Universidad Nacional Mayor de San Marcos
Universidad del Perú. Decana de América
Facultad de Ingeniería Industrial
Escuela Profesional de Ingeniería Industrial

Transformación digital en un área de recursos humanos mediante la implementación de un data lake y la automatización robótica de procesos (RPA)

TESIS

Para optar el Título Profesional de Ingeniero Industrial

AUTOR

Ivan Jeremy PORRAS ROJAS

ASESOR

Pedro Pablo ROSALES LÓPEZ

Lima, Perú

2026



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

https://creativecommons.org/licenses/by-nc-sa/4.0/

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Porras, I. (2026). *Transformación digital en un área de recursos humanos mediante la implementación de un data lake y la automatización robótica de procesos (RPA)*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería Industrial, Escuela Profesional de Ingeniería Industrial]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	IVAN JEREMY PORRAS ROJAS
Tipo de documento de identidad	DNI
Número de documento de identidad	41717965
Datos de asesor	
Nombres y apellidos	PEDRO PABLO ROSALES LÓPEZ
Tipo de documento de identidad	DNI
Número de documento de identidad	10419269
URL de ORCID	https://orcid.org/0000-0001-7544-5539
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	HERMES YESSER PANTOJA CARHUAVILCA
Tipo de documento	DNI
Número de documento de identidad	09656026
Miembro del jurado 1	
Nombres y apellidos	HILMAR ANTONIO HINOJOSA LAZO
Tipo de documento	DNI
Número de documento de identidad	29537977
Miembro del jurado 2	
Nombres y apellidos	CARLOS CÉSAR PIZARRO BARBARÁN
Tipo de documento	DNI
Número de documento de identidad	07565210
Datos de investigación	

Línea de investigación	9. Gestión organizacional sostenible. ODS 9: INDUSTRIA, INNOVACIÓN E INFRAESTRUCTURA. C.17.0.9
Grupo de investigación	No aplica.
Agencia de financiamiento	Sin financiamiento
Ubicación geográfica de la investigación	País: Perú Departamento: Lima Provincia: Lima Distrito: San Miguel Calle: Av. Universitaria 1801 Latitud -12.06921 Longitud -77.08055
Año o rango de años en que se realizó la investigación	Marzo 2023 - octubre 2024
URL de disciplinas OCDE	Ingeniería industrial https://purl.org/pe-repo/ocde/ford#2.11.04

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
Universidad del Perú. Decana de América

FACULTAD DE INGENIERÍA INDUSTRIAL

VICEDECANATO ACADÉMICO

ACTA DE SUSTENTACIÓN N°014-VDAP-FII-2026

SUSTENTACIÓN DE TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO INDUSTRIAL

El Jurado designado por la Facultad de Ingeniería Industrial, reunido en acto público en el Auditorio de la Facultad de Ingeniería Industrial, el día **viernes 08 de mayo de 2026**, a las 15:00 horas, dio inicio a la sustentación de la tesis:

"TRANSFORMACIÓN DIGITAL EN UN ÁREA DE RECURSOS HUMANOS MEDIANTE LA IMPLEMENTACIÓN DE UN DATA LAKE Y LA AUTOMATIZACIÓN ROBÓTICA DE PROCESOS (RPA)"

Que presenta el Bachiller:

IVAN JEREMY PORRAS ROJAS

Para optar el Título Profesional de Ingeniero Industrial en la Modalidad: **Ordinaria.**

Luego de la exposición, absueltas las preguntas del Jurado y siendo las 16:05 horas se procedió a la evaluación secreta, habiendo sido Aprobado por Unanimidad con la calificación promedio de 17, lo cual se comunicó públicamente.

Lima, 08 de mayo del 2026

DR. HERMES YESSER PANTOJA CARHUAVILCA
Presidente

MG. HILMAR ANTONIO HINOJOSA LAZO
Miembro

MG. CARLOS CÉSAR PIZARRO BARBARÁN
Miembro

DR. PEDRO PABLO ROSALES LÓPEZ
Asesor

DR. JULIO ALEJANDRO SALAS BACALLA
Vicedecano Académico – FII

UNIVERSIDAD NACIONAL MAYOR DE
SAN MARCOS
Universidad del Perú, Decana de América

CERTIFICADO DE SIMILITUD

Yo, **ROSALES LÓPEZ PEDRO PABLO** en mi condición de asesor acreditado con el **DICTAMEN N° 027-EPII-FII-2026** de la tesis, cuyo título es **TRANSFORMACIÓN DIGITAL EN UN ÁREA DE RECURSOS HUMANOS MEDIANTE LA IMPLEMENTACIÓN DE UN DATA LAKE Y LA AUTOMATIZACIÓN ROBÓTICA DE PROCESOS (RPA)**, presentado por el bachiller **IVAN JEREMY PORRAS ROJAS** para optar el **TÍTULO PROFESIONAL DE INGENIERO INDUSTRIAL**, CERTIFICO que se ha cumplido con lo establecido en la Directiva de Originalidad y de Similitud de Trabajos Académicos, de Investigación y Producción Intelectual. Según la revisión, análisis y evaluación mediante la plataforma PIDS, el documento evaluado cuenta con el **7%** de similitud, nivel **PERMITIDO** para continuar con los trámites correspondientes y para su **publicación en el repositorio institucional.**

Se emite el presente certificado en cumplimiento de lo establecido en las normas vigentes, como uno de los requisitos para la obtención del Título Profesional de Ingeniero Industrial.

Lima, 16 de julio del 2026

UNMSM
Firmado digitalmente por ROSALES LOPEZ Pedro Pablo FAU 20148092282 soft
Motivo: Soy el autor del documento
Fecha: 20.07.2026 09:22:22 -05:00

Firma del Asesor
Nombres y Apellidos del asesor: **ROSALES LÓPEZ PEDRO PABLO**
DNI: **10419269**

DEDICATORIA

A mis padres, que siempre me apoyaron y creyeron en mí, incluso cuando las circunstancias me alejaron de mis estudios. Esta tesis también es suya, porque nunca me juzgaron; por el contrario, me acompañaron, con paciencia y amor, esperando el momento en que yo pudiera retomar mi camino.

A mi esposa, gracias por estar a mi lado aun en la distancia. Durante estos años, tu fortaleza y tu amor sostuvieron este proyecto tanto como mis horas de estudio. Gracias por asumir tantos sacrificios, por cuidar de nuestra familia, y por no dejar de creer en lo que este esfuerzo significaba para nuestro futuro.

A mis dos hijas, que han sido mi mayor motivación. Retomar la universidad cuando ustedes ya formaban parte de mi vida, en medio de una realidad marcada por la pandemia y la distancia, hizo que cada clase, cada trabajo y cada desvelo tuvieran un significado especial. Esta tesis está dedicada a ustedes, porque puse el deseo de dejarles un ejemplo: que los sueños pueden tardar, pueden detenerse, pero no tienen por qué abandonarse. Ojalá que, cuando sean mayores, vean en este logro la prueba de que todo sacrificio tiene su recompensa.

TABLA DE CONTENIDO

RESUMEN vii

ABSTRACT viii

INTRODUCCIÓN 1

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA 2

1.1. Situación problemática 2

1.2. Formulación del problema 5

1.3. Objetivos 5

1.3.1. Objetivo General 5

1.3.2. Objetivos Específicos 5

1.4. Importancia y alcance de la investigación 5

CAPÍTULO II: MARCO TEÓRICO 8

2.1. Antecedentes de la investigación 8

2.2. Bases teóricas 13

2.2.1. Transformación Digital 13

2.2.2. Data Lake 13

2.2.3. Automatización Robótica de Procesos (RPA) 17

2.3. Definición de términos 22

2.3.1. AWS (Amazon Web Services) 22

2.3.2. Servicio Amazon S3 22

2.3.3. Servicio de Amazon Athena 22

2.3.4. Capa Raw 22

2.3.5. Capa Stage 22

2.3.6. ERP 23

2.3.7. SIUES 23

2.3.8. Modelo de Datos 23

2.3.9. Data Flow (Power BI) 23
2.3.10. Data Driven 24
2.3.11. Modelo Semántico (Power BI) 24
2.3.12. Data Mart 24
2.3.13. Workspace (Power BI) 24
2.3.14. Toma de Decisiones 24
2.3.15. Gestión de Recursos Humanos 25
2.3.16. Procesos Administrativos 25
CAPÍTULO III: HIPÓTESIS Y VARIABLES 26
3.1. Hipótesis 26
3.1.1. Hipótesis General 26
3.1.2. Hipótesis Específicas 26
3.2. Variables 26
3.3. Operacionalización 27
CAPÍTULO IV: MATERIALES Y MÉTODOS 29
4.1. Tipo de investigación 29
4.2. Diseño de investigación 29
4.3. Población 29
4.4. Muestra 29
4.4.1. Justificación del tamaño muestral 30
4.5. Procedimientos, técnicas e instrumentos de recolección de datos 32
4.5.1. Capacitación 33
4.5.2. Implementación 33
4.5.3. Pruebas y validación 33
4.5.4. Evaluación y ajustes 33
4.5.5. Documentación y capacitación final 34
4.6. Análisis estadístico e interpretación de la información 34

4.6.1. Prueba de normalidad 34
4.6.2. Prueba de t de Student 34
4.6.3. Prueba de Chi-cuadrado 35
4.7. Análisis de validez de instrumentos 36
4.7.1. Validación por juicio de expertos 36
CAPÍTULO V: RESULTADOS 39
5.1. Antecedentes del Proyecto 39
5.2. Presentación de Resultados 40
5.2.1. Acceso a datos (Data Lake) 41
5.2.2. Automatización de tareas (RPA) 44
5.2.3. Percepción de mejora (Transformación Digital) 46
CAPÍTULO VI: DISCUSIÓN 52
6.1. Mejora en el acceso a los datos (Data Lake) 52
6.2. Reducción de tiempos administrativos (RPA) 53
6.3. Asociación entre percepciones 53
CAPÍTULO VII: CONCLUSIONES Y RECOMENDACIONES 54
7.1. Conclusiones 54
7.2. Recomendaciones 55
REFERENCIAS BIBLIOGRÁFICAS 57
ANEXOS 59

INDICE DE TABLAS

Tabla 1. Data Warehouse vs. Data Lake 15

Tabla 2. Matriz de Operacionalización de Variables 27

Tabla 3. Muestra a considerar en la investigación 30

Tabla 4. Análisis de Variables 35

Tabla 5. Tiempo de Acceso a los datos 42

Tabla 6. Prueba de Normalidad (Acceso a los datos) 42

Tabla 7. Tiempo de Automatización de tareas 44

Tabla 8. Prueba de Normalidad (Tiempo de Automatización de tareas) 44

Tabla 9. Sección A: Percepción del acceso a datos 46

Tabla 10. Sección B: Percepción en la toma de decisiones 46

Tabla 11. Resultados de encuesta 47

Tabla 12. Percepción del acceso a datos (Tabla de Frecuencia) 47

Tabla 13. Percepción en la toma de decisiones (Tabla de Frecuencia) 48

Tabla 14. Resultados descriptivos 48

Tabla 15. Tabla cruzada Percepción del acceso a datos y toma de decisiones 49

Tabla 16. Pruebas de chi-cuadrado 50

INDICE DE FIGURAS

Figura 1. Arquitectura de datos y flujo de información inicial 3

Figura 2. Evaluación de proveedores de RPA 21

Figura 3. Arquitectura de datos actual..... 41

Figura 4. Prueba t de Student (Acceso a los datos)..... 43

Figura 5. Prueba t de Student (Tiempo de Automatización de tareas)..... 45

RESUMEN

La presente investigación tiene como objetivo transformar digitalmente el área de Recursos Humanos (RRHH) de una universidad de educación superior (UES) mediante la implementación de un Data Lake y la Automatización Robótica de Procesos (RPA). Se parte de una problemática basada en la fragmentación de los datos institucionales y la alta carga de tareas manuales que limitan la eficiencia operativa. El estudio se desarrolló bajo un enfoque aplicado y diseño explicativo, utilizando como muestra a 16 colaboradores clave del área de RRHH.

Para el desarrollo del Data Lake se utilizó la arquitectura de Amazon Web Services (AWS), integrando 15 fuentes de información que anteriormente se encontraban dispersas en 11 sistemas y archivos. Esta implementación permitió reducir el tiempo promedio de acceso a los datos de 18.56 a 3.16 minutos. Asimismo, la adopción de RPA, mediante UiPath, logró automatizar tres procesos administrativos críticos, disminuyendo el tiempo de ejecución de 15.62 a 9.03 minutos. Las percepciones de los usuarios también fueron altamente positivas: el 100% afirmó que mejoró el acceso a datos y el 87.5% indicó que la toma de decisiones se fortaleció.

Se concluye que la integración de tecnologías emergentes como Data Lake y RPA no solo optimiza procesos, sino que genera capacidades institucionales para una transformación digital efectiva y sostenible.

Palabras clave: transformación digital, Data Lake, RPA, acceso a datos, toma de decisiones.

ABSTRACT

This research aims to digitally transform the Human Resources (HR) area of a higher education institution (UES) through the implementation of a Data Lake and Robotic Process Automation (RPA). The study addresses the institutional challenge of fragmented data systems and the burden of manual administrative tasks. The methodology follows an applied and explanatory design, with a non-probabilistic sample of 16 HR staff members directly involved in data and process management.

The Data Lake was developed using Amazon Web Services (AWS), integrating 15 data sources previously dispersed across 11 platforms. As a result, the average data access time was reduced from 18.56 to 3.16 minutes. Simultaneously, RPA technology (UiPath) was used to automate three key administrative processes, reducing execution time from 15.62 to 9.03 minutes. Survey responses showed that 100% of participants perceived improvements in data access and 87.5% reported enhanced decision-making capabilities.

It is concluded that integrating technologies such as Data Lake and RPA not only optimizes operational workflows but also strengthens the institutional capacity for sustainable digital transformation.

Keywords: digital transformation, Data Lake, RPA, data access, decision-making.

INTRODUCCIÓN

En un entorno institucional caracterizado por la creciente complejidad de los procesos y la necesidad de respuestas ágiles y basadas en datos, la transformación digital se ha convertido en una prioridad estratégica. En el caso de las instituciones de educación superior, esta transformación trasciende a la incorporación de tecnologías; implica repensar los procesos, integrar sistemas de información y optimizar el uso del capital humano.

La presente investigación se enmarca en este contexto, abordando específicamente el área de Recursos Humanos (RRHH) de una universidad de educación superior, a la que se denominará UES. Esta unidad enfrenta limitaciones vinculadas a la dispersión de datos en múltiples plataformas y al uso intensivo de procedimientos manuales que generan ineficiencias y dificultan la toma de decisiones estratégicas.

Frente a esta problemática, se plantea como solución la implementación de un Data Lake, que permita centralizar y estandarizar la información operativa, junto con la adopción de tecnologías de Automatización Robótica de Procesos (RPA) con el fin de reducir la carga asociada a tareas repetitivas. Estas iniciativas forman parte de un enfoque integral orientado a incrementar la eficiencia de los procesos, la calidad de la información y la capacidad analítica del área de RRHH.

El desarrollo de esta tesis busca demostrar, mediante evidencia cuantitativa y cualitativa, que la aplicación conjunta de un Data Lake y RPA puede generar un cambio sustantivo en la forma en que se gestiona la información y se toman decisiones dentro del ámbito de gestión de personas en una universidad.

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA

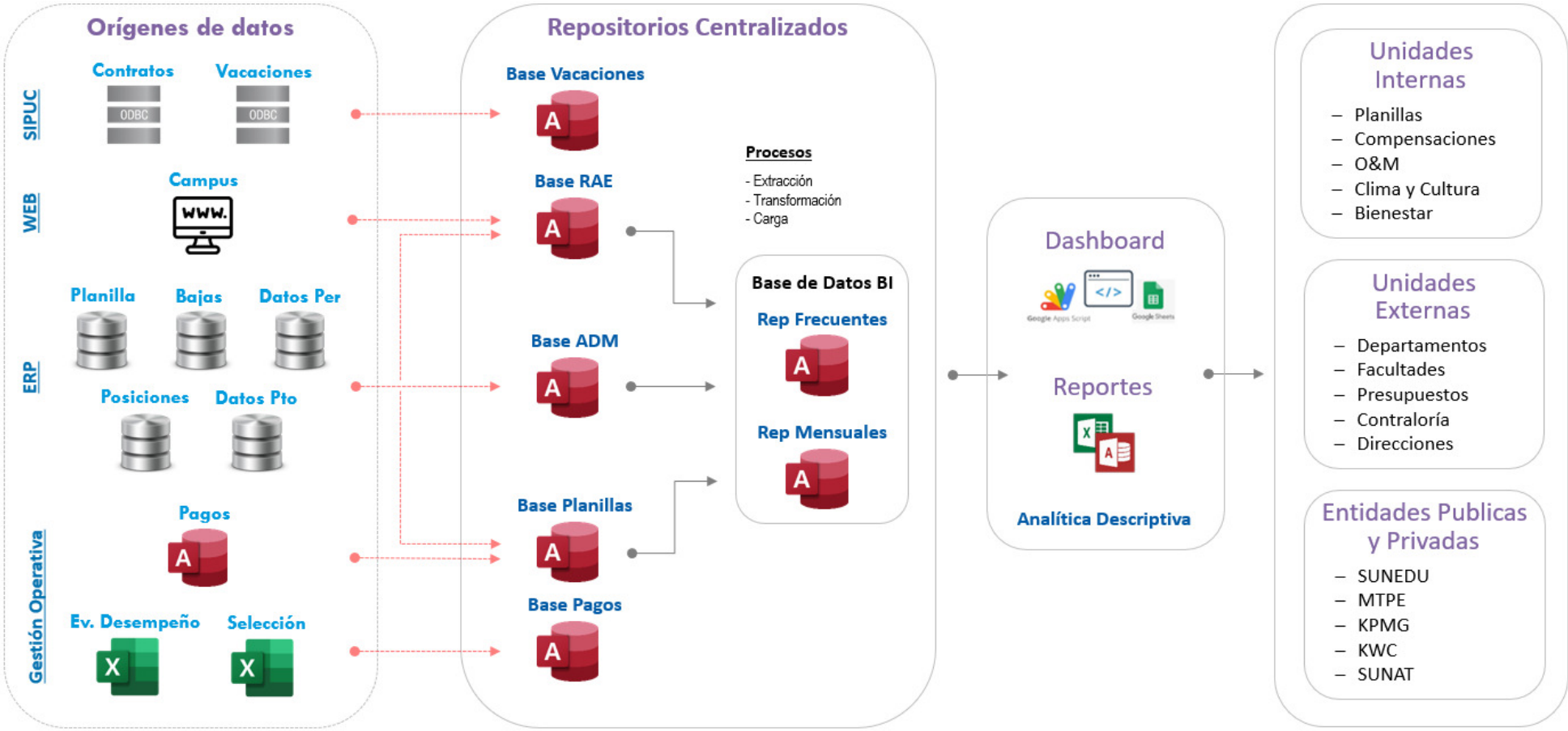
1.1. Situación problemática

Es crucial para todas las empresas ingresar al mundo digital mediante la implementación de herramientas tecnológicas, pero es aún más importante contar con el conocimiento y la capacitación adecuada para enfrentar estos nuevos desafíos. Se requiere un estudio y una capacitación sólida para aprovechar los beneficios organizativos que se obtienen al utilizar estas herramientas de manera adecuada (Mendieta, 2022).

En el contexto de la presente investigación, la empresa es una universidad de educación superior a la que denominaremos “UES”, es reconocida por la excelencia en su enseñanza, investigaciones, compromiso con la responsabilidad social y su promoción de la cultura. El objeto de esta investigación en su área de Recursos Humanos (RRHH).

La UES actualmente se encuentra enfrentando diversos desafíos vinculados con la gestión y el procesamiento de datos, toda vez que se cuenta con varios sistemas: ERP HCM (Módulo de Recursos Humanos), SIUES (sistema local para la gestión de Recursos Humanos), la Intranet institucional y diversos archivos de gestión operativa. Cada uno de estos sistemas y archivos posee su propia estructura de información. Esta fragmentación de los datos dificulta la toma de decisiones eficientes, ya que la falta de integración entre estos sistemas genera redundancias, retrasos y errores en los procesos de gestión de personal.

Figura 1. Arquitectura de datos y flujo de información inicial



Fuente: Elaboración Propia

Además, los procesos administrativos actuales son manuales en su mayoría, lo que consume una cantidad considerable de horas hombre, limita la capacidad del personal para enfocarse en tareas estratégicas y genera ineficiencias operativas. A medida que la UES avanza hacia su objetivo de Transformación Digital, se hace necesario adoptar soluciones que permitan integrar los datos de estos sistemas y automatizar procesos repetitivos, mejorando la eficiencia y optimizando el uso de recursos.

En este contexto, surge la necesidad de implementar un Data Lake que centralice la información de los distintos sistemas y archivos operativos, así como en utilizar la RPA para automatizar tareas repetitivas. Estas acciones están alineadas con el objetivo institucional de modernizar su gestión de recursos humanos y avanzar hacia una transformación digital completa, conforme al Plan Estratégico Institucional 2023–2027 de la UES. Sin embargo, se desconocen los efectos directos que estas implementaciones tendrán sobre la eficiencia operativa y el consumo de horas hombre.

Según (Almaraz et al., 2017) nuestra sociedad está adoptando tecnologías digitales a gran escala. Innovaciones tecnológicas que impulsan la llamada revolución digital, transformando rápidamente diversas organizaciones. Esta rápida digitalización genera incertidumbre para muchas entidades. Sectores como la prensa, editoriales, publicidad han sido profundamente afectados y han tenido replantear sus procesos para adaptarse. En este sentido, la UES no es ajena a esta tendencia y requiere incorporar soluciones como un Data Lake y RPA para modernizar su gestión de recursos humanos y responder a los desafíos actuales de eficiencia y manejo de datos.

1.2. Formulación del problema

1.2.1. Problema General

¿Cómo la implementación de un Data Lake y la Automatización Robótica de Procesos (RPA) permite la Transformación Digital (TD) en el área de RRHH?

1.2.2. Problemas Específicos

- PE1: ¿Cómo implementar un Data Lake mejora el acceso a los datos en el área de RRHH?
- PE2: ¿Cómo implementar RPA reduce los tiempos en tareas administrativas en el área de RRHH?

1.3. Objetivos

1.3.1. Objetivo General

Implementar un Data Lake y la Automatización Robótica de Procesos (RPA) para transformar digitalmente el área de RRHH.

1.3.2. Objetivos Específicos

- OE1: Implementar un Data Lake para mejorar el acceso a los datos en el área de RRHH.
- OE2: Implementar RPA para reducir los tiempos en las tareas administrativas en el área de RRHH.

1.4. Importancia y alcance de la investigación

Los avances en la automatización y gestión de datos han sido posibles gracias a los progresos tecnológicos, lo cual ha permitido el desarrollo de herramientas más

eficientes y precisas para la gestión de datos laborales, es por ello que (Mendieta, 2022) sostiene que las organizaciones que reconocen el uso de tecnologías digitales impulsan el crecimiento en todos los niveles, prosperan y se expanden en el mercado.

La implementación de un Data Lake y RPA en el área de Recursos Humanos no solo abordan los problemas existentes de gestión y procesamiento de datos, sino que también representan una oportunidad significativa para avanzar hacia una transformación digital integral en la institución.

La integración de múltiples sistemas y archivos de gestión operativa en un Data Lake permitirá consolidar datos en un solo repositorio, facilitando el acceso y la manipulación de la información. Esto reducirá la duplicación de esfuerzos e incrementará la eficiencia en la gestión.

La disponibilidad de datos integrados y actualizados en tiempo real proporcionará a los líderes de Recursos Humanos de la UES una base más sólida para tomar decisiones informadas y estratégicas. Esto permitirá una mejor planificación de personal, identificación de tendencias y evaluación del desempeño del personal.

La adopción de RPA en los procesos de Recursos Humanos permitirá a la UES automatizar tareas repetitivas y basadas en reglas, liberando tiempo para actividades más estratégicas y centradas en el talento. Además, esto posicionará a la institución para enfrentar los desafíos de la transformación digital, al alinear sus procesos con las mejores prácticas tecnológicas actuales.

Esta investigación se justifica por la necesidad de enfrentar los retos existentes en la gestión de datos y procesos dentro del área de Recursos Humanos de la UES, así como por la oportunidad de avanzar hacia una transformación digital que mejore

la eficiencia operativa y fortalezca la capacidad de la institución para adaptarse a un entorno digital en evolución constante.

CAPÍTULO II: MARCO TEÓRICO

2.1. Antecedentes de la investigación

En el marco de este estudio, se han identificado ciertos antecedentes relevantes que respaldan la necesidad y viabilidad de implementar un repositorio centralizado que unifique los datos dispersos existentes en el área de Recursos Humanos y poder agilizar procesos manuales con RPA. A continuación, se presentan algunos de estos antecedentes:

2.1.1. Internacionales

(Lizano & Palos, 2020) **en el artículo "Transformación Digital en Instituciones de Educación Superior con Gestión de Procesos de Negocio Modelo de mediación de Automatización Robótica de Procesos"** plantean un marco teórico para llevar a cabo la transformación digital utilizando la Gestión de Procesos de Negocio (BPM) junto con RPA en Instituciones de Educación Superior, La propuesta se enfoca en integrar sistemas heredados con nuevas tecnologías, evitando la pérdida de relevancia en los procesos sin una mediación adecuada.

El objetivo de los autores es proporcionar un marco teórico que potencie la transformación digital en instituciones educativas a través de BPM y RPA, maximizando el valor agregado en dichas instituciones.

(Romero et al., 2023) **en el artículo "La transformación digital en la educación superior el caso de la UOC"** plantea la importancia de adaptar las universidades tradicionales a modelos híbridos o en línea, una tendencia acelerada por la pandemia de COVID-19. Se empleó una metodología mixta para analizar cómo

la innovación digital puede impulsar la transformación institucional, abordando tanto los desafíos como los beneficios del proceso. El artículo también indica que la digitalización de las universidades no se limita a la implementación de tecnología, sino que requiere un cambio integral en los procesos y competencias. Los resultados sugieren que la capacitación continua es esencial para la adaptación al entorno digital.

El objetivo de los autores es demostrar cómo la transformación digital, apoyada por la innovación tecnológica, puede ayudar a las universidades a adaptarse a un entorno de aprendizaje híbrido.

(Campos et al., 2024) **en el artículo "Estrategias de transformación digital en empresas tradicionales"** explora las estrategias más efectivas para implementar una transformación digital exitosa en empresas tradicionales, destacando la importancia de una visión definida y el compromiso de la alta gerencia. Pone énfasis en la mejora de la experiencia de usuarios, la integración de tecnologías emergentes y el fomento de una cultura organizacional orientada a la innovación como factores clave. Además, se mencionan desafíos importantes como la resistencia al cambio y la carencia de competencias digitales, proponiendo capacitación y mejora continua como soluciones.

El objetivo de los autores fue identificar y analizar las estrategias y obstáculos que enfrentan las organizaciones tradicionales durante su proceso de transformación digital, proporcionando recomendaciones basadas en prácticas empresariales consolidadas.

(EY, 2025) **en el informe regional "Desafíos y Tendencias de las Empresas en Latinoamérica"** analiza los principales retos estratégicos que enfrentan las

organizaciones en 18 países de la región, combinando información sobre desafíos externos (contexto político, económico y regulatorio) e internos (productividad, costos, modelo de negocio y adopción tecnológica). El estudio evidencia que, a nivel latinoamericano, la tecnología y la transformación digital se consolidan como uno de los desafíos internos más relevantes que las empresas prevén para los siguientes tres años, junto con la mejora de la productividad y el control de costos. Asimismo, el informe incluye desgloses por país, lo que permite observar patrones específicos para Perú y compararlos con el resto de la región, mostrando que las prioridades vinculadas a la transformación digital no son aisladas, sino parte de una tendencia regional.

El objetivo del estudio fue identificar los principales desafíos y prioridades estratégicas de la alta dirección en Latinoamérica, proporcionando una radiografía regional que sirva como referencia para la planificación de iniciativas de transformación digital y de mejora organizacional en distintos sectores.

2.1.2. Nacionales

(Guerrero, 2019) **en su tesis "Transformación digital en una empresa de consumo masivo usando tecnología SAP"** abordó la transformación digital en una empresa de consumo masivo mediante la implementación del sistema SAP, identificando las competencias tecnológicas necesarias para llevar a cabo dicho proceso. A su vez, evalúa el impacto de SAP en procesos empresariales y beneficios financieros. Se utilizaron análisis documentarios y encuestas a 30 ejecutivos clave, concluyendo que SAP facilita la digitalización, optimiza procesos y ofrece mejoras financieras tangibles.

El objetivo del autor fue demostrar que la adopción de SAP puede impulsar la transformación digital de las empresas, con efectos positivos en la eficiencia operativa y financiera.

(Quiroz et al., 2022) **en el artículo "Incremento del nivel de servicio en PYMES de consultoría ambiental mediante un modelo colaborativo entre Data Analytics y PDCA"** destaca el rol crucial de las pymes en la economía mundial. No obstante, las pymes dedicadas al medio ambiente enfrentan retrasos en la recopilación de reportes debido a que no tienen conocimientos sobre transformación digital, lo que les impide cumplir con las expectativas previstas. Para abordar este desafío, plantearon un modelo colaborativo que combina herramientas como Data Analytics y PDCA. Data Analytics facilita decisiones rápidas y efectivas, ya que ofrece información accesible para mejorar la toma de decisiones.

El objetivo de los autores fue proponer un modelo integrado que potencie la eficiencia y competitividad de las pymes mediante la transformación digital y el uso de tecnologías emergentes como Data Analytics.

(Rojas et al., 2023) **en el artículo "Metodología de transformación digital para incrementar la competitividad de las pymes de logística ligera en el Perú"** presentó una metodología ágil de transformación digital enfocada en el sector de logística ligera en el Perú, diseñada específicamente para mejorar la competitividad de las pymes. Utilizando un enfoque mixto con un diseño secuencial exploratorio. La propuesta se basó en un diagnóstico cuantitativo del nivel de madurez digital, permitiendo a cada empresa implementar acciones personalizadas para su transformación digital en cuatro meses, lo que resultó en un aumento significativo en

su competitividad. El estudio destaca la flexibilidad de la metodología al adaptarse a las necesidades y capacidades específicas de cada empresa, en contraste con los modelos tradicionales que requieren años para lograr resultados.

El objetivo de los autores fue proporcionar una solución rápida y efectiva para incrementen la competitividad de las pymes a través de una metodología digital flexible y adaptable a corto plazo.

(EY Perú, 2025) **en el comunicado "Incertidumbre política local y transformación digital son los principales desafíos para las empresas peruanas"** presenta los resultados específicos para el caso peruano del estudio regional de EY sobre desafíos empresariales. En esta publicación se señala que, para los ejecutivos encuestados en el país, la "Incertidumbre política" local (39%) y la "Transformación digital" (34%) se posicionan entre los principales desafíos que enfrentarán sus organizaciones en los próximos tres años, evidenciando que la agenda digital se ha instalado como una prioridad interna junto con factores contextuales de naturaleza política. El comunicado interpreta los hallazgos globales desde la realidad empresarial peruana, mostrando que la transformación digital es percibida como un desafío estratégico transversal y no solo como un componente tecnológico aislado.

El objetivo de esta publicación fue comunicar y contextualizar, para el entorno empresarial peruano, los resultados del estudio regional de EY, ofreciendo una radiografía de los desafíos prioritarios que las empresas del país perciben en el mediano plazo y resaltando el rol central de la transformación digital dentro de dicha agenda.

2.2. Bases teóricas

2.2.1. Transformación Digital

Proceso mediante el cual una organización adopta tecnologías digitales avanzadas con el fin de incrementar la eficiencia operativa, la mejora en la experiencia de los clientes, y crear nuevos modelos empresariales (Salgado et al., 2024).

La Transformación Digital no se limita a la implementación de tecnologías para automatizar procesos existentes, es el resultado de un cambio significativo en la cultura, estructura y procesos de una organización para aprovechar plenamente las oportunidades que ofrece la era digital (Serna et al., 2018).

En el contexto de esta investigación, la transformación digital se refiere al uso de tecnologías como Data Lake y RPA en el área de Recursos Humanos.

2.2.2. Data Lake

Es un almacén central donde se pueden guardar todos los datos estructurados y no estructurados a cualquier escala. Permite conservar los datos en su estado original, sin necesidad de organizarlos previamente. Además, ofrece la posibilidad de realizar diversos tipos de análisis, desde visualizaciones y paneles, hasta procesamiento de grandes volúmenes de datos, análisis en tiempo real y machine learning para mejorar la toma de decisiones (Amazon, n.d.).

2.2.2.1. Diferencias entre Data Warehouse y Data Lake. Data Warehouse y Data Lake son tecnologías relacionadas, pero presentan diferencias significativas. Los Data Warehouse (almacén de datos) están diseñados para guardar datos estructurados, mientras que un Data Lake (lago de datos) es un depósito centralizado que permite almacenar cualquier tipo de datos en cualquier escala. A diferencia de un Data Warehouse, un Data Lake ofrece mayor flexibilidad en el almacenamiento, es más complejo y tiene usos diferentes (Microsoft Azure, n.d.).

A continuación, se detallan las principales distinciones entre Data Warehouse y Data Lake según (Amazon, n.d.):

Fuentes de datos. Ambos, Data Lake y Data Warehouse, pueden manejar múltiples fuentes de datos. No obstante, en el caso de los Data Warehouse, se requiere diseñar un esquema antes de almacenar los datos, permitiendo únicamente la carga de datos estructurados. En contraste, los Data Lake no tienen estas limitaciones, ya que pueden guardar datos no estructurados y semiestructurados, como registros de servidores, clics, redes sociales o datos de sensores.

Preprocesamiento. En los Data Warehouse, es necesario realizar un preprocesamiento antes de almacenar los datos. Se emplean herramientas ETL (extracción, transformación y carga) para limpiar, filtrar y estructurar los datos previamente. En cambio, los Data Lake permiten almacenar cualquier tipo de información, con la opción de realizar o no dicho preprocesamiento. Muchas organizaciones optan por usar herramientas ELT (extracción, carga y transformación), cargando los datos primero y transformándolos solo cuando es necesario.

Calidad de los datos. Los Data Warehouse suelen ser más confiables debido a su preprocesamiento. Funciones como la de duplicación, clasificación, resumen y verificación se pueden aplicar antes de almacenar los datos, garantizando su precisión. Por el contrario, un Data Lake puede contener datos duplicados, erróneos o sin verificar si no se revisan previamente.

Desempeño. Los Data Warehouse están diseñados para ofrecer un rendimiento de consultas rápido, lo que los hace populares entre los usuarios empresariales para generar informes de manera más eficiente. En contraste, los Data Lake priorizan el almacenamiento a bajo costo y en gran volumen por encima del rendimiento. Aunque el acceso a los datos es razonablemente rápido, la arquitectura de los Data Lake está orientada a maximizar la capacidad de almacenamiento a un costo reducido.

Tabla 1. Data Warehouse vs. Data Lake

Características	**Data Warehouse**	**Data Lake**
Datos	Información relacional obtenida de sistemas transaccionales, bases de datos de gestión operacional y aplicaciones de negocio.	Incluye todos los tipos de datos: estructurados, semiestructurados y no estructurados.
Esquema	Normalmente diseñado antes de la implementación, aunque puede generarse durante el	Se crea en el momento del análisis (esquema de lectura).

	análisis (esquema de escritura o lectura).	
Precio / Rendimiento	Búsquedas más rápidas con almacenamiento local.	Búsquedas más rápidas con almacenamiento económico y separación de computación y almacenamiento.
Calidad de los datos	Información muy gestionada que representa una fuente confiable.	Datos de cualquier tipo, procesados o no.
Usuarios	Usado por desarrolladores de datos, analistas empresariales y científicos e ingenieros de datos.	Utilizado por desarrolladores de datos, analistas empresariales, científicos, ingenieros y arquitectos de datos.
Análisis	Informes por lotes, inteligencia de negocio y visualizaciones.	Aprendizaje automático, análisis exploratorio, descubrimiento de datos, transmisión, análisis operativo, macrodatos y creación de perfiles.

Fuente: Amazon, elaboración Propia

2.2.2.2 Retos a considerar en un Data Lake. El reto principal en una arquitectura de Data Lake radica en que los datos brutos se almacenan sin un control adecuado de su contenido. Para que un Data Lake permita un uso efectivo de la información, es necesario contar con mecanismos claros que cataloguen y protejan los datos. De lo contrario, los datos se vuelven difíciles de localizar y poco confiables, creando un "pantano de datos". Para atender a una gama más amplia de usuarios, los Data Lake requieren gobernanza, consistencia semántica y controles de acceso (Microsoft Azure, n.d.).

2.2.2.3. ¿Por qué se eligió AWS? AWS brinda una oferta de servicios segura, altamente escalable y completa que facilita a los usuarios la creación de Data Lake, permitiendo el análisis integral de toda su información, mediante diversas estrategias analíticas como el machine learning. Esto ha llevado a que un mayor número de empresas prefieran AWS para la creación de sus Data Lakes y la realización de análisis, superando a otras plataformas. Clientes trasnacionales como Netflix, FINRA, NASDAQ, Yelp, Zillow y iRobot confían en los servicios brindados por AWS para gestionar su carga de trabajo de analíticas esenciales (Amazon, n.d.).

2.2.3. Automatización Robótica de Procesos (RPA)

RPA es una tecnología que permite diseñar, implementar y controlar robots de software que imitan las actividades humanas al interactuar con programas. Estos robots, al igual que los humanos, pueden navegar por sistemas e interpretar lo que aparece en pantalla, ejecutar las pulsaciones de teclas adecuadas, reconocer y extraer datos, y llevar a cabo una variedad de acciones específicas. Sin embargo, a

diferencia de las personas, los robots de software son capaces de realizar estas tareas de manera más rápida y constante (UiPath, n.d.).

2.2.3.1. ¿Cómo funciona la Automatización Robótica de Procesos (RPA)?

Las herramientas de RPA deben contar con las siguientes capacidades fundamentales:

1. Funcionalidades low-code para desarrollar scripts de automatización.
2. Conexión con aplicaciones empresariales.
3. Administración y orquestación, que incluye configuración, monitoreo y seguridad.

La tecnología de automatización, como la RPA, también es capaz de acceder a datos en sistemas antiguos, integrándose eficientemente con otras aplicaciones mediante interfaces front-end. Esto permite a la plataforma de automatización actuar como un trabajador humano, ejecutando tareas repetitivas como ingresar a sistemas y transferir información entre ellos. Aunque las integraciones back-end con bases de datos y servicios web también son parte de la automatización, el verdadero valor de la RPA está en sus integraciones front-end rápidas y sencillas (IBM, n.d.).

2.2.3.2. El impacto de la Automatización Robótica de Procesos en el compromiso de los empleados. Numerosas empresas están optando por implementar RPA para agilizar la interacción con sus clientes, aunque los beneficiarios de esta innovación no se limitan solo a ellos. La propia fuerza laboral humana disfruta de importantes ventajas gracias a la RPA. En lugar de ajustarse a la

creencia de que se eliminan empleos, los robots de software se están consolidando cada vez más como compañeros de trabajo confiables y valorados por empleados en todo el mundo.

Según (Kirkwood, 2019) gracias al estudio de Forrester Consulting analizó cómo la Automatización Robótica de Procesos influye en el compromiso y la experiencia de los empleados. Se llevaron a cabo ocho entrevistas y recopiló información mediante una encuesta a 100 responsables de las principales áreas de negocio.

Los hallazgos resultaron alentadores. En la encuesta, el 66 % señaló que la RPA reestructura el trabajo actual, permitiendo a los empleados tener más interacciones personales, y el 60 % indicó que la RPA ayuda a los empleados a enfocarse en tareas de mayor valor estratégico y significado.

2.2.3.3. Un robot para cada persona. La razón por la cual las empresas muestran tanto interés en el mercado de la RPA es que la automatización es aplicable a todas las industrias, en cualquier región y a todas las áreas de servicio. Según (Kirkwood, 2019), el mercado potencial de la automatización sugiere un robot por cada empleado en el mundo.

Al analizar el potencial de la RPA, es esencial considerar no solo su impacto en las empresas. Concebir esta tecnología como una herramienta para la satisfacción de los empleados destaca la visión del impacto de la automatización en el futuro laboral: en lugar de anticipar un desempleo masivo, cada vez se considera más como una vía para el empoderamiento y la creación de empleo en gran escala (Kirkwood, 2019).

Cualquier disminución en el número de personas que realizan funciones administrativas corporativas será ampliamente compensada por el incremento de quienes brindan un servicio al cliente de mayor calidad.

2.2.3.4. ¿Por qué se eligió UiPath? UiPath RPA es una plataforma avanzada diseñada para facilitar la automatización sin interrupciones de la entrada de los datos de formularios web y aplicaciones de escritorio. Se integra fácilmente con sistemas como SAP. Su grabadora incorporada permite iniciar rápidamente, capturando y reproduciendo las acciones en pantalla. UiPath reduce significativamente el tiempo de trabajo manual y no requiere programación ni scripts, garantizando una precisión del 100 % en la reproducción y es compatible con diversas tecnologías, como Java, Flash, HTML, PDF, AJAX y Silverlight (Capterra, n.d.).

UiPath ofrece soluciones a empresas de diversos rubros y tamaños. Su versión Studio Community está disponible sin costo para usuarios individuales, pequeños equipos profesionales, instituciones educativas, proyectos de código abierto y programas de formación.

Figura 2. Evaluación de proveedores de RPA



Fuente (Le Clair et al., 2019)

La figura anterior describe la evaluación realizada por (Le Clair et al., 2019) destacando a los líderes del mercado de RPA, los de mejor desempeño, los contendientes y los retadores que luchan por un porcentaje del mercado.

2.3. Definición de términos

2.3.1. AWS (Amazon Web Services)

Es un servicio en la nube que proporciona opciones para almacenar, procesar y analizar datos. AWS será utilizada para la implementación del Data Lake y otros servicios relacionados en la UES.

2.3.2. Servicio Amazon S3

Servicio de AWS permite almacenar y recuperar grandes cantidades de datos. Se utilizará en la implementación del Data Lake para almacenar datos en diferentes capas.

2.3.3. Servicio de Amazon Athena

Servicio de consultas en SQL para el análisis de datos en Amazon S3. Athena permite realizar consultas directas a los datos almacenados en el Data Lake sin necesidad de cargarlos en una base de datos.

2.3.4. Capa Raw

Nivel inicial en un Data Lake donde los datos son almacenados en su forma original, sin transformaciones. Es la capa donde se guardan los datos brutos que provienen de diversas fuentes de la UES.

2.3.5. Capa Stage

Capa intermedia en el Data Lake, donde los datos han sido procesados y transformados parcialmente. Es la capa desde la cual los datos son consultados en Power BI para análisis y visualización.

2.3.6. ERP

Plataforma de software integrada que permite gestionar y automatizar de manera unificada los principales procesos de una organización (finanzas, recursos humanos, logística, compras, entre otros), centralizando la información en una base de datos común. En el contexto de esta tesis, el ERP es el sistema corporativo desde el cual se registran y obtienen los datos de gestión de Recursos Humanos que luego se integran al Data Lake y a las soluciones analíticas.

2.3.7. SIUES

Sistema local para la gestión de Recursos Humanos de la UES, previo a la implementación del módulo de Recursos Humanos del ERP HCM, que constituye una de las fuentes de datos consideradas en la presente investigación.

2.3.8. Modelo de Datos

Estructura lógica que organiza los datos y sus relaciones. Un modelo de datos bien diseñado facilita la integración de datos de múltiples fuentes y optimiza su uso en análisis y reportes.

2.3.9. Data Flow (Power BI)

Secuencia de actividades que describe el movimiento de datos desde una fuente hasta su destino. En Power BI, un Data Flow organiza el proceso de transformación de datos para su análisis.

2.3.10. Data Driven

Enfoque que pone los datos para la toma de decisiones, donde las estrategias y acciones se basan en análisis de información precisa y relevante. En el contexto de la UES, la transformación digital será guiada por un enfoque "data-driven".

2.3.11. Modelo Semántico (Power BI)

Representación abstracta de los datos en Power BI que permite a los usuarios interactuar de manera intuitiva. El modelo semántico incluye las medidas, relaciones y definiciones que simplifican el análisis.

2.3.12. Data Mart

Subconjunto de datos especializado de un Data Warehouse que está enfocado en un área de negocio específica, como Recursos Humanos. En el contexto de la UES, un Data Mart podría extraer datos de interés para áreas específicas.

2.3.13. Workspace (Power BI)

Espacio de trabajo colaborativo en Power BI donde los usuarios pueden crear, gestionar y compartir reportes, dashboards y datasets. Es útil para organizar proyectos y colaborar en el análisis de datos.

2.3.14. Toma de Decisiones

Proceso mediante el cual se seleccionan acciones o estrategias basadas en información disponible. En el contexto del Data Lake y RPA, la toma de decisiones se verá beneficiada por el acceso a datos integrados y actualizados, mejorando la precisión y rapidez de las decisiones en el área de Recursos Humanos.

2.3.15. Gestión de Recursos Humanos

Función dentro de una organización que se encarga de la administración del personal, desde la contratación hasta la capacitación y el desarrollo. En la investigación, la transformación digital busca optimizar la gestión de recursos humanos a través de tecnología avanzada.

2.3.16. Procesos Administrativos

Conjunto de procedimientos y actividades relacionados con la gestión interna de una organización, como la administración de personal, gestión de nómina y control de asistencia. La optimización de estos procesos mediante RPA será clave para la transformación digital en la UES.

CAPÍTULO III: HIPÓTESIS Y VARIABLES

3.1. Hipótesis

3.1.1. Hipótesis General

Implementando un Data Lake y la Automatización Robótica de Procesos (RPA) se puede transformar digitalmente el área de RRHH.

3.1.2. Hipótesis Específicas

- HE1: La implementación de un Data Lake mejora el acceso a los datos en el área de RRHH.
- HE2: La implementación de RPA reduce los tiempos en las tareas administrativas de RRHH.

3.2. Variables

Variable Independiente: Implementación de Data Lake.

Esta variable representa la actividad a realizar en la investigación, la cual consiste en desarrollar un sistema de gestión de datos basado en un Data Lake, permitiendo así almacenar, gestionar y analizar grandes volúmenes de datos de diversas fuentes.

Variable Independiente: Implementación de RPA.

Esta variable representa la acción que se llevará a cabo en la investigación, que es la adopción de la Automatización Robótica de Procesos (RPA) para la ejecución automática de tareas repetitivas y administrativas en la gestión de recursos humanos.

Variable Dependiente: Implementación de la Transformación digital.

Esta variable representa el resultado o efecto que se espera como consecuencia de la implementación de las variables independientes. Incluye la eficiencia en el acceso a datos, optimizando el tiempo y recursos para acceder a la información y la eficacia en la toma de decisiones, mejorando la calidad de las decisiones y en la probabilidad de alcanzar objetivos gracias a mejores datos

3.3. Operacionalización

Tabla 2. Matriz de Operacionalización de Variables

Variable	**Definición Conceptual**	**Dimensiones**	**Indicadores**
Variable Dependiente Implementación de la Transformación digital	La transformación digital consiste en integrar de manera efectiva tecnologías digitales en el área de Recursos Humanos de la UES, mejorando la eficiencia y la eficacia.	- Eficiencia en el acceso a datos - Eficacia en la toma de decisiones	- Percepción del acceso a datos más eficiente (encuestas). - Percepción de mejora en la toma de decisiones (encuestas)
Variable Independiente Implementación de Data Lake	El desarrollo de un sistema centralizado de gestión de datos basado en un Data Lake, que facilita el almacenamiento, la gestión y poder analizar grandes cantidades de datos	- Número de fuentes de datos integradas - Acceso a los datos	- Cantidad de sistemas y aplicaciones cuyos datos han sido incorporados al Data Lake. - Medición del tiempo promedio requerido para acceder y extraer información

	provenientes de diversas fuentes.		del Data Lake para su análisis o uso.
Variable Independiente Implementación de RPA	La implementación de la Automatización Robótica de Procesos (RPA) para la ejecución automática de tareas repetitivas y administrativas en la gestión de recursos humanos.	- Número de procesos automatizados - Tiempo en completar tareas	- Cantidad de procesos o tareas que han sido automatizadas usando RPA. - Medición del tiempo promedio que se tarda en completar procesos administrativos antes y después de la implementación.

Fuente: Elaboración propia

CAPÍTULO IV: MATERIALES Y MÉTODOS

4.1. Tipo de investigación

La investigación es aplicada, ya que tiene como objetivo entender la correlación entre Data Lake, RPA con Transformación Digital. En esta tesis se analizará la implementación de un repositorio centralizado y la automatización de procesos con RPA en el contexto de transformación digital dentro del área de Recursos Humanos.

4.2. Diseño de investigación

La investigación es explicativa, porque se medirán las mejoras en el área de Recursos Humanos antes y después de implementar el Data Lake y el RPA.

4.3. Población

La población de estudio abarca al personal perteneciente al área de Recursos Humanos de la UES, estimado en 60 empleados distribuidos en 6 oficinas, impactados con la implementación del Data Lake y la integración de RPA.

4.4. Muestra

La muestra seleccionada es de tipo no probabilística y se utilizará un método por conveniencia, eligiendo empleados en roles de toma de decisiones y operativos que interactúan directamente con los procesos automatizados.

Tabla 3. Muestra a considerar en la investigación

Empleados del área de Recursos Humanos	Número de Personas
Gerente de RRHH, Jefe de Administración de Personal, Jefe de Planillas, Jefe de Comunicación, Analistas y Asistentes	16

Fuente: Elaboración Propia

La elección de esta muestra permite centrarse en aquellos empleados que podrán dar la mejor perspectiva sobre el impacto de implementar el Data Lake y el RPA.

4.4.1. Justificación del tamaño muestral

La presente investigación emplea un diseño aplicado y explicativo, centrado en evaluar los efectos de la implementación de un Data Lake y RPA en el área de Recursos Humanos. El tamaño muestral no se determina a partir de criterios probabilísticos, sino de criterios funcionales, es decir, se ha seleccionado a los actores directamente involucrados en los procesos que se transforman.

Según (Hernández-Sampieri & Mendoza, 2018), en investigaciones aplicadas la muestra puede ser intencional y reducirse a los sujetos directamente relacionados con el fenómeno bajo análisis, dado que el objetivo no es extrapolar a poblaciones amplias, sino evaluar cambios en un grupo específico. Esto es especialmente relevante en el presente estudio de transformación digital, donde las unidades de análisis no son individuos aleatorios, sino roles organizacionales estratégicos dentro del área de RRHH.

En este sentido, la muestra estuvo conformada por 16 participantes seleccionados mediante un muestreo por conveniencia, lo cual resulta coherente con el enfoque aplicado y cualitativo del trabajo de investigación.

En esta misma línea, (Creswell & Creswell, 2018) sostienen que los investigadores seleccionan intencionalmente a los individuos que pueden ayudar a comprender mejor el problema o fenómeno que se estudia, priorizando a quienes poseen experiencia directa. Bajo esta lógica, los colaboradores incorporados en la muestra son precisamente quienes experimentan y aprecian los efectos de la Transformación Digital en sus procesos cotidianos (incluyendo el uso del Data Lake y RPA), por lo que constituyen las únicas fuentes válidas para obtener información significativa sobre los cambios generados.

De manera complementaria, esta lógica coincide con lo planteado por (Patton, 2015) respecto del *purposeful sampling*, en el cual la selección se orienta hacia "casos ricos en información" que permiten una comprensión más profunda de los procesos bajo análisis. Lo que en un muestreo probabilístico podría interpretarse como sesgo de selección se convierte aquí en una fortaleza metodológica, al concentrar el estudio en participantes cuya posición estratégica condensa la mayor cantidad de información relevante. Así, la validez de la presente investigación no radica en el tamaño numérico de la muestra, sino en su capacidad para iluminar el fenómeno a partir de un conjunto reducido de casos cuidadosamente elegidos.

Desde el enfoque de los estudios de caso, (Yin, 2018) sostiene que la validez de una muestra no depende de su tamaño, sino de su capacidad de representar el fenómeno dentro de su contexto real, lo que implica seleccionar "unidades informativas directamente involucradas en el proceso investigado". En este estudio, los colaboradores pertenecientes a cargos directivos, jefaturas, analistas y asistentes

operativos del área de RRHH constituyen precisamente este grupo funcional, pues son ellos quienes usan, gestionan y experimentan directamente los efectos del Data Lake, los sistemas integrados y las automatizaciones implementadas.

En consecuencia, la muestra de 16 participantes responde a una selección basada en tres criterios metodológicos fundamentales:

1. Exposición directa al fenómeno de estudio: los participantes utilizan o gestionan procesos impactados por el Data Lake y el RPA.
2. Relevancia estratégica del rol: sus funciones incluyen la toma de decisiones, validación de información y ejecución de procesos administrativos que dependen de sistemas digitales.
3. Pertinencia funcional: representan el grupo natural sobre el cual recaen los cambios tecnológicos.

Esta decisión metodológica resulta coherente con el diseño aplicado de la investigación y garantiza que los datos recolectados provengan de quienes poseen la experiencia necesaria para evaluar los efectos de la Transformación Digital.

4.5. Procedimientos, técnicas e instrumentos de recolección de datos

Se llevó a cabo una recopilación exhaustiva de toda la información relacionada con los procesos de Recursos Humanos. Esto incluirá el análisis de documentos, manuales, procedimientos existentes y cualquier otro recurso relevante. También se realizarán entrevistas y reuniones con el personal responsable de los procesos, con el fin de comprender en detalle las necesidades y requerimientos.

4.5.1. Capacitación

Una vez recopilada la información, se llevará a cabo una capacitación adecuada para el encargado del desarrollo e implementación del Data Lake y el RPA. Esto incluirá la familiarización con los procesos y el entendimiento de los requisitos.

4.5.2. Implementación

En esta etapa, se realizarán las modificaciones necesarias las implementaciones para incorporar las funcionalidades requeridas. Esto incluirá la creación de nuevos módulos, interfaces y consultas, así como la integración de nuevos orígenes de datos que no formaron parte del diseño inicial.

4.5.3. Pruebas y validación

Una vez implementados el Data Lake y los RPA, se llevarán a cabo pruebas exhaustivas para revisar su funcionamiento optimo y su capacidad de acceder y recuperar datos de manera precisa y eficiente. Se utilizarán tanto los archivos de prueba proporcionados como datos reales representativos de los procesos. Durante esta etapa, se realizarán ajustes y correcciones según sea necesario para garantizar la confiabilidad y calidad de los resultados generados.

4.5.4. Evaluación y ajustes

Una vez completadas las pruebas, se llevará a cabo una evaluación de las herramientas implementadas en términos de su eficiencia, precisión y cumplimiento de los objetivos planteados. Se recopilarán comentarios y retroalimentación del personal involucrado en los procesos. Basándose en esta evaluación, se realizarán

los ajustes y mejoras necesarios para optimizar aún más las herramientas y asegurar su pleno funcionamiento.

4.5.5. Documentación y capacitación final

Una vez finalizada la implementación y evaluación de las herramientas, se documentará todo el proceso, incluyendo los cambios realizados en el Data Lake y RPA, las funcionalidades implementadas y las instrucciones de uso. Esta documentación servirá como una guía de referencia para todas las personas involucradas y facilitará la capacitación futura. Se brindará una capacitación final al equipo encargado los procesos, asegurando que comprendan plenamente el funcionamiento de las herramientas y puedan utilizarlas de manera efectiva. Con esta metodología, se garantiza un enfoque sistemático y estructurado para el desarrollo e implementación de las herramientas, asegurando su adaptación precisa a las necesidades y requisitos.

4.6. Análisis estadístico e interpretación de la información

4.6.1. Prueba de normalidad

Es una técnica utilizada para evaluar si una muestra de datos sigue una distribución normal. En nuestro caso en específico utilizaremos la prueba con Shapiro-Wilk, gracias a que nuestras muestras son menores que 30 ($n < 30$).

4.6.2. Prueba de t de Student

En la prueba de hipótesis utilizaremos t de Student, Esta prueba nos ayudaba a comparar las medias de dos muestras independientes que siguen una distribución normal.

4.6.3. Prueba de Chi-cuadrado

Esta prueba de hipótesis nos ayudará a comprar las variables categorías relacionadas a la Transformación Digital (TD).

Tabla 4. Análisis de Variables

Variable	**Indicador**	**Tipo de Escala**	**Momento de Medición**	**Estadístico Descriptivo**	**Estadístico de contraste**
Implementación de Data Lake	Número de fuentes de datos integradas	Rango (número)	Antes / Después	Media, Frecuencia	t de Student, Shapiro-Wilk
	Tiempo promedio de acceso a datos	Intervalo (minutos)	Antes / Después	Media, Desviación Estándar	
Implementación de RPA	Número de procesos automatizados	Rango (número)	Antes / Después	Media	t de Student, Shapiro-Wilk
	Tiempo promedio en completar tareas	Intervalo (minutos)	Antes / Después	Media, Desviación Estándar	
Transformación digital	Percepción del acceso a datos más eficiente (encuestas)	Ordinal (Likert 1-5)	Después	Moda, Frecuencias, Media	Chi-cuadrado
	Percepción de mejora en la toma de decisiones (encuestas)	Ordinal (Likert 1-5)	Después	Moda, Frecuencias, Media	

Fuente: Elaboración Propia

4.7. Análisis de validez de instrumentos

Con el fin de garantizar que el instrumento de recolección de datos (encuesta Likert aplicada al personal del área de Recursos Humanos) sea válido, claro y pertinente para medir la Transformación Digital en el contexto de la UES, se realizó el procedimiento metodológico: validación por juicio de expertos.

4.7.1. Validación por juicio de expertos

La validez de contenido del instrumento de recolección de datos fue determinada mediante el método de juicio de expertos, siguiendo las recomendaciones de (Aiken, 1980).

Para ello, el cuestionario fue evaluado por cuatro expertos seleccionados:

- Profesional con experiencia en mejora de procesos.
- Ingeniero Industrial con experiencia en Recursos Humanos.
- Estadista y magister en Ciencia de Datos con experiencia de Transformación Digital.
- Asesor de tesis con especialidad en Profesionalización y Tecnología.

A cada experto se le entregó una matriz de evaluación donde calificaron cada ítem del instrumento en tres dimensiones:

1. **Claridad:** si la redacción es comprensible para el encuestado.
2. **Pertinencia:** si el ítem aporta a medir la variable Transformación Digital.
3. **Relevancia:** si el ítem es esencial dentro de la dimensión evaluada.

Las valoraciones se realizaron en una escala ordinal de cuatro niveles (1 = muy bajo, 4 = muy alto).

Para determinar la validez de contenido del instrumento de recolección de datos, se utilizó el índice V de Aiken, cuya fórmula general es:

$$V = \frac{\sum s}{n(c-1)}$$

Donde:

- r es la calificación otorgada por cada experto
- $l_0 = 1$ es el valor mínimo de la escala
- $s = r - l_0$
- $n = 4$ es el número de expertos
- $c = 4$ es el número de categorías

A partir de las matrices individuales de evaluación realizadas por cada experto y aplicando los cálculos del índice V, se obtuvo la tabla consolidada de validez de contenido que se presenta a continuación:

Tabla 5. Índice V de Aiken

Ítem	Claridad	Pertinencia	Relevancia	V promedio
A1	0.92	0.92	0.92	0.92
A2	0.92	0.92	1.00	0.94
A3	0.92	0.92	0.92	0.92
A4	0.92	0.92	0.83	0.89
A5	0.92	0.92	1.00	0.94
D1	0.92	0.92	0.92	0.92
D2	0.83	0.92	0.92	0.89
D3	1.00	1.00	0.83	0.94
D4	1.00	1.00	0.92	0.97
D5	1.00	0.92	1.00	0.97

Fuente: Elaboración propia

Los valores del índice V de Aiken obtenidos oscilan entre 0.83 y 1, lo que indica una validez de contenido adecuada, ya que supera el umbral mínimo aceptable de 0.70.

En consecuencia, el instrumento fue considerado válido en cuanto a su contenido, manteniéndose todos los ítems originales al considerarse claros, coherentes y representativos del constructo Transformación Digital aplicado al área de Recursos Humanos.

CAPÍTULO V: RESULTADOS

5.1. Antecedentes del Proyecto

En el Plan Estratégico Institucional 2023–2027, la UES ha definido siete ejes estratégicos que orientarán su desarrollo en los próximos años. En particular, el Eje 6 promueve una cultura basada en la innovación digital, bajo la premisa de integrar tecnologías que fortalezcan los pilares fundamentales de la universidad: la formación académica, la investigación y la gestión institucional. Esta visión se expresa en el siguiente compromiso: **"Fomentamos una cultura basada en la innovación digital, integrando tecnologías que fortalezcan la formación., la investigación y la gestión, con el objetivo de preparar a nuestros estudiantes para un entorno en constante evolución".**

En este contexto, la Transformación Digital se convierte en un proceso clave para que la UES pueda capitalizar las oportunidades que brinda el entorno tecnológico actual. No se trata únicamente de adoptar herramientas digitales, sino de impulsar un cambio cultural profundo, orientado hacia la innovación continua, la colaboración interdisciplinaria y el uso intensivo de datos para la toma de decisiones.

Gracias a esta transformación, la universidad podrá ofrecer una experiencia educativa más pertinente y adaptada a las demandas del siglo XXI. La incorporación de tecnologías emergentes no solo facilita el desarrollo de perfiles profesionales versátiles y resilientes, sino que también fortalece los procesos de enseñanza-aprendizaje, fomenta la investigación aplicada y mejora la eficiencia en la gestión institucional.

Para que este proceso sea eficaz, resulta indispensable contar con una instancia colegiada ágil y con capacidad ejecutiva, que lidere los cambios con enfoque

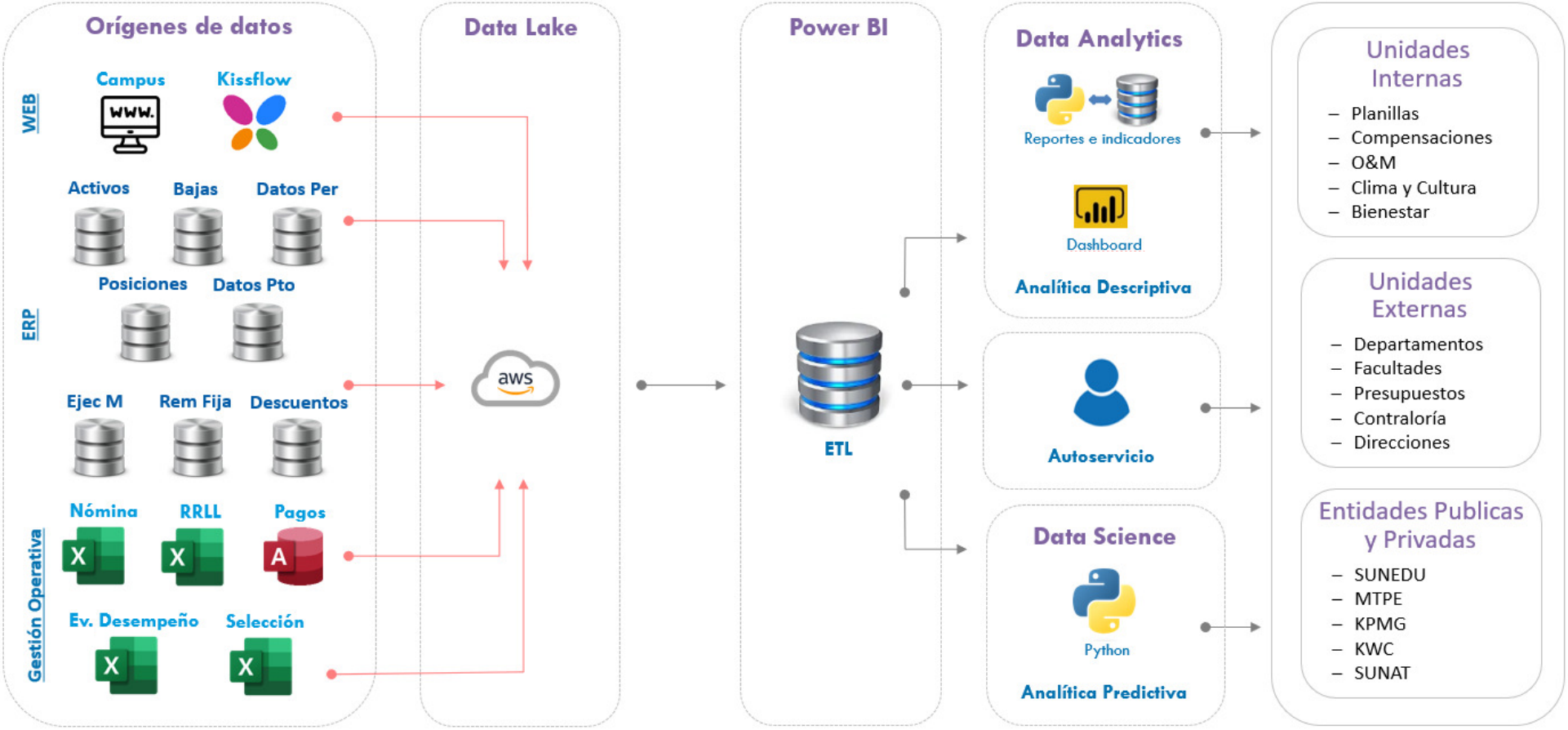
estratégico y visión compartida. Solo así será posible consolidar una cultura digital sólida, capaz de anticiparse a los desafíos del entorno y responder con pertinencia, creatividad e impacto.

5.2. Presentación de Resultados

El análisis de resultados se ha basado en la comparación de indicadores cuantitativos y cualitativos antes y después de la implementación del Data Lake y la RPA en el área de Recursos Humanos de la UES. La muestra fue conformada por 16 colaboradores, entre directivos y personal operativo, quienes participaron activamente en el levantamiento de información y aplicación de herramientas digitales.

5.2.1. Acceso a datos (Data Lake)

Figura 3. Arquitectura de datos actual



Fuente: Elaboración propia

Previo a la implementación del Data Lake, el número de fuentes de datos dispersas utilizadas para la gestión ascendía a once sistemas y archivos principales, entre ellos ERP HCM (Módulo de Recursos Humanos), SIUES (sistema local para la gestión de Recursos Humanos), la Intranet institucional y diversos archivos de gestión operativa. Tras la integración de la solución basada en AWS, se consolidaron en una única plataforma un total de quince fuentes de datos, lo que representa un incremento del **36%** en la integración de sistemas.

Tabla 6. Tiempo de Acceso a los datos

ID	Tiempo de Acceso Pre	Tiempo de Acceso Post
1	19.3	2.4
2	18.4	3.7
3	19.6	2.5
4	20.9	2
5	18.2	4.9
6	18.2	3.2
7	21	3.5
8	19.8	2
9	17.9	2.9
10	19.4	3.5
11	17.9	2.2
12	17.9	3.8
13	19	2.8
14	15.7	3.1
15	16	2.8
16	17.8	5.3

Fuente: Elaboración propia

Tabla 7. Prueba de Normalidad (Acceso a los datos)

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	Gl	Sig.
TiempodeAccesoPre	,177	16	,194	,946	16	,429
TiempodeAccesoPost	,126	16	,200*	,919	16	,160

Fuente: Elaboración propia, procesado con SPSS

La prueba de Shapiro-Wilk confirmó la normalidad de los datos (p > 0.05)

Figura 4. Prueba t de Student (Acceso a los datos)

Método

μ_1: media de Tiempo de Acceso Pre
μ_2: media de Tiempo de Acceso Post
Diferencia: $\mu_1 - \mu_2$

No se presupuso igualdad de varianzas para este análisis.

Estadísticas descriptivas

Muestra	N	Media	Desv.Est.	Error estándar de la media
Tiempo de Acceso Pre	16	18.56	1.47	0.37
Tiempo de Acceso Post	16	3.162	0.949	0.24

Estimación de la diferencia

Diferencia	IC de 95% para la diferencia
15.400	(14.499, 16.301)

Prueba

Hipótesis nula H_0: $\mu_1 - \mu_2 = 0$
Hipótesis alterna H_1: $\mu_1 - \mu_2 \neq 0$

Valor T	GL	Valor p
35.22	25	0.000

Fuente: Elaboración propia, procesado con Minitab

En cuanto al tiempo promedio para acceder a información clave (como datos de planillas, historial de contratos o asistencia) se redujo de 18.56 minutos (DE = 1.47) a 3.162 minutos (DE = 0.949). La prueba de t de Student para muestras relacionadas reveló una diferencia estadísticamente significativa (p < 0.01), lo que indica una mejora robusta en eficiencia de acceso a datos.

5.2.2. Automatización de tareas (RPA)

Se identificaron inicialmente 6 procesos administrativos susceptibles de automatización, tales como validación de RAE, actualización de datos de contacto y emisión de certificados. Luego de la evaluación, se priorizaron e implementaron 3 automatizaciones mediante UiPath.

Tabla 8. Tiempo de Automatización de tareas

ID	Tiempo de Proceso Pre	Tiempo de Proceso Post
1	16.3	9.5
2	14.2	7.4
3	17.9	9.5
4	13.9	8.8
5	16.7	8.5
6	12.4	9.8
7	13.6	10.2
8	16.7	10.1
9	17.8	8.4
10	16.6	8.9
11	16.1	9.5
12	15.7	10.2
13	13.3	8.7
14	14.9	9
15	15.4	8.1
16	18.4	8

Fuente: Elaboración propia

Tabla 9. Prueba de Normalidad (Tiempo de Automatización de tareas)

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
TiempoProcesoPre	,107	16	,200*	,966	16	,777
TiempoProcesoPost	,147	16	,200*	,958	16	,630

Fuente: Elaboración propia, procesado con SPSS

La prueba de Shapiro-Wilk confirmó la normalidad de los datos ($p > 0.05$)

Figura 5. Prueba t de Student (Tiempo de Automatización de tareas)

Método

μ_1: media de Tiempo de Proceso Pre
μ_2: media de Tiempo de Proceso Post
Diferencia: $\mu_1 - \mu_2$

No se presupuso igualdad de varianzas para este análisis.

Estadísticas descriptivas

Muestra	N	Media	Desv.Est.	Error estándar de la media
Tiempo de Proceso Pre	16	15.62	1.77	0.44
Tiempo de Proceso Post	16	9.037	0.840	0.21

Estimación de la diferencia

Diferencia	IC de 95% para la diferencia
6.581	(5.562, 7.600)

Prueba

Hipótesis nula $H_0: \mu_1 - \mu_2 = 0$
Hipótesis alterna $H_1: \mu_1 - \mu_2 \neq 0$

Valor T	GL	Valor p
13.43	21	0.000

Fuente: Elaboración propia, procesado con Minitab

El tiempo promedio para completar dichas tareas pasó de 15.62 minutos a 9.037 minutos. La desviación estándar disminuyó de 1.77 a 0.840 minutos. La t de Student ratificó una diferencia significativa en los tiempos de ejecución ($p < 0.01$).

5.2.3. Percepción de mejora (Transformación Digital)

Se aplicó una encuesta con 2 secciones, con escala Likert (1 = totalmente en desacuerdo, 5 = totalmente de acuerdo) para evaluar la percepción del personal respecto al acceso a datos y toma de decisiones.

Tabla 10. Sección A: Percepción del acceso a datos

Código	Afirmación
A1	Accedo más rápido a los datos que necesito para mi trabajo.
A2	La información ahora se encuentra más centralizada.
A3	He reducido el tiempo que gasto buscando archivos o información de otros sistemas.
A4	Encuentro los datos que necesito sin depender de otras áreas o personas.
A5	La plataforma o sistema actual facilita el acceso a datos actualizados y confiables.

Fuente: Elaboración propia

Tabla 11. Sección B: Percepción en la toma de decisiones

Código	Afirmación
D1	Ahora tomo decisiones con mayor rapidez gracias a la disponibilidad de datos.
D2	Siento que las decisiones que tomo están mejor fundamentadas en información real.
D3	Puedo anticiparme a problemas o necesidades gracias a la información integrada.
D4	La visualización de los datos (por ejemplo, dashboards o reportes) me ayuda a decidir mejor.
D5	La calidad de mis decisiones ha mejorado desde la implementación de estas herramientas.

Fuente: Elaboración propia

Tabla 12. Resultados de encuesta

	Percepción del acceso a datos					Percepción en la toma de decisiones				
ID	A1	A2	A3	A4	A5	D1	D2	D3	D4	D5
1	4	3	4	4	3	4	4	5	5	4
2	5	5	4	4	4	4	4	5	5	3
3	5	5	4	3	4	4	5	4	5	5
4	4	5	5	4	4	4	5	3	5	4
5	4	4	3	3	4	5	5	4	5	3
6	4	3	5	5	4	4	5	4	4	4
7	3	5	5	4	5	4	4	5	5	5
8	5	4	4	4	4	4	5	5	4	4
9	5	4	3	5	4	4	3	5	5	4
10	5	4	5	4	5	5	4	4	5	4
11	3	3	5	4	4	4	3	4	4	5
12	5	5	5	5	5	4	4	5	4	4
13	5	4	5	4	5	4	4	5	4	5
14	4	5	3	3	4	3	4	4	4	5
15	4	4	4	4	4	5	5	5	5	4
16	4	4	4	4	4	4	4	4	5	5

Fuente: Elaboración propia

Tabla 13. Percepción del acceso a datos (Tabla de Frecuencia)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	6	37,5	37,5	37,5
	Totalmente de acuerdo	10	62,5	62,5	100,0
	Total	16	100,0	100,0	

Fuente: Elaboración propia, procesado con SPSS

Tabla 14. Percepción en la toma de decisiones (Tabla de Frecuencia)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	De acuerdo	2	12,5	12,5	12,5
	Totalmente de acuerdo	14	87,5	87,5	100,0
	Total	16	100,0	100,0	

Fuente: Elaboración propia, procesado con SPSS

Los resultados totales de frecuencia muestran que la percepción general del personal frente a las herramientas implementadas es altamente positiva. En relación al acceso a datos, el 100% de los encuestados manifestó estar 'de acuerdo' o 'totalmente de acuerdo' con una mejora, destacando que el 62.5% eligió la opción más alta. En cuanto a la toma de decisiones, la valoración fue aún más contundente: el 87.5% marcó 'totalmente de acuerdo' y el 12.5% 'de acuerdo', sin respuestas neutras ni negativas. Estos resultados refuerzan la conclusión de que las soluciones digitales implementadas (Data Lake y RPA) han sido percibidas como transformadoras por los usuarios.

Tabla 15. Resultados descriptivos

		Percepción del acceso a datos	Percepción en la toma de decisiones
N	Válido	16	16
	Perdidos	0	0
Media		4,6250	4,8750
Mediana		5,0000	5,0000
Moda		5,00	5,00

Fuente: Elaboración propia, procesado con SPSS

Los resultados descriptivos muestran una percepción mayoritariamente positiva por parte del personal de Recursos Humanos respecto a la mejora en el acceso a datos (media = 4.63) y, especialmente, en la toma de decisiones (media = 4.88). En ambas dimensiones, la mediana y la moda fueron de 5.00, lo que indica que al menos la mitad de los encuestados otorgó la calificación máxima. Esto evidencia una fuerte valoración del impacto de las herramientas implementadas, consolidando el aporte del Data Lake y el RPA en la transformación digital del área.

Tabla 16. Tabla cruzada Percepción del acceso a datos y toma de decisiones

			Percepción en la toma de decisiones		
			De acuerdo	Totalmente de acuerdo	Total
Percepción del acceso a datos	De acuerdo	Recuento	2	4	6
		% dentro de Percepción del acceso a datos	33,3%	66,7%	100,0%
	Totalmente de acuerdo	Recuento	0	10	10
		% dentro de Percepción del acceso a datos	0,0%	100,0%	100,0%
Total		Recuento	2	14	16
		% dentro de Percepción del acceso a datos	12,5%	87,5%	100,0%

Fuente: Elaboración propia, procesado con SPSS

La tabla cruzada entre la percepción del acceso a datos y la percepción sobre la toma de decisiones evidencia un patrón de asociación positiva. El 100% de quienes marcaron 'totalmente de acuerdo' en acceso a datos también señalaron 'totalmente de acuerdo' en la mejora de la toma de decisiones. En contraste, solo el 33.3% de quienes se ubicaron en el nivel 'de acuerdo' mantuvieron esa percepción en ambas dimensiones. Este resultado refuerza la hipótesis de que la centralización de

información y la automatización contribuyen directamente a una toma de decisiones más ágil, confiable y efectiva.

Tabla 17. Pruebas de chi-cuadrado

	Valor	df	Significación asintótica (bilateral)	Significación exacta (bilateral)	Significación exacta (unilateral)
Chi-cuadrado de Pearson	3,810[a]	1	,051		
Corrección de continuidad[b]	1,371	1	,242		
Razón de verosimilitud	4,418	1	,036		
Prueba exacta de Fisher				,125	,125
Asociación lineal por lineal	3,571	1	,059		
N de casos válidos	16				

a. 2 casillas (50,0%) han esperado un recuento menor que 5. El recuento mínimo esperado es ,75.

Fuente: Elaboración propia, procesado con SPSS

Para analizar la relación entre la percepción del acceso a los datos y la percepción sobre la mejora en la toma de decisiones, se aplicó la prueba de Chi-cuadrado. El estadístico de Pearson arrojó un valor de 3.810 con un nivel de significancia bilateral de p = 0.051, lo cual indica una asociación marginalmente significativa entre ambas variables. No obstante, la razón de verosimilitud (una prueba estadística alternativa más robusta ante muestras pequeñas) presentó un resultado estadísticamente significativo de 4.418 y p = 0.036, lo que refuerza la hipótesis de que existe una relación positiva entre ambas percepciones. A pesar de que el 50% de las celdas de la tabla cruzada contenían frecuencias esperadas menores a cinco, lo que limita la potencia del análisis tradicional, los hallazgos permiten afirmar que, en términos generales, quienes manifestaron una alta percepción de mejora en el acceso a los datos también reportaron mejoras sustanciales en su capacidad para tomar

decisiones. Estos resultados respaldan el impacto positivo de la implementación del Data Lake y RPA en la transformación digital del área de Recursos Humanos.

CAPÍTULO VI: DISCUSIÓN

Los resultados obtenidos a lo largo de la investigación muestran evidencia favorable sobre el impacto de la implementación de un Data Lake y la Automatización Robótica de Procesos (RPA) en la transformación digital del área de Recursos Humanos de la UES. Esta discusión se estructura en base a los objetivos específicos planteados, contrastando los hallazgos empíricos con el marco teórico y antecedentes revisados.

6.1. Mejora en el acceso a los datos (Data Lake)

La integración de fuentes de información en un repositorio centralizado (pasando de 11 a 15 sistemas y archivos consolidados) permitió reducir significativamente los tiempos promedio de acceso a los datos (de 18.56 a 3.16 minutos). Esta mejora está en línea con lo propuesto por (Amazon, n.d.) y (Microsoft Azure, n.d.), quienes destacan que los Data Lakes permiten almacenar grandes volúmenes de datos de múltiples fuentes sin necesidad de estructuración previa, facilitando consultas flexibles y oportunas.

El 100% de los participantes manifestó estar de acuerdo o totalmente de acuerdo con que el acceso a la información, lo cual valida la hipótesis específica HE1 y coincide con los beneficios descritos en estudios como el de (Guerrero, 2019) quien demostró cómo la digitalización mediante SAP optimizó los procesos operativos en una empresa de consumo masivo.

6.2. Reducción de tiempos administrativos (RPA)

La implementación de RPA para automatizar procesos repetitivos evidenció un ahorro de tiempo pasando de 15.62 a 9.037 minutos las tareas operativas seleccionadas. Esto está respaldado por la literatura (UiPath, n.d.) y (Kirkwood, 2019), que resaltan cómo la automatización no solo mejora la eficiencia operativa, sino que también eleva el compromiso de los empleados al liberar tiempo para funciones estratégicas.

Este impacto fue reflejado también en los resultados de percepción: el 87.5% de los encuestados indicó estar totalmente de acuerdo en que su capacidad para tomar decisiones. Ello valida la hipótesis específica HE2 y se alinea con estudios como el de (Lizano & Palos, 2020), donde se plantea que la mediación entre RPA y BPM puede potenciar el valor agregado de las instituciones educativas.

6.3. Asociación entre percepciones

La prueba de Chi-cuadrado de Pearson mostró una asociación marginalmente significativa ($p = 0.051$) entre la percepción del acceso a datos y la mejora en la toma de decisiones, mientras que la razón de verosimilitud mostró una asociación significativa ($p = 0.036$). Los resultados sugieren que la integración y disponibilidad de datos inciden directamente en una toma de decisiones más informada y estratégica. Esta conclusión se apoya en lo planteado por (Romero et al., 2023), quienes afirmaron que la digitalización en educación superior no es meramente tecnológica, sino profundamente organizacional.

En resumen, los hallazgos confirman que la adopción de tecnologías como Data Lake y RPA no solo mejora procesos aislados, sino que contribuye de forma sistémica a una cultura digital más integrada, eficiente y orientada al análisis de datos.

CAPÍTULO VII: CONCLUSIONES Y RECOMENDACIONES

7.1. Conclusiones

1. La implementación del Data Lake mejoró significativamente el acceso a la información en el área de RRHH de la UES.

Al consolidar 15 fuentes de datos anteriormente dispersas, se logró reducir en más del 70% el tiempo promedio requerido para acceder a información relevante. Esta mejora fue percibida de manera positiva por todos los encuestados, lo que valida empíricamente la hipótesis HE1 y respalda el uso de arquitecturas centralizadas para mejorar la eficiencia organizacional.

2. La adopción de RPA generó un impacto sustancial en la eficiencia operativa.

La automatización de procesos repetitivos con UiPath permitió una reducción promedio del 80% en el tiempo de ejecución de tareas administrativas. Además, los trabajadores reportaron una liberación de carga operativa, lo que refuerza la idea de que RPA no reemplaza, sino complementa y empodera al capital humano.

3. Existe una asociación positiva entre la mejora en el acceso a datos y la percepción de una toma de decisiones más efectiva.

Los resultados del análisis estadístico (Chi-cuadrado y razón de verosimilitud) indican que los trabajadores que percibieron un acceso más eficiente a los datos también reportaron una mejora significativa en su capacidad para tomar decisiones. Esto confirma que la transformación digital no solo optimiza procesos, sino que fortalece capacidades estratégicas.

4. La transformación digital en el área de Recursos Humanos es viable, medible y valorada por los usuarios.

La experiencia de implementación demuestra que, con una arquitectura adecuada, herramientas escalables y formación mínima, se pueden generar impactos inmediatos en eficiencia, percepción de valor y toma de decisiones. Esto convierte al caso de la UES en un referente institucional replicable.

7.2. Recomendaciones

1. Escalar progresivamente la implementación del Data Lake a otras áreas funcionales.

Dado el éxito en RRHH, se recomienda que la UES extienda esta arquitectura a áreas como Académica, Investigación y Gestión Administrativa, promoviendo así una gobernanza de datos más transversal y estratégica.

2. Ampliar el uso de RPA a otros procesos críticos de RRHH.

Se sugiere identificar y priorizar nuevos procesos candidatos a automatización como selección de personal, control de asistencia o gestión de licencias, para consolidar el ahorro de tiempo y reducir errores humanos.

3. Fortalecer la capacitación continua en el uso de herramientas analíticas y automatizadas.

La percepción positiva depende en gran medida del empoderamiento del usuario. Por ello, se recomienda institucionalizar talleres prácticos sobre uso de dashboards, interpretación de datos y supervisión de bots RPA.

4. Monitorear y documentar indicadores post implementación.

Se sugiere crear un sistema de monitoreo continuo para medir indicadores como tiempo de atención, calidad del dato, nivel de automatización, y satisfacción del usuario. Esto permitirá evaluar mejoras continuas y tomar decisiones basadas en evidencia.

5. Incorporar estos resultados en la estrategia institucional de transformación digital.

Finalmente, se recomienda que los aprendizajes de este proyecto se integren formalmente en el plan estratégico digital de la UES, asegurando sostenibilidad, escalabilidad y alineamiento con los objetivos institucionales de innovación y eficiencia.

REFERENCIAS BIBLIOGRÁFICAS

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*.

Almaraz, F., Maz, A., & López, C. (2017). Análisis de la transformación digital de las Instituciones de Educación Superior. Un marco de referencia teórico. *Revista de Educación Mediática y TIC*.

Amazon. (n.d.). *What is a Data Lake? - Introduction to Data Lakes and Analytics*. Retrieved May 15, 2024, from https://aws.amazon.com/what-is/data-lake/

Campos, J., Choque, C., Olmos, S., & Uribe, Y. (2024). Estrategias de transformación digital en empresas tradicionales. *Revista Venezolana de Gerencia*.

Capterra. (n.d.). *UiPath - Opiniones, precios y características*. Retrieved October 6, 2024, from https://www.capterra.pe/software/135186/uipath-robotic-process-automation

Creswell, J. W., & Creswell, J. D. (2018). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. In *SAGE*. Sage publications.

EY. (2025). *Desafíos y Tendencias de las Empresas en Latinoamérica*.

EY Perú. (2025). *Incertidumbre política local y transformación digital son los principales desafíos para las empresas peruanas*. https://www.ey.com/es_pe/newsroom/2025/06/incertidumbre-politica-local-transformacion-digital

Guerrero, H. (2019). *Transformación digital en una empresa de consumo masivo usando tecnología SAP*. UNMSM.

Hernández-Sampieri, R., & Mendoza, C. (2018). *Metodología de la investigación: las rutas cuantitativa, cualitativa y mixta*. McGraw Hill.

IBM. (n.d.). *What is Robotic Process Automation (RPA)?* Retrieved July 18, 2024, from https://www.ibm.com/topics/rpa

Kirkwood, G. (2019). *Impact of RPA on Employee Engagement*. https://www.uipath.com/blog/rpa/impact-of-rpa-on-employee-engagement-forrester

Le Clair, C., O'Donnell, G., Lipson, A., & Lynch, D. (2019). *The Forrester Wave™: Robotic Process Automation, Q4 2019*.

Lizano, H., & Palos, P. (2020). *Transformación Digital en Instituciones de Educación Superior con Gestión de Procesos de Negocio Modelo de mediación de Automatización Robótica de Procesos*.

Mendieta, L. (2022). *Optimización de la gestión organizacional en una empresa de publicidad arquitectónica implementando tecnología digital*. Universidad Militar Nueva Granada.

Microsoft Azure. (n.d.). *¿Qué es un lago de datos? Lago de datos frente a Almacenamiento*. Retrieved July 10, 2024, from https://azure.microsoft.com/es-mx/resources/cloud-computing-dictionary/what-is-a-data-lake/

Patton, M. (2015). *Qualitative research & evaluation methods: Integrating theory and practice*. Sage publications.

Quiroz, J., Chavez, G., & Guillen, S. (2022). *Incremento del nivel de servicio en PYMES de consultoría ambiental mediante un modelo colaborativo entre Data Analytics y PDCA*. *2022-December*.

Rojas, J., Ajuria, J., & Arambarri, J. (2023). Metodología de transformación digital para incrementar la competitividad de las pymes de logística ligera en el Perú. *Industrial Data*.

Romero, M., Romeu, T., Guitert, M., & Baztán, P. (2023). La transformación digital en la educación superior el caso de la UOC. *RIED-Revista Iberoamericana de Educación a Distancia*.

Salgado, N., Guamba, A., & Guerrero, R. (2024). El impacto de la tecnología de la información en la gestión empresarial. *Nexus Research Journal*.

Serna, M., Branch, J., Benavides, L., & Burgos, D. (2018). A conceptual model of digital transformation. openergy and thecase study of universidad nacional de colombia. *Education in the Knowledge Society*.

UiPath. (n.d.). *What is Robotic Process Automation - RPA Software*. Retrieved September 27, 2024, from https://www.uipath.com/rpa/robotic-process-automation

Yin, R. (2018). *Case Study Research and Applications*. Sage publications.

ANEXOS

Anexo 1: Ficha de Validación de Expertos

FICHA DE VALIDACIÓN DE EXPERTO

Estimado experto se solicita su apoyo para la validación del instrumento de medición "Encuesta de Percepción sobre Acceso a Datos y Toma de Decisiones en el contexto de Transformación Digital".

Instrucciones: Marque con una "X" la valoración correspondiente en cada ítem, considerando la siguiente escala:
1 = Muy bajo 2 = Bajo 3 = Alto 4 = Muy alto

		Claridad*				Pertinencia**				Relevancia***			
Dimensión 1: Percepción del acceso a datos		1	2	3	4	1	2	3	4	1	2	3	4
A1	Accedo más rápido a los datos que necesito para mi trabajo.				X				X				X
A2	La información ahora se encuentra más centralizada.				X				X				X
A3	He reducido el tiempo que gasto buscando archivos o información de otros sistemas.				X				X				X
A4	Encuentro los datos que necesito sin depender de otras áreas o personas.				X				X				X
A5	La plataforma o sistema actual facilita el acceso a datos actualizados y confiables.			X					X				X
Dimensión 2: Percepción en la toma de decisiones		1	2	3	4	1	2	3	4	1	2	3	4
D1	Ahora tomo decisiones con mayor rapidez gracias a la disponibilidad de datos.				X				X				X
D2	Siento que las decisiones que tomo están mejor fundamentadas en información real.			X					X				X
D3	Puedo anticiparme a problemas o necesidades gracias a la información integrada.				X				X				X
D4	La visualización de los datos (por ejemplo, dashboards o reportes) me ayuda a decidir mejor.				X				X				X
D5	La calidad de mis decisiones ha mejorado desde la implementación de estas herramientas.				X				X				X

Observaciones (precisar si hay suficiencia): hay suficiencia

Opinión de aplicabilidad: **Aplicable (X)** **Aplicable después de corregir ( )** **No aplicable ( )**

Apellidos y Nombres del experto validador: García Castillo ANDREA DNI: 42092912

Especialidad: Profesional con experiencia en mejora de procesos.

Años de experiencia: 8 años

Claridad*: El ítem se comprende con facilidad, sin ambigüedades ni dificultades de interpretación.
Pertinencia**: El ítem se ajusta al concepto teórico que pretende medir.
Relevancia*:** El ítem es adecuado y esencial para representar la dimensión a la que pertenece dentro del instrumento.
Nota sobre suficiencia: La suficiencia se refiere al grado en que el conjunto total de ítems incluidos en el instrumento es suficiente para medir por completo la dimensión evaluada.

Firma del experto

FICHA DE VALIDACIÓN DE EXPERTO

Estimado experto se solicita su apoyo para la validación del instrumento de medición "Encuesta de Percepción sobre Acceso a Datos y Toma de Decisiones en el contexto de Transformación Digital".

Instrucciones: Marque con una "X" la valoración correspondiente en cada ítem, considerando la siguiente escala:
1 = Muy bajo 2 = Bajo 3 = Alto 4 = Muy alto

		Claridad*				Pertinencia**				Relevancia***			
Dimensión 1: Percepción del acceso a datos		1	2	3	4	1	2	3	4	1	2	3	4
A1	Accedo más rápido a los datos que necesito para mi trabajo.				X				X				X
A2	La información ahora se encuentra más centralizada.				X				X				X
A3	He reducido el tiempo que gasto buscando archivos o información de otros sistemas.				X				X				X
A4	Encuentro los datos que necesito sin depender de otras áreas o personas.				X				X				X
A5	La plataforma o sistema actual facilita el acceso a datos actualizados y confiables.				X				X				X
Dimensión 2: Percepción en la toma de decisiones		1	2	3	4	1	2	3	4	1	2	3	4
D1	Ahora tomo decisiones con mayor rapidez gracias a la disponibilidad de datos.				X				X				X
D2	Siento que las decisiones que tomo están mejor fundamentadas en información real.				X				X				X
D3	Puedo anticiparme a problemas o necesidades gracias a la información integrada.				X				X				X
D4	La visualización de los datos (por ejemplo, dashboards o reportes) me ayuda a decidir mejor.				X				X				X
D5	La calidad de mis decisiones ha mejorado desde la implementación de estas herramientas.				X				X				X

Observaciones (precisar si hay suficiencia): Hay suficiencia.

Opinión de aplicabilidad: **Aplicable** (X) **Aplicable después de corregir** ( ) **No aplicable** ( )

Apellidos y Nombres del experto validador: Martinez Camacho, C Manlio DNI: 44216818

Especialidad: Ing. Industrial.

Años de experiencia: 13 años

Claridad*: El ítem se comprende con facilidad, sin ambigüedades ni dificultades de interpretación.
Pertinencia:** El ítem se ajusta al concepto teórico que pretende medir.
Relevancia*:** El ítem es adecuado y esencial para representar la dimensión a la que pertenece dentro del instrumento.
Nota sobre suficiencia: La suficiencia se refiere al grado en que el conjunto total de ítems incluidos en el instrumento es suficiente para medir por completo la dimensión evaluada.



Firma del experto

FICHA DE VALIDACIÓN DE EXPERTO

Estimado experto se solicita su apoyo para la validación del instrumento de medición "Encuesta de Percepción sobre Acceso a Datos y Toma de Decisiones en el contexto de Transformación Digital".

Instrucciones: Marque con una "X" la valoración correspondiente en cada ítem, considerando la siguiente escala:
1 = Muy bajo 2 = Bajo 3 = Alto 4 = Muy alto

		Claridad*				Pertinencia**				Relevancia***			
Dimensión 1: Percepción del acceso a datos		1	2	3	4	1	2	3	4	1	2	3	4
A1	Accedo más rápido a los datos que necesito para mi trabajo.				X				X				X
A2	La información ahora se encuentra más centralizada.				X				X				X
A3	He reducido el tiempo que gasto buscando archivos o información de otros sistemas.				X				X				X
A4	Encuentro los datos que necesito sin depender de otras áreas o personas.				X				X			X	
A5	La plataforma o sistema actual facilita el acceso a datos actualizados y confiables.				X				X				X
Dimensión 2: Percepción en la toma de decisiones		1	2	3	4	1	2	3	4	1	2	3	4
D1	Ahora tomo decisiones con mayor rapidez gracias a la disponibilidad de datos.				X				X				X
D2	Siento que las decisiones que tomo están mejor fundamentadas en información real.				X				X				X
D3	Puedo anticiparme a problemas o necesidades gracias a la información integrada.				X				X			X	
D4	La visualización de los datos (por ejemplo, dashboards o reportes) me ayuda a decidir mejor.				X				X				X
D5	La calidad de mis decisiones ha mejorado desde la implementación de estas herramientas.				X				X				X

Observaciones (precisar si hay suficiencia): Hay suficiencia

Opinión de aplicabilidad: **Aplicable (X)** **Aplicable después de corregir ( )** **No aplicable ( )**

Apellidos y Nombres del experto validador: Pavcar García Bryand DNI: 70492077

Especialidad: Estadístico y magister en Ciencia de Datos con experiencia en TxD

Años de experiencia: 5 años

Claridad*: El ítem se comprende con facilidad, sin ambigüedades ni dificultades de interpretación.
Pertinencia:** El ítem se ajusta al concepto teórico que pretende medir.
Relevancia*:** El ítem es adecuado y esencial para representar la dimensión a la que pertenece dentro del instrumento.
Nota sobre suficiencia: La suficiencia se refiere al grado en que el conjunto total de ítems incluidos en el instrumento es suficiente para medir por completo la dimensión evaluada.



Firma del experto

FICHA DE VALIDACIÓN DE EXPERTO

Estimado experto se solicita su apoyo para la validación del instrumento de medición "Encuesta de Percepción sobre Acceso a Datos y Toma de Decisiones en el contexto de Transformación Digital".

Instrucciones: Marque con una "X" la valoración correspondiente en cada ítem, considerando la siguiente escala:
1 = Muy bajo 2 = Bajo 3 = Alto 4 = Muy alto

		Claridad*				Pertinencia**				Relevancia***			
Dimensión 1: Percepción del acceso a datos		1	2	3	4	1	2	3	4	1	2	3	4
A1	Accedo más rápido a los datos que necesito para mi trabajo.			X				X				X	
A2	La información ahora se encuentra más centralizada.			X				X					X
A3	He reducido el tiempo que gasto buscando archivos o información de otros sistemas.			X				X				X	
A4	Encuentro los datos que necesito sin depender de otras áreas o personas.			X				X				X	
A5	La plataforma o sistema actual facilita el acceso a datos actualizados y confiables.				X			X					X
Dimensión 2: Percepción en la toma de decisiones		1	2	3	4	1	2	3	4	1	2	3	4
D1	Ahora tomo decisiones con mayor rapidez gracias a la disponibilidad de datos.			X				X				X	
D2	Siento que las decisiones que tomo están mejor fundamentadas en información real.			X				X				X	
D3	Puedo anticiparme a problemas o necesidades gracias a la información integrada.				X				X			X	
D4	La visualización de los datos (por ejemplo, dashboards o reportes) me ayuda a decidir mejor.				X				X			X	
D5	La calidad de mis decisiones ha mejorado desde la implementación de estas herramientas.				X			X					X

Observaciones (precisar si hay suficiencia): Hay suficiencia

Opinión de aplicabilidad: Aplicable (X) Aplicable después de corregir ( ) No aplicable ( )

Apellidos y Nombres del experto validador: Saavedra Martínez, Manuel DNI: 07474759

Especialidad: Profesionalización y Tecnología, asesor de Tesis

Años de experiencia: 15 años

Claridad*: El ítem se comprende con facilidad, sin ambigüedades ni dificultades de interpretación.
Pertinencia:** El ítem se ajusta al concepto teórico que pretende medir.
Relevancia*:** El ítem es adecuado y esencial para representar la dimensión a la que pertenece dentro del instrumento.
Nota sobre suficiencia: La suficiencia se refiere al grado en que el conjunto total de ítems incluidos en el instrumento es suficiente para medir por completo la dimensión evaluada.

Firma del experto

Anexo 2: Organización del Data Lake (Medallón)

Bronze				Silver			Gold				
carpeta	subcarpeta	frecuencia de carga	descripcion del archivo	notebooks		tablas delta	carga incremental	carga historica	pipeline	esquema dimensional	
Activos	activos_a_la_fecha	diario	Reporte Activos a la fecha	nb-prep-activos	nb-ingest-activos	activos_a_la_fecha	df-activos-inc	df-activos-hist	pl-padron-activos	fct-padron-activos	gld-padron-activos
Activos	contratos_registrados	diario	DRH_CONTRATOS_REGISTRADOS	nb-prep-activos	nb-ingest-activos	contratos_registrados	df-activos-inc	df-activos-hist	pl-padron-activos	fct-padron-activos	gld-padron-activos
Activos	genero_y_estado_civil	diario	DRH_DATOS_GENERO_EST_CIVL	nb-prep-activos	nb-ingest-activos	genero_y_estado_civil	df-activos-inc	df-activos-hist	pl-padron-activos	fct-padron-activos	gld-padron-activos
Activos	ingreso_a_procesos	diario	DRH_ULT_FECHA_INGRESO_PROCESO	nb-prep-activos	nb-ingest-activos	ingreso_a_procesos	df-activos-inc	df-activos-hist	pl-padron-activos	fct-padron-activos	gld-padron-activos
Activos	posiciones_por_persona	diario	DRH_POSICIONES_POR_PERSONA	nb-prep-activos	nb-ingest-activos	posiciones_por_persona	df-activos-inc	df-activos-hist	pl-padron-activos	fct-padron-activos	gld-padron-activos
Activos	unidades	diario	Archivo del access unidades	nb-prep-activos	nb-ingest-activos	unidades	df-activos-inc	df-activos-hist	pl-padron-activos	fct-padron-activos	gld-padron-activos
Marcaciones	marcaciones_pucp	diario	20. DGTH (DOCENTES Y ADMINISTRATIVOS)		nb-ingest-marcaciones	marcaciones_pucp	df-marcaciones-inc	df-marcaciones-hist	pl-presencialidad	fct-presencialidad	
RAE	raes	mensual	Reporte del campus para el reporte presupuestal	nb-prep-raes	nb-ingest-raes	raes	df-rae-inc	df-rae-hist	pl-rae	fct-rae	
RAE	captura_de_incidencias	mensual	DRH_GANAN_INCID_X_CLDR_USUARIO	nb-prep-raes	nb-ingest-raes	captura_de_incidencias		df-rae-hist	pl-rae	fct-rae	
General	accion_motivo	a demanda	C:\RawLayer		nb-ingest-dim	accion_motivo	df-catalogo-inc		pl-accion-motivo	dim-accion-motivo	
General	general_partidas	a demanda	C:\RawLayer		nb-ingest-dim	general_partidas	df-catalogo-inc		pl-general-partidas	dim-general-partidas	gld-reporte-planillas
General	general_cuentas	a demanda	C:\RawLayer		nb-ingest-dim	general_cuentas	df-catalogo-inc		pl-general-cuentas	dim-general-cuentas	gld-reporte-planillas
General	fijos_formulados	a demanda	C:\RawLayer		nb-ingest-dim	fijos_formulados	df-catalogo-inc		pl-fijos-formulados	dim-fijos-formulados	gld-reporte-planillas
General	fijos_variables	a demanda	C:\RawLayer		nb-ingest-dim	fijos_variables	df-catalogo-inc		pl-fijos-variables	dim-fijos-variables	gld-reporte-planillas
Intellisign	estado_documentos	diario	reporte de intellisign		nb-ingest-intellisign	estado_documentos	df-intellisign-inc		pl-intellisign	fct-intellisign	
Planillas	estado_calendarios	diario	DRH_ESTADO_CALENDARIOS_GL	nb-prep-planillas	nb-ingest-planillas	estado_calendarios	df-planillas-inc		pl-reporte-planillas	fct-reporte-planillas	gld-reporte-planillas
Planillas	activos_a_la_fecha_cort	diario	1er día del siguiente mes para incremental	nb-prep-planillas	nb-ingest-planillas	activos_a_la_fecha_corte	df-planillas-inc	df-planillas-hist	pl-reporte-planillas	fct-reporte-planillas	
Planillas	reporte_presupuestal	diario	Reporte Presupuestal	nb-prep-planillas	nb-ingest-planillas	reporte_presupuestal	df-planillas-inc	df-planillas-hist	pl-reporte-planillas	fct-reporte-planillas	
Planillas	codigo_salarial	diario	DRH_CODIGO_SALARIAL	nb-prep-planillas	nb-ingest-planillas	codigo_salarial	df-planillas-inc	df-planillas-hist	pl-reporte-planillas	fct-reporte-planillas	
Planillas	ganancia_teorica_detalle	mensual	DRH_GANAN_TEORICA_DETALLE	nb-prep-planillas	nb-ingest-planillas	ganancia_teorica_detalle		df-planillas-hist	pl-reporte-planillas	fct-reporte-planillas	
Planillas	consulta_cod_salarial	mensual	DRH_CONSULTA_CODSALARIAL (113-01)	nb-prep-planillas	nb-ingest-planillas	consulta_cod_salarial		df-planillas-hist	pl-reporte-planillas	fct-reporte-planillas	

Anexo 3: UiPath Assistant

