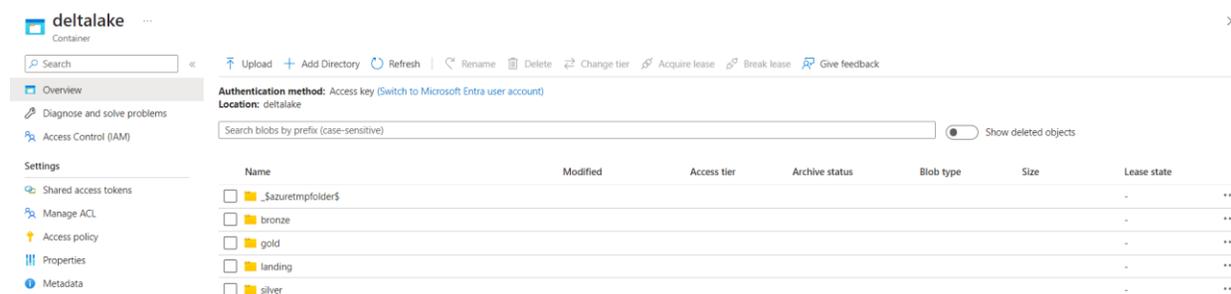
En la fase de almacenamiento físico, se gestionó la ubicación de los datos en la máquina virtual, conectando a un Delta Lake para asegurar la coherencia y fiabilidad de la información. La selección de Delta Lake ofrece beneficios significativos, incluida su capacidad para manejar grandes volúmenes de datos y llevar a cabo operaciones atómicas para mantener la integridad de los datos y mejorar la eficiencia en las operaciones.

En la capa lógica de Databricks, se crea una estructura organizada y abstracta de los datos mediante un proceso ETL. Este sistema cataloga los datos en tablas, establece relaciones y aplica restricciones, simplificando la consulta y el análisis de estos. Esta abstracción actúa como un puente entre la complejidad del almacenamiento físico y la simplicidad necesaria para el análisis, mejorando así la comprensión y usabilidad de los datos.

Se establecieron cuatro bases de datos que hemos designado como nuestros pilares fundamentales:

Figura 13

Estructura de Carpetas en Data Lake Storage



The screenshot shows the Azure Data Lake Storage interface for a container named 'deltalake'. The interface includes a search bar, navigation options (Upload, Add Directory, Refresh, Rename, Delete, Change tier, Acquire lease, Break lease, Give feedback), and a table of folders. The table has columns for Name, Modified, Access tier, Archive status, Blob type, Size, and Lease state. The folders listed are .azurempfolder\$, bronze, gold, landing, and silver.

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
<input type="checkbox"/> .azurempfolder\$						- ...
<input type="checkbox"/> bronze						- ...
<input type="checkbox"/> gold						- ...
<input type="checkbox"/> landing						- ...
<input type="checkbox"/> silver						- ...

Nota. Esta figura ilustra la organización de carpetas en el Data Lake Storage, reflejando la implementación de una estrategia de almacenamiento por niveles (landing, bronze, silver, gold) para la gestión eficiente de datos en el proyecto. Fuente: Autoría Propia.

- Landing

Esta área se reserva para preservar la información en su estado original y sin alteraciones. Sirve como el punto inicial de entrada de datos al sistema, donde se reciben archivos y se almacenan sus metadatos para su procesamiento posterior. Estos metadatos provienen de archivos provenientes del servidor de nuestra máquina virtual, representando así la primera etapa en el flujo de datos hacia nuestro sistema.

- Bronze

Diseñado para alojar datos que han sido refinados y están listos para su posterior manipulación. Esta fase proporciona la flexibilidad necesaria para la ingestión de datos crudos antes de someterlos a procesos de transformación y mejora de calidad.

- Silver

En este punto, los datos han sido procesados y están listos para ser presentados en nuestros paneles de visualización. Esta área se dedica al tratamiento y estandarización de los datos, donde se consolidan y aplican reglas de calidad para garantizar su uniformidad y fiabilidad. Los datos se preparan y transforman según estándares y reglas predefinidos, asegurando así su disponibilidad para su uso en modelos y proporcionando una base de datos completa y de alta calidad.

- Gold

Este espacio ha sido creado para permitir una posible expansión y una mejora adicional en el refinamiento de datos. En muchos casos, estos espacios están dedicados a llevar a cabo

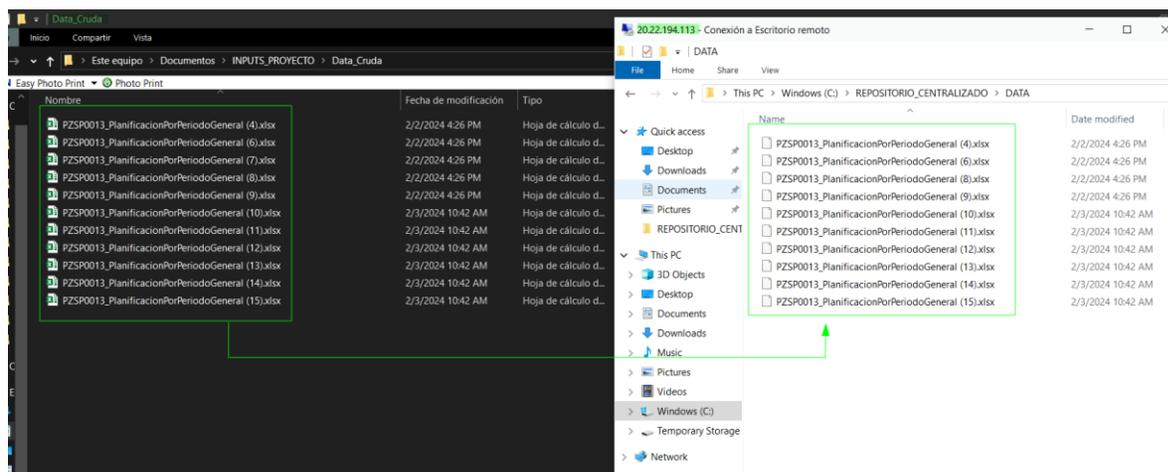
un análisis avanzado de los datos, permitiendo explorar nuevas perspectivas y descubrir insights más profundos.

Funcionamiento y aplicación práctica del proceso de implementación

Para aprovechar al máximo nuestra guía paso a paso, es crucial comprender su forma de uso. Simplemente, accedemos a nuestra máquina virtual y procedemos a copiar nuestros archivos.

Figura 14

Acceso y Copia de Archivos en la Máquina Virtual



Nota. Esta sección detalla la manera en que se utiliza la guía paso a paso proporcionada, indicando cómo acceder a la máquina virtual y copiar los archivos necesarios para llevar a cabo el proceso de implementación. Fuente: Autoría Propia.

Análisis de Costos y Gestión Eficiente de Recursos en la Adopción de Servicios en la Nube

El análisis minucioso de los costos vinculados a la implementación de servicios en la nube constituye un elemento clave para una gestión de recursos empresariales eficaz. En este estudio, se realizó una evaluación detallada de los costos asociados con la adopción de servicios en la nube a través de la plataforma Azure de Microsoft. El enfoque se centró en el

examen de una factura específica, que reflejaba los costos acumulados durante un período de nueve días de operación continua.

Los datos recopilados revelaron un patrón ascendente en los costos a lo largo del período analizado. Inicialmente, el costo fue de \$15.83 USD y aumentó gradualmente hasta alcanzar los \$43.08 USD al final del estudio. Este aumento progresivo se atribuye al uso continuo de los servicios en la nube, lo que resultó en un consumo progresivo de recursos y, en consecuencia, en un incremento de los costos asociados.

Nuestro análisis revela que el costo diario promedio es de aproximadamente \$4.79 USD durante un período de nueve días. Esta cifra provee importante luz acerca del costo medio diario de los servicios de Azure. Es innegable la importancia de esta información, especialmente en términos de planificación y manejo eficaz de recursos en el actual contexto empresarial contemporáneo.

Tabla 3

Costos de Implementación en Azure

Día	Costo acumulado (USD)
Día 1	\$15.83
Día 2	\$17.09
Día 3	\$19.39
Día 4	\$23.62
Día 5	\$26.50
Día 6	\$28.67

Día 7	\$33.06
Día 8	\$39.81
Día 9	\$43.08

Nota. Tabla que presenta los costos asociados con la implementación de la solución en la nube de Azure diario.

Costo diario promedio = Costo total / Número de días

$$= \$43.08 / 9$$

$$\approx \mathbf{\$4.79 \text{ USD/día}}$$

Análisis del Costo de una Máquina Virtual para la Investigación

La evaluación del costo asociado al mantenimiento de una máquina virtual durante un periodo específico es esencial para la gestión eficiente de recursos en proyectos de investigación. Utilizando el precio por hora proporcionado por el proveedor de servicios en la nube, se ha llevado a cabo un análisis del costo de tener una máquina virtual activa durante los días laborables, de lunes a viernes, en el intervalo horario de 8am a 2pm.

Según el cálculo realizado, el costo por hora de la máquina virtual es de 0.1170 USD. Durante los días laborables mencionados, la máquina virtual se encuentra activa durante un total de 6 horas cada día, sumando así 30 horas semanales y 120 horas al mes. Por lo tanto, el costo mensual de mantener la máquina virtual funcionando durante este periodo específico se estima en aproximadamente 14.04 USD.

Figura 15

Costos máquina virtual

[Home](#) > [Virtual machine](#) >

Create a virtual machine ...

✔ Validation passed

Basics Disks Networking Management Monitoring Advanced Tags [Review + create](#)

ⓘ Cost given below is an estimate and not the final price. Please use [Pricing calculator](#) for all your pricing needs.

Price

1 X Standard DS1 v2
by Microsoft
[Terms of use](#) | [Privacy policy](#)

Subscription credits apply ⓘ

0.1170 USD/hr
[Pricing for other VM sizes](#)

TERMS

Nota. Pago proporcionado de la máquina virtual Fuente: Autoría Propia.

Capítulo IV

Resultados

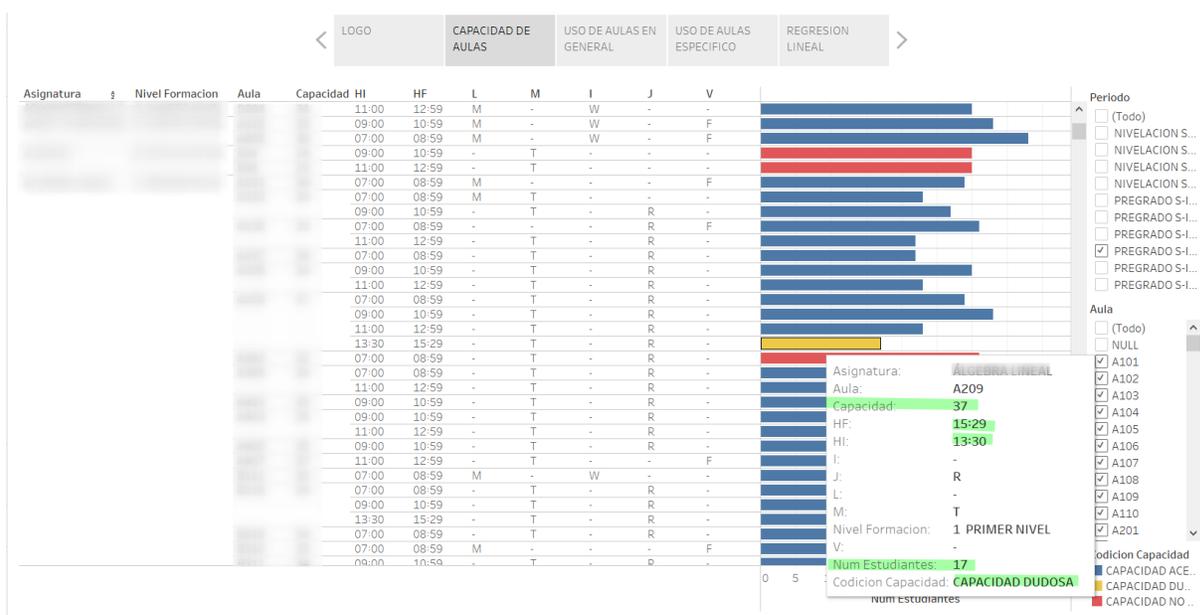
En este capítulo se presentan los resultados originarios de la implementación de la arquitectura de ingeniería de datos en la nube para la Unidad de Educación Presencial (UEP) de la Universidad de las Fuerzas Armadas ESPE. Se destacan dos componentes clave: la visualización de datos mediante Tableau y el cumplimiento de los objetivos establecidos en el diseño e implementación de esta arquitectura.

Resultados de Tableau

La visualización de datos a través de Tableau ha proporcionado una representación clara y dinámica de la información académica y administrativa en la Unidad de Educación Presencial (UEP).

Figura 16

Análisis de las Capacidades de las Aulas



Nota. Esta figura representa la capacidad de aulas mediante indicadores sobre las capacidades del aula y el número de estudiantes. Fuente: Autoría Propia.

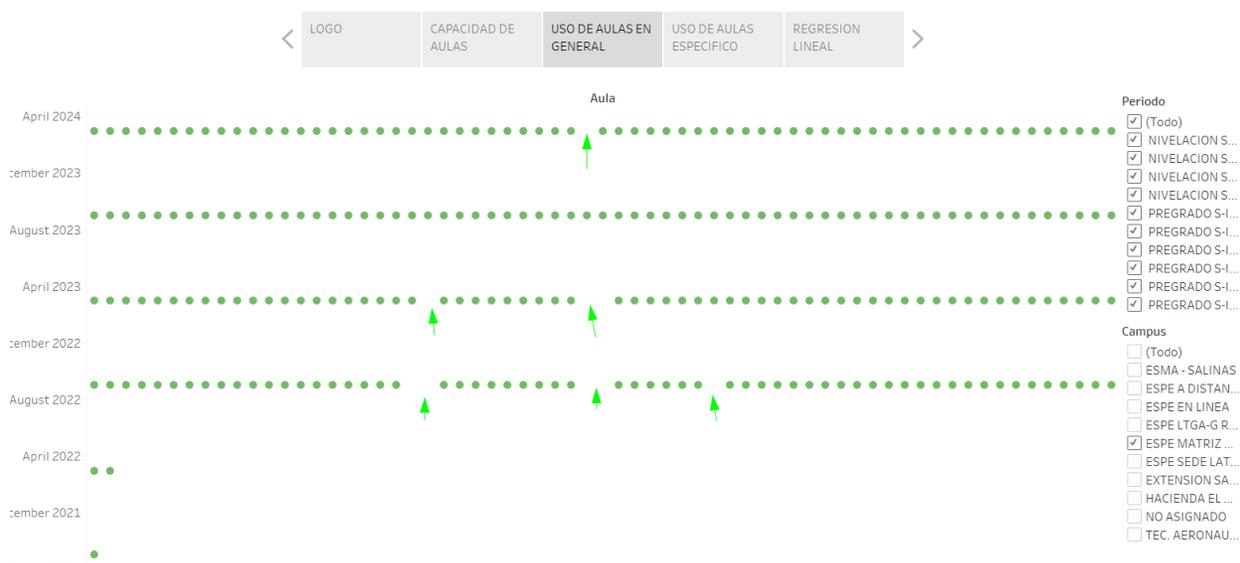
Aulas Aceptadas: Se refiere a aquellas aulas cuya capacidad es adecuada para el espacio físico disponible. En este análisis, se identifican y se muestran las aulas cuya capacidad se considera óptima y proporcionada al tamaño del espacio físico. Estas aulas son las más adecuadas para el desarrollo de actividades académicas.

Capacidad No Aceptada: Este aspecto se centra en aquellas aulas cuya capacidad es superada por el número de estudiantes matriculados en la asignatura correspondiente. En este análisis, se identifican y se muestran las aulas cuya capacidad resulta insuficiente para albergar a todos los estudiantes matriculados. Esta situación puede generar problemas de comodidad y eficiencia en el desarrollo de las actividades académicas.

Capacidad Dudosa: Se refiere a las aulas de gran capacidad en las que el número de estudiantes matriculados es notablemente inferior a su capacidad máxima. En este análisis, se identifican y se muestran aquellas aulas que, a pesar de tener una capacidad considerable, son utilizadas por un número reducido de estudiantes. Esto puede indicar una subutilización de recursos y una posible optimización en la asignación de aulas.

Figura 17

Uso de aulas en la Universidad de las Fuerzas Armadas (ESPE)



Nota. En esta figura observamos un detallado análisis sobre la utilización de las aulas en general. El análisis se centra en esas aulas que lamentablemente no han sido aprovechadas durante todo el semestre académico. Fuente: Autoría Propia

Este análisis es de gran importancia para evaluar las condiciones y validar la falta de uso de estas aulas. La visualización proporciona una representación gráfica de las aulas que han permanecido inactivas a lo largo del semestre, lo que permite identificar rápidamente aquellas instalaciones que no han sido aprovechadas. Este análisis es crucial para comprender las razones detrás de la falta de uso de estas aulas y puede ayudar a tomar decisiones informadas sobre su asignación y gestión en el futuro.

Al identificar las aulas que han estado inactivas durante todo el semestre, se pueden investigar las posibles causas de su subutilización, como problemas de ubicación, condiciones físicas inadecuadas, falta de demanda para ciertos cursos o programas, entre otros factores.

Esta información es valiosa para mejorar la eficiencia en la asignación de recursos y optimizar el uso de las instalaciones académicas.

Figura 18

Detalle del Uso de Aulas por Hora y Día



Nota. En esta figura se presenta un análisis detallado del uso de cada una de las aulas cuando se está especificando el uso por hora y día a lo largo del semestre académico. Fuente: Autoría Propia.

Este nivel de detalle permite obtener una visión completa del patrón de uso de cada aula individualmente.

La visualización muestra la distribución del uso de cada aula en función de las horas del día y los días de la semana. Esto proporciona información valiosa sobre los momentos de mayor y menor actividad en cada aula, así como los días de la semana en los que se concentra el uso.

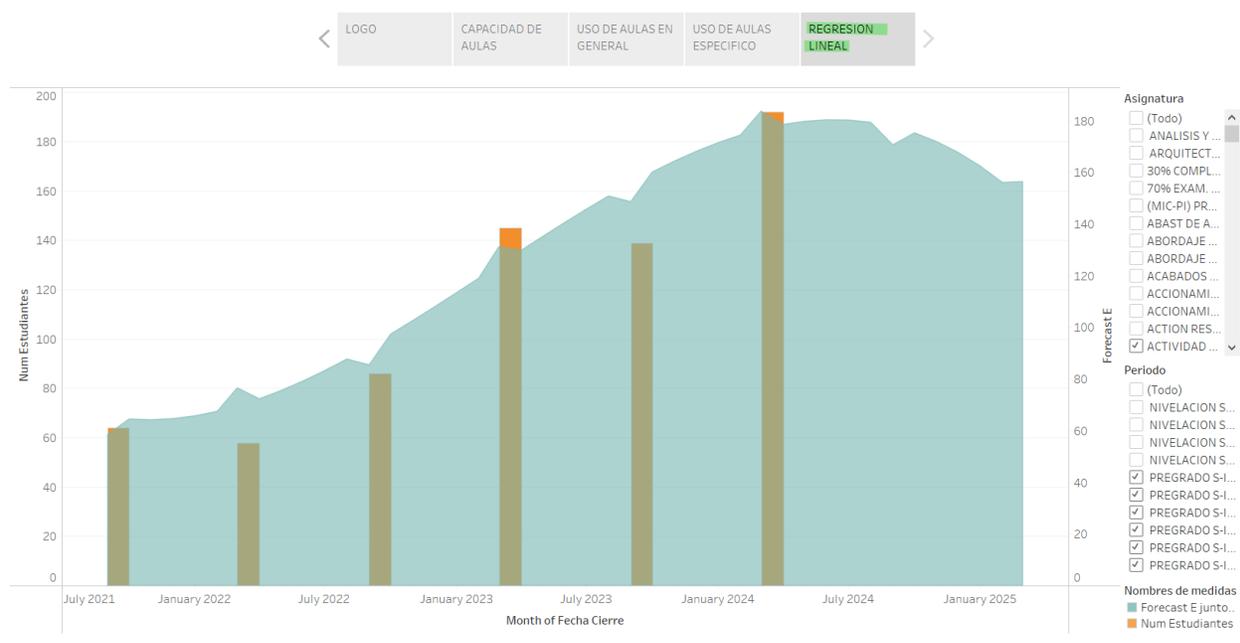
Este análisis detallado permite identificar patrones de uso específicos para cada aula, lo que puede ayudar a comprender mejor las necesidades de programación y asignación de

recursos. Esto permite identificar problemas potenciales o incoherencias en el uso de las instalaciones como por ejemplo aulas que pueden estar siendo subutilizadas en determinados horarios o días.

La información que brinda esta herramienta es vital para optimizar la programación y asignación de aulas, asegurando un uso eficiente de los recursos y ofreciendo una mejor experiencia para los estudiantes y docentes.

Figura 19

Análisis Avanzado: Modelo de Machine Learning con uso de Regresión Lineal de Estudiantes por Materia



Nota. Se presenta un análisis de machine learning avanzado realizado mediante un modelo de regresión línea la predicción del número de estudiantes por materia. Fuente: Autoría Propia.

Este análisis nos da la enseñanza que utilizando técnicas estadísticas se logra identificar las interrelaciones y tendencias entre variables para así predecir la cantidad de estudiantes matriculados en cada materia.

Este modelo de regresión se ha desarrollado a partir de datos históricos de la cantidad de estudiantes matriculados en diferentes cursos a lo largo de varios periodos académicos. A través de estos datos nuestro modelo establece una relación entre variables como la materia, el semestre académico, el programa académico, entre otros, y la cantidad de estudiantes matriculados. Dando como resultado tener una proyección de análisis para el cálculo de capacidades de aulas en un futuro.

La visualización que ahora la tecnología nos ofrece es lograr los resultados planteados en el ámbito de predicción sobre el número esperado de estudiantes para cada materia en periodos académicos futuros. Estas predicciones son fundamentales para la planificación y asignación de recursos permitiendo a las instituciones educativas prever la demanda de estudiantes y tomar decisiones acertadas sobre la oferta de cursos y la capacidad de las aulas.

El análisis avanzado que se realiza mediante el modelo de regresión lineal proporciona un enfoque intuitivo para la administración eficiente de los recursos hacia un futuro y la optimización de la planificación académica que nos contribuye al mejoramiento de la calidad de la experiencia educativa que la institución ofrece.

A continuación, se brindan detalles sobre las pruebas que se han realizado y cómo estas han contribuido al logro de los objetivos establecidos:

Pruebas Realizadas y Contribución al Logro de Objetivos

1. Integración de Datos:

Las pruebas se realizaron para validar la habilidad del sistema en integrar datos de numerosas hojas de cálculo en un repositorio centralizado. Comprobando que la eficiencia del proceso de integración así también como la coherencia datos y fiabilidad de los datos almacenados.

2. Acceso a la Información:

Se realizaron pruebas exhaustivas como el acceder de forma correcta y de formas incorrectas que se puede acceder a la información relevante para la Unidad de Educación Presencial (UEP). Se comprobó que el sistema permite a los usuarios consultar y analizar los datos de manera intuitiva y efectiva.

3. Seguridad de los Datos:

Se realizaron pruebas rigurosas para validar los mecanismos de seguridad diseñados para proteger la confidencialidad, integridad y disponibilidad de los datos, como por ejemplo accediendo con otras computadoras a los datos a pesar de tener un sistema similar, dando como resultado en acceso únicamente a las ip establecidas para el acceso. Se garantizó la protección de la información sensible y el cumplimiento de los estándares de seguridad.

4. Análisis Avanzados:

Se realizaron propuestas de machine learning dentro de Tableau la su visualización de proyecciones hacia un futuro evaluando la cantidad de estudiantes por materias. De igual forma evaluó la habilidad del sistema en generar percepciones significativas gracias a la historia creada por los datos proporcionados.

5. Automatización de Procesos:

Se evaluó la automatización de procesos como la extracción, transformación y carga de datos (ETL). El proceso que se automatizo fueron varios empezando el encendido y apagado de la máquina virtual automáticamente para luego en la carga de los archivos el procesamiento se totalmente automático y finalmente la visualización cuando termine el proceso sea actualizada automáticamente. Logrando esta automatización que los funcionales de la Unidad

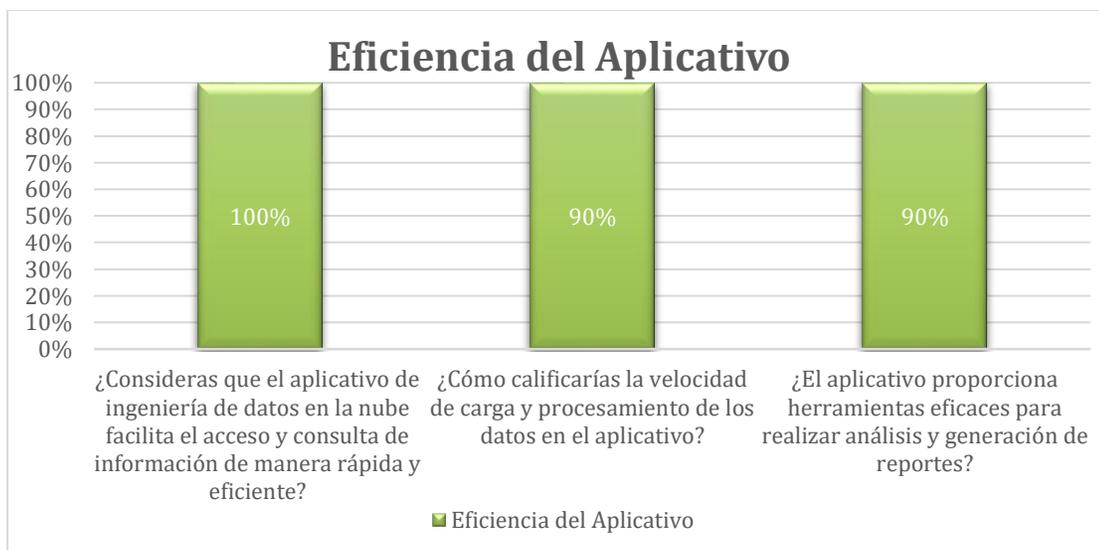
de Educación Presencial únicamente realicen la carga de archivos y ya solo realicen los análisis correspondientes con los resultados en Tableau.

6. Escalabilidad y Flexibilidad:

Se realizaron implementaciones de lagos de datos como una implementación de servicios como bases de datos en la nube para tener una mayor escalabilidad y también configuraciones internas que permitan acceder de forma eficiente a una escalabilidad y de igual forma se propone mantener a expertos en datos el nuevas adaptaciones que puedan existir en algún futuro y sean necesidades cambiantes de la Unidad de Educación Presencial (UEP) y de soportar volúmenes variables de datos sin comprometer el rendimiento. La escalabilidad que tiene es poder aumentar más procesos del área y de igual forma el aumento de data por que el sistema creado en la nube es dinámico ayudando a tener mayor flexibilidad

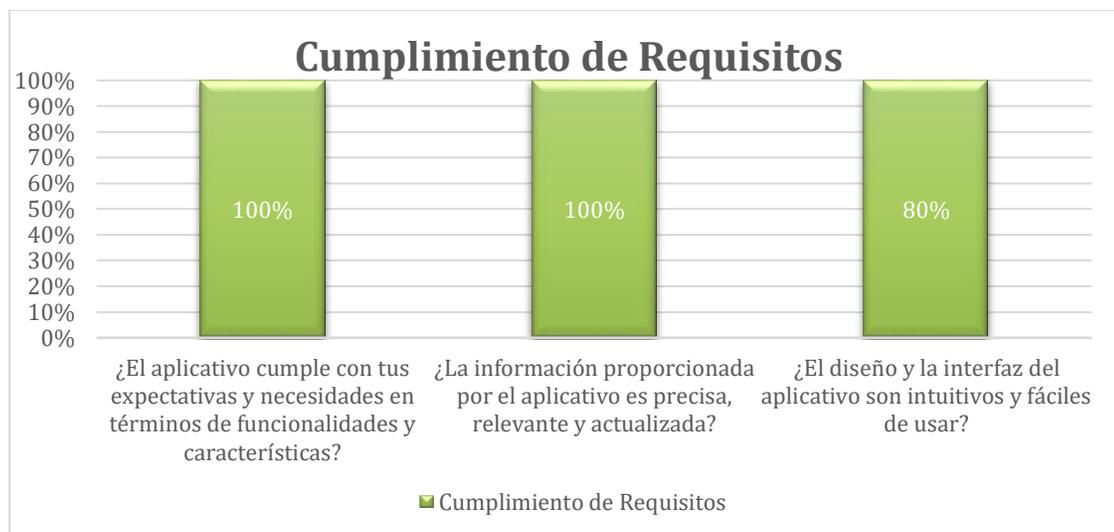
Evaluación del aplicativo:

En la Sección 1, evaluamos la eficiencia del aplicativo. Los usuarios han reconocido la accesibilidad y consulta de información rápida y eficiente. Sin embargo, la velocidad de carga y procesamiento de los datos se calificó como "moderada". Y mientras se confirmó que el aplicativo proporciona herramientas eficaces para analizar y generar reportes, los encuestados indicaron que estas podrían mejorarse.

Figura 20*Eficiencia del Aplicativo*

Nota. Encuesta realizada para evaluar la eficiencia del aplicativo.

En la Sección 2, referente al cumplimiento de los requisitos, el aplicativo ha cumplido en su mayoría con las expectativas y necesidades de funcionalidades y características de los usuarios, proporcionando información precisa, relevante y actualizada. Sin embargo, se ha señalado que el diseño y la interfaz, aunque son intuitivos, podrían ser mejorados explorando otras herramientas de visualización.

Figura 21*Cumplimiento de Requisitos*

Nota. Encuesta realizada para evaluar los cumplimientos de los requisitos.

Para la Sección 3, se destacó la necesidad de mejorar la visualización de datos y la inclusión de reportes desglosados por factor en lugar de reportes generales. También, se sugirió la combinación con otra aplicación que permita una visualización de reportes dinámica y adaptada a la necesidad institucional.

Estos resultados nos guían en la optimización de nuestros procesos de planificación académica semestral en la asignación de infraestructura universitaria a través de la automatización y el análisis de datos. Esto servirá para trabajos futuros mejorando la eficiencia, cumpliendo con las expectativas y proporcionando una experiencia de usuario satisfactoria.

Capítulo V

Conclusiones

La propuesta de arquitectura de ingeniería de datos en la nube para la Unidad de Educación Presencial (UEP) de la Universidad de las Fuerzas Armadas (ESPE) como un hito importante para mejorar la gestión de datos. El proyecto, denominado "Diseño e implementación de una propuesta de arquitectura de ingeniería de datos en la nube para la construcción de un repositorio centralizado con capacidades (ETL)" se destaca por tener una serie de mejoras que optimizarán el proceso de toma de decisiones y la eficiencia operacional si se implementa con éxito.

La automatización de procesos ETL sugerida en el proyecto, se identifica como la unión de varios elementos esenciales logrando una óptima gestión de datos. Permitiendo un ahorro significativo de tiempo y recursos para la Unidad de Educación Presencial, de igual manera el incremento de la precisión y confiabilidad de los datos procesados. Adicionalmente la propuesta de herramientas de visualización como Tableau, será de gran ayuda para generar historias de datos valiosos y facilitar decisiones informadas por parte de los usuarios finales.

La integración eficaz de un repositorio en la nube de acuerdo con parámetros de negocio preestablecidos fue clave para el éxito potencial del proyecto. Al lograr completar todos los desafíos propuestos dando solución a cada uno de ellos, al igual que se logró que la plataforma seleccionada sea escalable, segura y se adapte a las necesidades cambiantes de procesamiento de datos, proporcionando una base sólida para la extracción, transformación y carga de datos.

Las implicaciones de esta propuesta son significativas, no solo en la mejora de la gestión de datos de la Unidad de Educación Presencial (UEP), sino también como un camino de modalidad de trabajo dejando atrás los trabajos operativos y arquitecturas locales, para la

toma de decisiones en la distribución de aulas y laboratorios para tener una mejor administración y coordinación de uso. La implementación exitosa de esta arquitectura de ingeniería de datos en la nube podría tener un precedente para futuros proyectos tecnológicos en la Universidad fomentando una nueva cultura de innovación y eficiencia en la gestión de datos. Usando la tecnología que nos brindan como estudiantes al tener accesos como al portal de Azure gratis y otras nubes más.

En el transcurso de nuestro trabajo, hemos entendido la importancia de recoger y considerar las opiniones y experiencias de nuestro público objetivo. Como parte de nuestra metodología cascada, hemos realizado una encuesta exhaustiva, donde los resultados y los comentarios obtenidos han sido sumamente valiosos.

La encuesta ha generado reacciones muy positivas y la receptividad a las soluciones han sido altas. Esto nos llena de orgullo y nos motiva aún más a seguir ofreciendo soluciones innovadoras y efectivas para la automatización de procesos de planificación académica en la Unidad de Educación Presencial (UEP).

Recomendaciones

Basándose en los aprendizajes y la experiencia lograda durante este proyecto, se sugieren las siguientes recomendaciones para futuras investigaciones y desarrollos en el campo de la ingeniería de datos:

1. Evaluación Continua del sistema implementado: Se aconseja realizar un seguimiento de las nuevas tecnologías y evaluación constante una vez que la arquitectura de ingeniería de datos en la nube tanto en el trabajo propuesto como en nuevas áreas de mejora y optimización.
2. Capacitación del Personal de ingeniería de datos: Es importante tener el personal adecuado con la experiencia en el campo porque este tipo de arquitectura se necesita que se

encuentre involucrado en el manejo y mantenimiento de la arquitectura de datos en la nube para aprovechar al máximo las funcionalidades y herramientas disponibles.

3. **Monitoreo y Mantenimiento:** Se sugiere implementar un sistema de monitoreo continuo con personal de experiencia y mantenimiento proactivo para garantizar un rendimiento óptimo y la seguridad de la infraestructura en la nube.
4. **Actualización Tecnológica:** Es importante estar al día con las últimas tendencias en tecnología de la nube y considerar actualizaciones o mejoras en los proyectos de arquitecturas de datos para incorporar nuevas funcionalidades u optimizaciones.
5. **Comentarios de los usuarios con respecto al uso y resultados del aplicativo:** Se recomienda recoger de forma regular la retroalimentación de los usuarios finales sobre la usabilidad y efectividad de la arquitectura en la nube, esto importante porque nos ayuda a mantener una amplia perspectiva de cómo se está usando y si está sirviendo o que datos podrían ahora servir.
6. **Colaboración Interdepartamental:** Fomentar la colaboración entre varias áreas como tener un equipo de datos teniendo un gobierno de datos para esquematizar todos equipos de tecnología de la información como administración académica y otros departamentos.

Bibliografía

- Alfaro, H. (2023). *Documento final de Proyecto de Investigación Aplicada 2*.
- Alshdadi, A., Yahya, F., Wills, G., & Walters, R. (2015, mayo 20). *An Overview of Cloud Services Adoption Challenges in Higher Education Institutions*.
<https://doi.org/10.5220/0005529701020109>
- Benavente Toro, A. A. (2023). *Visualizador de datos de atenciones ambulatorias del hospital Guillermo Grant Benavente de Concepción*.
<http://repositorio.udec.cl/jspui/handle/11594/11480>
- Chavarría Romero, B. B. (2021). Implementación de un modelo de Business Intelligence para mejorar el proceso de toma de decisiones en el área de Trade marketing de la marca Puma, Lima, 2019. *Universidad Católica Sedes Sapientiae*.
<https://repositorio.ucss.edu.pe/handle/20.500.14095/1223>
- Dixson, Y. R., & Maturel, L. N. (2015). La inteligencia de negocio como apoyo a la toma de decisiones en el ámbito académico. *GECONTEC: Revista Internacional de Gestión del Conocimiento y la Tecnología*, 3(2), Article 2. <https://doi.org/10.5281/zenodo.7469263>
- Ghosh, P. (Guha). (2022, febrero 22). Data Architecture Trends in 2022. *DATAVERSITY*.
<https://www.dataversity.net/data-architecture-trends-in-2022/>
- Guerrero, C. A., & Londoño, J. M. (2016). Revisión de la Problemática de la Calidad del Software para el Desarrollo de Aplicaciones de Computación en la Nube. *Información tecnológica*, 27(3), 61–80. <https://doi.org/10.4067/S0718-07642016000300007>
- Gutierrez Arana, R. A., & Asca Arevalo, B. C. (2023). Migración de sistema de pipeline de datos de un ambiente On Premise a servicios basados en Cloud Computing. *Universidad Peruana de Ciencias Aplicadas (UPC)*.
<https://repositorioacademico.upc.edu.pe/handle/10757/671048>

- Juma, M. K., & Tjahyanto, A. (2019). Challenges of Cloud Computing Adoption Model for Higher Education Level in Zanzibar (the Case Study of SUZA and ZU). *Procedia Computer Science*, 161, 1046–1054. <https://doi.org/10.1016/j.procs.2019.11.215>
- Khan, A., Ghosh, S., & Ghosh, S. K. (2018). eDWaaS: A Scalable Educational Data Warehouse as a Service. En A. Abraham, P. Kr. Muhuri, A. K. Muda, & N. Gandhi (Eds.), *Intelligent Systems Design and Applications* (pp. 998–1007). Springer International Publishing. https://doi.org/10.1007/978-3-319-76348-4_96
- Loza, J. (2023, julio 5). (17) *Arquitectura de Datos en la Nube: Tendencias y Perspectivas Futuras* | LinkedIn. <https://www.linkedin.com/pulse/arquitectura-de-datos-en-la-nube-tendencias-y-futuras-jose-loza/?originalSubdomain=es>
- Paredes-Parada, W. (2021). *TecnologíES Tecnologías para las instituciones de educación superior (IES) y sus experiencias*. CEDIA. <https://repositorio.cedia.edu.ec/handle/123456789/4>
- Prieto, E. (2023, noviembre 16). *¿Cuáles son las etapas del Desarrollo de Software?* - Tiffin University. <https://global.tiffin.edu/noticias/cuales-son-las-etapas-del-desarrollo-de-software>
- Priyadarshini, A., & Rath, S. K. (2020). *Data Warehousing and Data Mining in the Education System*. 10(09).
- Proaño, M. F., Orellana, S. Y., & Martillo, I. O. (2018). Los sistemas de información y su importancia en la transformación digital de la empresa actual. *Revista ESPACIOS*, 39(45). <https://es.revistaespacios.com/a18v39n45/18394503.html>
- Rafika Helaimia. (2023, enero). *CLOUD COMPUTING IN HIGHER EDUCATION INSTITUTIONS: PROS AND CONS*. <https://as-proceeding.com/index.php/ijanser/article/view/381>
- Rodzi, N. A. H. M., Othman, M. S., & Yusuf, L. M. (2015). Significance of data integration and ETL in business intelligence framework for higher education. *2015 International*

Conference on Science in Information Technology (ICSITech), 181–186.

<https://doi.org/10.1109/ICSITech.2015.7407800>

Whale Design. (2023, enero 1). *Modelo de cascada*. Vecteezy. <https://es.vecteezy.com/artefectorial/8423055-el-modelo-de-cascada-infografico-vector-se-utiliza-en-ingenieria-de-software-o-procesos-de-desarrollo-de-software-la-ilustracion-tiene-6-pasos-como-metodologia-agil-o-pensamiento-de-diseno-para-sistema-de-aplicaciones>

Yahuarcani, H. J. S., & Devia, J. S. M. (2022). *PROTOTIPO DE UN SISTEMA DE INFORMACIÓN PARA LA GESTIÓN DE DATOS FISIOLÓGICOS MEDIANTE ETL*.

Yulianto, A. A. (2019). Extract Transform Load (ETL) Process in Distributed Database Academic Data Warehouse. *Aptikom Journal on Computer Science and Information Technologies*, 4(2), Article 2. <https://doi.org/10.11591/APTIKOM.J.CSIT.36>