



**Diseño e implementación de un mecanismo ETL para nutrir la plataforma DM de la
Dirección de Talento Humano del GADCG.**

Suntaxi Morales, Diego Fabricio

Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

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

Msc. Rea Guamán, Ángel Marcelo

28 de febrero del 2024



Plagiarism and AI Content Detection Report

tesis_SuntaxiDiegoV10.pdf

Scan details

Scan time: February 28th, 2024 at 20:38 UTC Total Pages: 74 Total Words: 18271

Plagiarism Detection



Types of plagiarism		Words
Identical	0.1%	25
Minor Changes	0%	1
Paraphrased	1.5%	273
Omitted Words	2%	361

AI Content Detection



Text coverage		Words
AI text	4.1%	744
Human text	95.9%	17166

[Learn more](#)

Firma:



Firma electrónica por:
ANGEL MARCELO REA
GUAMAN

Rea Guaman, Angel Marcelo

Director



Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Certificación

Certifico que el trabajo de integración curricular: **"Diseño e implementación de un mecanismo ETL para nutrir la plataforma DM de la Dirección de Talento Humano del GADCG"** fue realizado por el señor **Suntaxi Morales, Diego Fabricio**, el mismo que cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, además fue revisado y analizada en su totalidad por la herramienta de prevención y/o verificación de similitud de contenidos; razón por la cual me permito acreditar y autorizar para que se lo sustente públicamente.

Sangolquí, 28 febrero del 2024

Firma:



Rea Guaman, Ángel Marcelo

C. C.: 020106877-2



Departamento de Ciencias de la Computación
Carrera de Tecnologías de la Información

Responsabilidad de Autoría

Yo, **Suntaxi Morales, Diego Fabricio**, con cédula de ciudadanía n° 172340169-9, declaro que el contenido, ideas y criterios del trabajo de integración curricular: **Diseño e implementación de un mecanismo ETL para nutrir la plataforma DM de la Dirección de Talento Humano del GADCG** es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos, y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 28 de febrero del 2024

Firma

Suntaxi Morales, Diego Fabricio

C.C.: 172340169-9



Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Autorización de Publicación

Yo **Suntaxi Morales, Diego Fabricio**, con cédula de ciudadanía n° 172340169-9, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de integración curricular: **Título: Diseño e implementación de un mecanismo ETL para nutrir la plataforma DM de la Dirección de Talento Humano del GADCG** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 28 de febrero del 2024

Firma

Suntaxi Morales, Diego Fabricio

C.C.: 172340169-9

Dedicatoria

En primer lugar, este logro está dedicado a Dios por su guía durante la consecución de este logro.

Este logro es dedicado a mis padres Jorge y María, quienes con su esfuerzo incalculable y su trabajo incansable han apoyado mi camino universitario. Todo su esfuerzo y sacrificio ha sido recompensado con este logro, el cual se lo debo a ustedes por su dedicación y sacrificio.

Quiero dedicar este la consecución de este logro a mis hermanos Juan y Paul quienes mediante su ejemplo han sido una parte fundamental en este camino, mi agradecimiento por siempre estar conmigo. A mis sobrinos Juan Sebastián y Gabriela, se merecen la dedicación por que mediante su ternura y sus ocurrencias han sido una fuente de motivación a lo largo de este trayecto universitario.

A mis tíos, Beatriz, Elvira y Rodrigo el agradecimiento y dedicación de este logro, ya que en los momentos más difíciles fueron piezas importantes para la continuación de este objetivo.

A mis primos que son mis hermanos Santiago y José, quienes con sus palabras atinadas de apoyo han sido un respaldo incondicional en todos los momentos de mi vida, la dedicación de este éxito en mi vida.

A mis compañeros y con el paso del tiempo amigos, Paul, José, Sebastián y Mateo con quienes compartí el proceso de este desafiante camino universitario desde el día uno en la nivelación. Donde se presentaron múltiples momentos de risas y enojos, pero siempre prevaleciendo la amistad, siempre los recordare con afecto.

A mi mejor amiga, María Fernanda quien ha estado en todos los momentos de este proceso universitario siendo un apoyo incondicional mi profundo agradecimiento.

Por último, mi agradecimiento y dedicación de este logro a mis profesores con los cuales he formado una amistad, quienes mediante sus palabras han sido parte importante de este logro.

DIEGO SUNTAXI

Agradecimientos

Inicialmente quiero expresar mi alegría y agradecimiento con Dios, por ser el guía en cada paso de esta etapa universitaria y pilar esencial en mi vida con sus bendiciones.

Mi reconocimiento y agradecimiento a la Universidad de las Fuerzas Armadas ESPE y a los docentes del Departamento de Ciencias de la Computación, quienes mediante sus enseñanzas han sido indispensables en mi formación profesional. Mi agradecimiento eterno a los docentes con los que compartí en las aulas, quienes mediante sus enseñanzas y consejos aportaron guiaron mi camino universitario.

El agradecimiento especial a mis compañeros con los cuales al pasar del tiempo se han formado lazos de amistad que no tengo duda prevalecerá con el tiempo. Ya que en este camino llamado universidad se tuvo experiencias buenas y no tan buenas, pero nos llenamos de valor para saberlas superar.

A mi tutor de tesis, el agradecimiento especial por su guía en este proyecto de integración curricular, el cual es un trabajo crucial que marca el final de este camino universitario. Sin dejar pasar que desde el primer momento con su carisma me brindó su amistad y consejos que han sido inspiradores en mi persona.

No puedo dejar pasar este momento para dar mi agradecimiento a mis conocid@s y amig@s, quienes a lo largo de este camino universitario mediante sus sabias palabras me demostraron su constante respaldo.

Finalmente, a toda mi familia quienes vivieron el día a día de este arduo camino universitario el agradecimiento fraterno. Sus atinadas palabras de respaldo y aliento en los diversos momentos por los que cursaba, fueron el impulso que necesitaba para seguir. En especial el agradecimiento a mis Padres, Hermanos y Sobrinos quienes con su comprensión, paciencia, motivación y afecto diario han inspirado la culminación de este objetivo.

DIEGO SUNTAXI

Índice de contenido

Dedicatoria	6
Agradecimientos.....	7
Resumen.....	16
Abstract.....	17
Capítulo I: Introducción.....	18
Introducción.....	18
Antecedentes	19
Problemática	21
Justificación.....	22
Objetivos	23
Objetivo General.....	23
Objetivos Específicos.....	23
Alcance.....	23
Planteamiento de la hipótesis	23
Capítulo II: Marco Teórico	24
Conceptos Generales.....	24
Big Data.....	25
Tipos de datos según su formato.....	27
Minería de datos – Data Mining	28
Ciclo de vida de la información	29
Etapas del ciclo de vida de los datos	29

Inteligencia de Negocios - Business Intelligence	31
Funcionamiento del Business Intelligence	32
Data Warehouse.....	32
Componentes Data Warehouse.....	33
Data Mart.....	34
Estructura de un Data Mart.....	36
Tipos de data Mart.....	37
Diferencias entre Data Warehouse y Data Mart.....	38
ETL.....	39
Extracción.....	40
Transformación.....	40
Carga.....	40
Características ETL	41
Capítulo III: Herramientas.....	42
Cuadrante Mágico de Gartner	42
Funcionamiento	43
Herramientas para la integración de datos	43
Talend	44
IBM InfoSphere DataStage	45
Python	47
Herramientas para el almacenamiento de datos.....	48

	10
MySQL	48
PostgreSQL	49
Oracle Database.....	50
Herramientas para la visualización de datos.....	52
Power BI.....	52
Tableau	54
Qlik	55
Criterios de selección de las herramientas	56
Herramienta de integración de datos	56
Herramienta de almacenamiento de datos.....	57
Herramienta de visualización de datos	58
Capítulo IV: Metodología	60
Metodología de Hefesto.....	60
Fases de la metodología de Hefesto.....	62
Características de la metodología de Hefesto.....	62
Metodología Ralph Kimball	63
Principios de la metodología de Ralph Kimball	63
Fases de la metodología de Ralph Kimball	64
Comparación entre las metodologías de Hefesto y Ralph Kimball	65
Justificación de la metodología a seguir	66
Capítulo V: Desarrollo	68

	11
Descripción de la metodología a seguir	68
Fase 1: Análisis de requerimientos	68
Fase 2: Análisis de los OLTP	69
Fase 3: Modelo lógico del Data Mart o Data Warehouse	70
Fase 4: Integración de datos.....	72
Caso práctico.....	73
Fase 1: Análisis de requerimientos	73
Fase 2: Análisis de los OLTP.....	75
Fase 3: Modelo lógico del Data Mart o Data Warehouse	80
Fase 4: Integración de datos.....	82
Cubos de información.....	84
Resultados obtenidos	87
Cumplimiento de objetivos.....	90
Revisión de Objetivos:	90
Evaluación del Cumplimiento de Objetivos:	91
Validación de la hipótesis	95
Capítulo VI: Conclusiones y recomendaciones.....	99
Conclusiones.....	99
Recomendaciones.....	99
Referencias	101
Apéndices	108

Índice de tablas

Tabla 1 Las 5 “V” del Big Data	266
Tabla 2 Tipos de datos, según su estructura.....	277
Tabla 3 Clasificación de los Data Mart, de acuerdo a la fuente de obtención de los datos .	355
Tabla 4 Ventajas y desventajas de la clasificación de Data Mart según la fuente de obtención de la Data	366
Tabla 5 Estructura de los Data Mart.....	377
Tabla 6 Tipos de Data Mart.....	377
Tabla 7 Comparativa entre Data Warehouse y Data Mart	388
Tabla 8 Comparativa Herramientas Integración de Datos	566
Tabla 9 Comparativa Herramientas Almacenamiento de Datos	577
Tabla 10 Comparativa Herramientas de Visualización	588
Tabla 11 Metodología Hefesto vs Metodología Ralph Kimball.....	655
Tabla 12 Indicadores características metodologías.....	677
Tabla 13 Diccionario de datos de la tabla uniformes	788
Tabla 14 Diccionario de datos de la tabla uniformes-personal	79
Tabla 15 Diccionario de datos de la tabla rol.....	79
Tabla 16 Resumen de tablas desarrolladas	911
Tabla 17 Resumen de relaciones generadas	922

Índice de figuras

Figura 1 Dato, información y conocimiento	255
Figura 2 Componentes de un Data Warehouse	333
Figura 3 Esquema de un proceso ETL.....	400
Figura 4 Cuadrante Mágico de Gartner.....	422
Figura 5 Cuadrante Mágico Gartner Integración de Datos	433
Figura 6 Cuadrante Mágico Gartner Business Intelligence	522
Figura 7 Fases de la metodología de Hefesto.....	611
Figura 8 Fases de la metodología de Ralph Kimball	644
Figura 9 Modelo Conceptual.....	69
Figura 10 Modelo conceptual Ampliado	700
Figura 11 Esquema de una tabla de dimensiones	711
Figura 12 Esquema de una tabla de hechos.....	722
Figura 13 Modelo Conceptual del caso de estudio.....	755
Figura 14 Modelo Entidad Relación Encontrado	766
Figura 15 Modelo Entidad Relación Actualizado.....	777
Figura 16 Relación entre el Modelo Conceptual y Modelo Entidad Relación	788
Figura 17 Modelo Conceptual Ampliado del caso de estudio.....	79
Figura 18 Modelo Copo Nieve	800
Figura 19 Tabla de dimensiones uniformes	811
Figura 20 Tabla de dimensiones Categoría Uniformes	811
Figura 21 Tabla de dimensiones Rol	811
Figura 22 Tabla de hechos Uniformes-Personal	822
Figura 23 Uniones realizadas en el modelo	822
Figura 24 Esquema de tablas en MySQL	833
Figura 25 Arquitectura ETL.....	844

Figura 26 Panel de control del Personal	888
Figura 27 Panel de control de Uniformes.....	888
Figura 28 Panel de control de Partidas Presupuestarias	89
Figura 29 Panel de control de Roles de los empleados	89
Figura 30 Panel de control de Beneficios de Ley	900
Figura 31 Panel de control Horas Extras	900
Figura 32 Distribución del personal por Rol	955
Figura 33 Gasto de vestimenta por Rol.....	966
Figura 34 Valor a pagar por las horas extras	977

Índice de Funciones DAX

Función 1 Empleados por categoría	855
Función 2 Prendas entregadas por rol.....	866
Función 3 Gasto empleado en vestimenta	866
Función 4 Gasto en salarios por rol.....	866
Función 5 Promedio de años de permanencia en la institución de los empleados	877
Función 6 Contabilización del personal según su nivel académico.....	877
Función 7 Promedio de antigüedad del personal por el rol que ocupa.....	877

Resumen

El presente trabajo de integración curricular se enfoca en el diseño e implementación de un mecanismo basado en la Extracción, Transformación y Carga o conocido por sus siglas como (ETL) a través del manejo de Python, siendo esta una herramienta de programación versátil que, de manera simultánea con sus librerías, facilita el adecuado tratamiento de los datos. Una vez completado el proceso ETL, los datos se almacenarán en un Data Mart (DM) en MySQL perteneciente a la Dirección de Talento Humano del Gobierno Autónomo Descentralizado del cantón Guaranda (GADCG) y el mismo contendrá la información que previamente ha sido tratada en el proceso ETL. Este mecanismo desarrollado permitirá ubicarse al área a la vanguardia en temas tecnológicos relacionados a la integración de datos, permitiendo de esta manera mejorar la recuperación y recopilación de la información que el área genere en sus distintas actividades. Cabe mencionar que el proceso previo de tratamiento de la información se efectuaba de manera manual, lo que originaba la posibilidad de errores.

La implementación de este mecanismo será de gran utilidad para los miembros del área de Talento Humano del GADCG, ya que garantizará la disponibilidad, confiabilidad e integridad de la información. Además, permitirá tomar decisiones informadas, ya que ofrecerá datos actualizados y precisos en base a la información que maneja el área de las actividades que se encuentran bajo su cargo. Una vez que la información esté almacenada en el Data Mart del área, será posible crear paneles de control (visualizaciones) utilizando la herramienta Microsoft Power BI Desktop en su versión gratuita permitiendo presentar la información recopilada de una manera visual, lo cual permitirá la mejor interpretación de los mismos. Esta implementación de bajo costo, pero de gran utilidad, proporcionará al área de Talento Humano la información necesaria de manera oportuna y en el lugar adecuado, facilitando así la toma de decisiones estratégicas basadas en datos confiables y actualizados.

Palabras Clave: ETL, Talento Humano, Data Mart, paneles de control, toma de decisiones

Abstract

This curricular integration work focuses on the design and implementation of a mechanism based on Extraction, Transformation and Loading (ETL) through the use of Python, a versatile programming tool that, simultaneously with its libraries, facilitates the proper processing of data. Once the ETL process is completed, the data will be stored in a Data Mart (DM) in MySQL belonging to the Human Resources area of the Decentralized Autonomous Government of Guaranda Canton (GADCG) and it will contain the information that has been previously treated in the ETL process. This developed mechanism will allow the area to be at the forefront of technological issues related to data integration, thus improving the recovery and collection of information that the area generates in its various activities. It is worth mentioning that the previous process of information processing was carried out manually, which led to the possibility of errors.

The implementation of this mechanism will be very useful for the members of the Human Resources area of the GADCG, since it will guarantee the availability, reliability and integrity of the information. In addition, it will allow informed decision making, since it will provide updated and accurate data based on the information handled by the area on the activities under its responsibility. Once the information is stored in the area's Data Mart, it will be possible to create control panels (visualizations) using the Microsoft Power BI Desktop tool in its free version, allowing to present the information collected in a visual way, which will allow a better interpretation of them. This low-cost but very useful implementation will provide the Human Resources area with the necessary information in a timely manner and in the right place, thus facilitating strategic decision making based on reliable and updated data.

Keywords: ETL, Human Resources, Data Mart, control panels, decision making.

Capítulo I: Introducción

El presente capítulo del trabajo de integración curricular está compuesto de los siguientes elementos: introducción, antecedentes, exposición de la problemática, justificación, objetivo general, objetivos específicos, alcance y el planteamiento de la hipótesis.

Introducción

La información desde sus inicios ha sido de utilidad para cada una de las personas, debido a que cada dato que perciba el ser humano permitirá su desarrollo en la vida cotidiana y desde el enfoque empresarial la manipulación de la información facilitara a cada sector productivo tomar decisiones acordes a su realidad. (Benítez, 2009)

A medida que los datos se incrementan se vuelven intratables, lo cual conlleva al procesamiento de la información de manera parcial, lo que puede incurrir a largo plazo en la toma de decisiones erróneas, por lo cual, la necesidad de implementar procesos ETL (extracción, transformación y carga) se han visto necesarios y de gran utilidad, para que de esta manera se pueda nutrir a las plataformas Data Mart conocidas también como DM (almacenes de datos de una unidad específica de la organización) o Data Warehouse conocidas también como DW (almacenes de datos de múltiples unidades de la organización) con información necesaria que será de gran ayuda para la toma de decisiones en las instituciones.

El Gobierno Autónomo Descentralizado del Cantón Guaranda (GADCG) se enfrenta a desafíos significativos en el manejo de la información debido a la influencia de la globalización y el incremento sustancial de los servicios brindados a la ciudadanía. Esta realidad ha vuelto obsoleto el enfoque tradicional en la gestión de datos, abriendo así nuevas oportunidades para mejorar los procesos de Tecnologías de la Información (TI) en la municipalidad. En este contexto, la situación en el Gobierno Autónomo Descentralizado del Cantón Guaranda no es ajena a los cambios inducidos por la globalización y el incremento de datos producto de la implementación de nuevos servicios a la ciudadanía, lo que subraya la necesidad de

adaptación y modernización en el tratamiento de los datos y poder generar información relevante y oportuna.

Bajo lo expuesto, este trabajo de integración curricular se verá enfocado en la generación e implementación de un mecanismo ETL que sea capaz de nutrir la plataforma Data Mart de la Dirección de Talento Humano del Gobierno Autónomo Descentralizado del Cantón Guaranda, con lo cual se busca que los procesos de toma de decisiones dentro de la mencionada área se vean beneficiados por contar con información acertada y destacada.

Antecedentes

Ortiz (2021), en su trabajo de titulación hace alusión a la implementación de un almacén de datos que va de la mano con procesos ETL, el mismo que se encargara de someter a los datos a los distintos procesos de depuración y limpieza, lo cual permitirá aportar con información en el proceso de toma de decisiones al directorio y a las personas administradoras, dicha información posibilitará la combinación de variables con la finalidad de realizar análisis estadísticos y generar informes, facilitando así la capacidad de analizar la información de manera más profunda y detallada, el autor desarrolla la investigación mediante la metodología de Hefesto.

Mientras que para Trujillo (2018), manifiesta que debido a la evolución organizacional a nivel mundial se ha visto la necesidad de almacenar y procesar grandes volúmenes de datos. El procesamiento de la información no debe afectar en las funciones cotidianas de las organizaciones, obteniendo como consecuencia la necesidad de establecer nuevas formas de acumular la información a través de almacenes de datos, en la cual la información requerida para los distintos procesos de tratamiento de la información se encuentran consolidados en un repositorio, con ese objetivo el autor en su trabajo plantea el uso de la metodología Ralph Kimball, y ciertas fases en la metodología Hefesto como respaldo.

Para Velasco (2020), en su trabajo determina la necesidad de desarrollar un Data Mart para el proceso de "Administración Tributaria del área de Rentas y Finanzas en la Municipalidad

Distrital de Puente Piedra – Perú”, donde concluye con la existencia de carencias con respecto a los distintos escenarios que permiten evaluar indicadores de morosidad. Permitiendo visualizar esta implementación como una oportunidad de contemplar el cumplimiento del objetivo de identificar que el desarrollo e implementación del Data Mart, ocasiona una reducción del índice de morosidad por contribuyente del 86% al 60%, lo cual permitió entender que estas nuevas formas de tecnología permiten complementar y mejorar los procesos estratégicos de la organización. Para el mencionado trabajo se utilizó la metodología de Hefesto.

Por otro lado, Ayala (2021), sustenta su investigación en el desarrollo de un Data Mart dirigido hacia el departamento de Análisis de la Instancia Técnica, con el objetivo de gestionar la información y mejorar la toma de decisiones. Cabe destacar que el autor implemento en su trabajo la metodología de Ralph Kimball, permitiéndose detallar todas las fases de la mencionada metodología.

Es importante destacar que la implementación del Data Mart puede ser gestionada con software de código abierto, como Pentaho Data Integration y MySQL Workbench. Además, las herramientas de Excel y Power BI sirven como extensiones para la elaboración de reportes e informes. En síntesis, la automatización de procesos da como consecuencia una importante disminución en los tiempos de respuesta.

Finalmente, Vento (2020), manifiesta a través de su trabajo de investigación el cual hace referencia a la evaluación del impacto de un Data Mart en el mecanismo de cobranza predial en la Municipalidad Distrital de Ate – Perú, aplicando la metodología de Ralph Kimball, elegida por su ajuste a las necesidades y fases del desarrollo, la cual le permite optimizar el manejo de información histórica de manera dinámica. En la elaboración del Data Mart se empleó herramientas como SQL Server 2019 siendo este un gestor de base de datos y la herramienta de BI en Analysis Services en Visual Studio 2019, la misma que permite la creación de cubos de datos, los cuales fueron dirigidos dependiendo los roles de los usuarios.

Problemática

Los instrumentos de inteligencia de negocios son capaces de respaldar la toma de decisiones, los cuales han quedado lejanos para las diferentes entidades gubernamentales, debido a la priorización de procesos operativos dejando de lado procesos estratégicos del negocio, uno de los elementos que apoya a los procesos estratégicos es el desarrollo de un ETL, debido a que la falta de este proceso en las distintas organizaciones ha generado desafíos significativos en la disponibilidad y calidad de los datos, afectando la capacidad de la organización para tomar decisiones basadas en información oportuna y relevante.

Estos procesos que eran realizados de forma manual, los cuales eran propensos a errores humanos, dificultaban la toma de decisiones debido a la no disponibilidad de información relevante, complicaban la capacidad de respuesta ante las necesidades institucionales.

El Gobierno Autónomo Descentralizado del Cantón Guaranda, perteneciente a la capital de la Provincia de Bolívar en Ecuador, se sitúa a una altitud de 2.668 metros sobre el nivel del mar. Se encuentra a 220 km al sur de Quito y a 150 km de Guayaquil. Con una extensión de 1.898 km cuadrados, la ciudad experimenta un clima agradable con temperaturas que varían entre los 15 y 21 grados centígrados. Su población, compuesta por aproximadamente 65,000 habitantes, se comunica en español y kichwa.

Este trabajo de integración curricular propone realizar el diseño e implementación de un mecanismo ETL para nutrir la plataforma Data Mart del Gobierno Autónomo Descentralizado del Cantón Guaranda, con la información de las diferentes fuentes de información de la Dirección de Talento Humano, que incluye nómina del personal, valores de beneficios de ley, horas extras, subsidios, alimentación y vestimenta, para poder entregar información consolidada a la Dirección y responsables del área de Talento Humano para la toma de decisiones y cumplan con los indicadores de gestión. Este proyecto no solo busca resolver los problemas actuales de gestión de datos, sino también establecer las bases para un enfoque

más integral en la gestión del talento municipal y tener información útil disponible en todo momento.

Justificación

La elaboración del presente trabajo de integración curricular aborda desafíos actuales en cuanto a la gestión de los datos, que incurrirá en los distintos procesos para la toma de decisiones, lo cual permitirá dar un enfoque proactivo al área de Talento Humano mejorando notablemente la calidad del servicio que brinde el área a los colaboradores de la municipalidad.

Los beneficios potenciales que permitirá obtener la implementación de un mecanismo ETL personalizado a la realidad de la municipalidad se enfoca a la mejora en los informes en base a los requerimientos que se han solicitado por parte de la unidad de Talento Humano. En este contexto, se planea clasificar y agrupar la información, presentándola en paneles de control que simplifiquen la identificación precisa de los requerimientos evaluados.

La correlación de las tecnologías de la información y la implementación de almacenes de datos, como el Data Mart, promete proporcionar un acceso más rápido y directo a información procesada y relevante. Este proceso ágil de tratamiento de los datos brinda la oportunidad de involucrar a todos los miembros de la unidad de Talento Humano en la gestión de toma de decisiones. La incorporación de estas tecnologías en el entorno organizacional como lo es la Municipalidad de Guaranda representa una evolución clave en la gestión oportuna de la información para la toma de decisiones.

La gestión eficiente de la información que maneja la unidad de Talento Humano del Gobierno Autónomo Descentralizado del Cantón Guaranda ayudará a mejorar la visión estratégica del área, siendo ejemplo y sobresaliendo entre las demás áreas de la mencionada municipalidad para procesos de toma de decisiones organizativas.

Objetivos

Objetivo General

Diseñar e implementar el Proceso ETL, para alimentar el Data Mart de la Dirección de Talento Humano del GADCG

Objetivos Específicos

- Consolidar la información de la Dirección de Talento Humano
- Desarrollar cubos de información requeridos
- Generar información para la dirección

Alcance

El alcance del presente trabajo de integración curricular radica en el diseño e implementación del proceso de ETL para alimentar el Data Mart siendo aplicable para el área Talento Humano del Gobierno Autónomo Descentralizado del Cantón Guaranda y destinadas a los responsables en tomar decisiones en base a la información que se espera obtener.

Planteamiento de la hipótesis

Los Dashboard o paneles de control que se desarrollaran para representar de manera visual los datos almacenados dentro del Data Mart, mejorarán sustancialmente el proceso relacionado a la toma de decisiones, debido a que el análisis de la información presentada en el mecanismo posibilite el desarrollo de estrategias permitiendo aportar de manera positiva al desarrollo institucional. Además, permitirá presentar los datos de manera intuitiva y de igual forma permitirá que los gráficos presentados se ajusten dinámicamente conforme incremente la data del Data Mart.

Capítulo II: Marco Teórico

El presente capítulo del trabajo de integración curricular se compone de los siguientes elementos: revisión del estado del arte, que incluye una investigación exhaustiva de los componentes idóneos para el trabajo mediante la presentación de un marco teórico sustancial.

Conceptos Generales

Los conceptos relevantes para este trabajo de integración curricular, son los siguientes:

Dato. La Editorial Etece (2021) determina que un dato es una representación simbólica sea esta números, letras o símbolos que pueden representar cualidades cuantitativas o cuantitativas.

A lo cual se puede considerar que el dato es la unidad fundamental generadora de información como se puede observar en la figura 1, ya que, un dato por sí solo no tiene un contexto o un significado el cual nos deslumbre una realidad. Un dato puede ser un número, palabra, símbolo, entre otros. Por ejemplo “540” por sí mismo carece de utilidad para la toma de decisión ya que no nos menciona nada en concreto.

Información. Carrión (s.f), sostiene que la información por su parte es el conjunto de datos que se han procesado, calculado, corregido, filtrado permitiendo así organizar y dar un significado a los datos obtenidos, lo cual posibilitará comprender el contexto en el que los ha recopilado, dándoles así un propósito y relevancia de acuerdo a las necesidades obtenidas. Por ejemplo, el dato “540” puede ser “el salario del personal de la Dirección de Talento Humano de una empresa”. Por su parte la información no representa el conocimiento, más bien es el preámbulo para formar el mismo, como se puede apreciar en la figura 1.

Conocimiento. El conocimiento corresponde a la experiencia, valores, y el actuar en la búsqueda del saber o interpretar la información generada de manera continua, lo cual nos permitirá tener una visualización previa de las nuevas experiencias, siendo estas de vital importancia para la toma de decisiones. Las organizaciones no solo se enfocan en generar conocimiento en base a documentación o bases de datos de información, sino que también

orientan a adquirir conocimiento en base a las rutinas y flujos de trabajo diario que realizan.

(Davenport y Prusak, 1999m como se cita en Carrión, s.f)

Para Carrión (s.f), la información nace del tratamiento de los datos y el conocimiento nace de la información que se ha obtenido, como se puede observar en la figura 1.

Figura 1

Dato, información y conocimiento.



Nota. El grafico representa, el proceso por el que atraviesan los datos.

Tras adquirir conocimientos sobre los conceptos generales y considerando la abundancia de información generada en el Departamento de Talento Humano, surge la relevancia de términos emergentes, entre ellos, el concepto de Big Data.

Big Data

Hace referencia a la agrupación de datos extremadamente extensos y de gran complejidad de gestión, procesamiento o análisis, para lo cual se desarrollaron herramientas que permiten el correcto procesamiento de la información. Sin embargo, el paso del tiempo a permitido visualizar que dichas herramientas se vuelven obsoletas.

El Big Data rompe todo impedimento relacionado a la cantidad de datos e incorpora cualidades específicas como: diversidad, velocidad y calidad de la información. Su naturaleza se enfoca en la capacidad de transmutar los datos en información relevante para la toma de

decisiones. Este campo se encuentra en continúa evolucionando con el surgimiento de nuevas tecnologías y desafíos.

Vega (2020), se refiere al Big Data como "datos masivos" o "macrodatos", siendo estos conjuntos de datos que superan la capacidad de gestión de los softwares convencionales. Englobando el volumen, la diversidad, velocidad, confiabilidad y utilidad de los datos, conocidas como las "5V" del Big Data. Estos datos proporcionan información valiosa para las distintas áreas de la organización lo cual contribuyen a mejorar en la toma de decisiones, optimización de resultados, reducción de costos y desarrollar políticas que beneficien a la sociedad.

Para Gómez (2020), manifiesta que para una gestión eficiente de extensos volúmenes de datos es valedero conocer los 5 aspectos importantes del Big Data, los cuales se detallan en la tabla 1.

Tabla 1

Las 5 "V" del Big Data

Aspecto	Descripción
Volumen	Anteriormente, la obtención de datos se realizaba manualmente, sin el respaldo de máquinas o tecnología. Analizar el volumen masivo de datos de esta manera sería lento e impreciso, con alta probabilidad de pérdida de información. Sin embargo, con el Big Data, actualmente es posible extraer, almacenar y clasificar automáticamente estos datos de manera eficiente.
Velocidad	La velocidad con la que se producen datos estimula que muchos de ellos se vuelvan obsoletos y pierdan relevancia rápidamente al ser superados por información más reciente.
Variedad	Los datos son generados por distintas fuentes por lo cual nace la necesidad de implementar diferentes formatos de recepción.
Veracidad	Es importante utilizar mecanismos basados en Big Data que permitan depurar los datos erróneos que se reciben, de esta manera se podrá obtener veracidad en la información a analizar.
Aspecto	Descripción

Valor

La eficacia que provea la información para las necesidades que se tengan de la misma, va a depender del valor que aporte, teniendo como premisa el cumplimiento de las "4V" antes tratadas.

Nota. En esta tabla se muestra los conceptos claves de las 5 "V" del Big Data.

Tipos de datos según su formato

Los datos pueden agrupar en tres categorías principales según su estructura: estructurados, no estructurados y semiestructurados, que a continuación se detallaran en la tabla 2 (Perez, 2020):

Tabla 2

Tipos de datos, según su estructura.

Tipo	Descripción
Estructurados	Los datos estructurados se encuentran ordenados y bien definidos, lo que facilita su análisis. Se suelen extraer de bases de datos o hojas de cálculo y se caracteriza por su simpleza en su almacenamiento y recuperación eficiente.
Semi Estructurados	Poseen una estructura delimitada, evitando la necesidad de que se encuentren ordenados. Requieren algún procesamiento adicional para obtener la información requerida, pero debido a su etiquetado estandarizado, la extracción se vuelve relativamente sencilla. Ejemplos de datos semiestructurados incluyen formatos como HTML, JSON o XM. Se caracteriza por tener cierto grado de flexibilidad y adaptarse a los cambios en su estructura.
No Estructurados	Los datos no estructurados son aquellos cuya forma o estructura es desconocida, generalmente de gran tamaño, lo que expone desafíos en su procesamiento para obtener valor. Un ejemplo típico incluye fuentes de datos heterogéneas que abarcan diversos formatos, como archivos de texto, imágenes y videos. Se caracterizan por ser difíciles en su análisis, mediante herramientas tradicionales debido a ello se hace necesario el uso enfoques más avanzados para el análisis.

Nota. En esta tabla se muestra los conceptos de cada tipo de clase de los datos.

La gestión de datos masivos en el ámbito de la Dirección de Talento Humano va más allá de simplemente buscar patrones o tendencias; implica asegurar que la información disponible sea

valiosa para la toma de decisiones. Por esta razón, es crucial comprender conceptos fundamentales como la minería de datos.

Minería de datos – Data Mining

La Minería de Datos es considerada la fase inicial de extracción de información, la misma que consiste en agrupar datos a través de procesos computacionales que buscan ser identificados dentro de un almacén de datos, con la finalidad de que aporte valor para las organizaciones. Siendo su esencia el análisis de extensos grupos de datos cuyo objetivo es visualizar tendencias o patrones que respalden las decisiones tomadas en la organización. Para el proceso en mención se hace uso de herramientas matemáticas, estadísticas y analíticas las cuales permiten realizar un proceso eficaz.

Beltrán (2018), menciona que la tecnología ha ido evolucionando, han ido desarrollando software y hardware idóneo para el procesamiento de grandes cantidades de información que tienen las empresas, las cuales han visto la necesidad de verse competitivos en el mercado. A tal efecto, las organizaciones han optado por nuevos métodos de análisis de la información, como lo es la minería de datos, debido a que este proceso busca identificar patrones, desviaciones o comportamientos atípicos de los datos analizados, los mismos que servirán para garantizar el proceso de implantación de planes estratégicos y toma de decisiones. A su vez, es valedero indicar que el proceso de minería de datos se ubica en el más alto de los procesos tecnológicos relacionados al análisis de los datos.

Mediante la minería de datos se puede deslumbrar patrones o tendencias ocultas que pueden encontrarse dentro de los extensos volúmenes de la Big Data. A través, del uso de instrumentos y herramientas particulares las organizaciones convierten los datos en conocimiento, los cuales son empleados por las mismas para decisiones a tomar, permitiendo potenciar el rendimiento organizacional. (AWS, s.f.)

Tras adquirir información valiosa del extenso ámbito de Talento Humano, es esencial reconocer que dicha información debe pasar por un ciclo de vida adaptado a las necesidades

de la organización. Por lo tanto, comprender en qué consiste este ciclo y sus diversas etapas se convierte en un aspecto de importancia fundamental.

Ciclo de vida de la información

Corresponde a una serie de pasos que los datos experimentan durante su vigencia, los cuales pueden ser clasificados en varios aspectos y en base a tareas específicas se puede determinar su avance. Este ciclo abarca desde la generación hasta la eliminación o reutilización de los datos en diferentes repositorios, y se denomina así porque los conocimientos adquiridos en un proyecto suelen servir de base para el siguiente, creando así una retroalimentación entre la última y la primera fase del proceso. (Narvaez, s.f.)

Para Mejias (2023), menciona que al ciclo de vida de los datos como un proceso recurrente que está compuesto de las siguientes fases: captura, guardado, utilización, archivo y borrado. El determinar un avance dentro de las fases deberá ser determinado en relación al cumplimiento de distintos requisitos, los mismos que son definidos por cada etapa. Este proceso relevante debido a que las organizaciones tienen la capacidad de utilizar y reutilizar la información como estimen pertinente buscando tener datos relevantes.

Etapas del ciclo de vida de los datos

Las etapas correspondientes al ciclo de vida de los datos son divididas, según diversos criterios. A continuación, se tratarán cada una de estas etapas:

Captura

Es la primera etapa del ciclo, en la cual se crea el dato en bruto que es aquel que se obtiene de diversas técnicas de recolección en diferentes formatos. Las técnicas por las cuales se pueden obtener son:

Adquisición. es cuando las empresas adquieren como su nombre lo dice, los datos a otras empresas, siendo de esta manera datos originados externamente.

Entrada. es cuando el personal de la empresa crea los datos de manera manual, siendo de esta manera datos originados internamente.

Creación. es cuando los datos de la organización han sido generados en los distintos procesos que realizan en sus funciones cotidianas en relación al giro del negocio.

Guardado

Esta etapa es una de las más delicadas debido a que en base al dato en bruto que ha sido obtenido se debe buscar la manera idónea de almacenarlos. A lo cual depende el tipo de dato para las distintas maneras de almacenaje:

Datos estructurados. estos datos poseen una estructura definida, siendo estos más accesibles para programas y humanos. El almacenamiento ideal se lo podría realizar en bases de datos relacionales, donde permiten organizarlos en tablas.

Datos no estructurados. estos datos no tienen una estructura definida por lo cual no tienen un modelo de datos definido. Por ende, un almacenamiento idóneo se lo podría realizar en una base de datos no relacional o No SQL.

Utilización

Dentro del ciclo de vida de los datos esta etapa se podría considerar como la más importante, debido a que se da una funcionalidad a la información, para lo cual, es imprescindible que a las personas en las que recaigan la toma de decisiones tengan acceso a la información.

Archivado

En esta etapa no es necesario ningún tipo de mantenimiento, más bien como su nombre lo dice es la encargada del archivo de la información una vez esta haya dejado de ser útil para la toma de decisiones dentro de la organización, pero se debe asegurar que en caso de que se necesite consultar dicha información esta pueda ser restaurada.

Borrado

Esta etapa es la final del ciclo, radica en la depuración y eliminación de los datos que han sido almacenados, estos datos deben haber cumplido cierto tiempo en el archivo y cerciorarse

que no vayan a ser utilizados en la empresa para poder cumplir esta fase. Lo que se debe cerciorar que una vez haya sido realizada la etapa de eliminación esta se elimine por completo.

Para que de esta manera esta información generada proporcione aporte al proceso de toma de decisiones relacionado a procesos de inteligencia de negocios. Este enfoque implica proporcionar una visión informada y estratégica que respalde la gestión.

Inteligencia de Negocios - Business Intelligence

La Inteligencia de Negocios (BI) es una línea preferencial que permiten modernizar los distintos procedimientos y estrategias organizacionales. Ya que, mediante la integración y análisis de la información, genera información útil, concreta y real, lo cual permite conocer datos reales para tener bases más confiables para toma de decisiones estratégicas, buscando de esta manera encontrar ventajas competitivas en el mercado, permitiendo de esta manera sobresalir en su nicho de negocio.

Para Giraldo (2022), en su trabajo de investigación menciona que la Inteligencia de Negocios agrupa diversos métodos para la recolección de información, buscando con esto recopilar, analizar, integrar y presentar datos significativos. El objetivo del BI radica en respaldar y simplificar la toma de decisiones en el ámbito empresarial. Además, facilita a las organizaciones el acceso a información crítica para el éxito en diversas.

Una correcta implementación de un mecanismo BI impulsa la generación de datos procesados, ofreciendo valiosas estrategias referente a las tendencias de la industria y generando un beneficio de enfoque en la toma de decisiones.

Por otro lado, Tableau (s.f.) manifiesta que, Business Intelligence minimiza los riesgos ante la toma de decisiones a través de un riguroso análisis de datos. partiendo de la minería de datos, posterior visualización de los mismo en conjunto con buenas prácticas y concluyendo con la generación de herramientas de datos. En la práctica, implica comprender a fondo los datos organizativos y utilizar esta información para promover cambios, eliminar ineficiencias y adaptarse ágilmente a las variaciones del mercado o la demanda.

Funcionamiento del Business Intelligence

Tableau (s.f.), explica el funcionamiento en que las organizaciones plantean preguntas y metas, y para evaluar su rendimiento y alcanzar dichas metas, recolectan, analizan datos relevantes y determinan acciones necesarias, explicándolo en el aspecto técnico los datos que se encuentran sin procesar y han sido recopilados de las actividades cotidianas del giro del negocio, son procesados y almacenados, los cuales pueden ser accedidos por los usuarios que lo necesiten para su análisis buscando cumplir las metas de la organización.

La información que ha sido seleccionada para formar parte de los procesos de inteligencia de negocios, debe ser resguardada en almacenes de datos. Estos se dividen en dos grupos que son Data Warehouse, cuyo enfoque es general de toda la organización, y el Data Mart cuyo enfoque es específico a un departamento. A continuación, se explorarán en detalle estos conceptos.

Data Warehouse

Zas (2018), en su estudio hace referencia a que un Data Warehouse busca un análisis óptimo para lo cual agrupa datos de múltiples áreas de la organización. Este agrupamiento se considera un punto de partida para procesos de inteligencia empresarial.

Un Data Warehouse, se caracteriza por ser un almacén de información, dando resguardo a la información que proviene de distintas fuentes (Data Mart y Fuentes Externas). Brindado de esta manera una disponibilidad inmediata a la información, a través de los almacenes de datos.

Por otro lado, Ávila, C. & Chiquito, J. (2022) expresa que un Data Warehouse corporativo, permite el almacenaje de extensas cantidades de datos, los mismo que fueron recolectados e integrados desde fuentes diversas. Es fundamental que estos datos mantengan un formato coherente y sean de fácil acceso, permitiendo a las organizaciones realizar análisis y generar informes. La información contenida dentro de los Data Warehouse juega un papel preponderante en la toma de decisiones, ya que centralizan y garantizan la disponibilidad de la información.

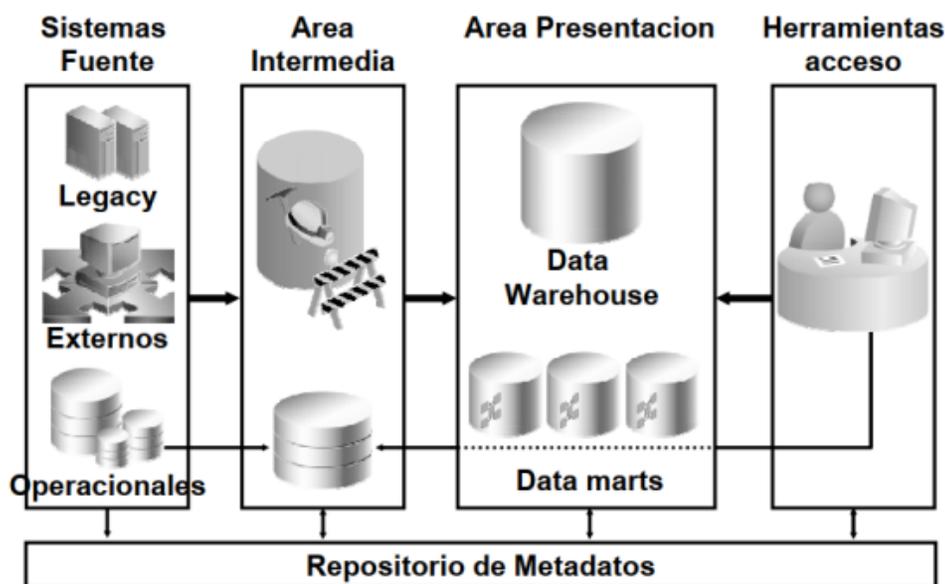
Los Data Warehouse tienen la capacidad de resguardar y almacenar enormes volúmenes de información, los cuales para tener una mejor interpretación se busca que tengan un formato estándar. Además, pueden integrar información de diversas fuentes y estandarizarla para facilitar su análisis. Permitiendo mantener registros históricos, que pueden ser comparados de ser necesarios. Asimismo, garantizan la seguridad de los datos al almacenarlos en un único lugar, permitiendo el acceso solo a personas autorizadas que necesitan información específica.

Componentes Data Warehouse

Los componentes de un Data Warehouse se muestran en la figura 2.

Figura 2

Componentes de un Data Warehouse



Nota. La figura representa los elementos de un Data Warehouse. Tomado de "Implementación de un Data Mart para la ayuda en la toma de decisiones de la Gerencia de Talento Humano", por Rossana Traiman, 2021, Universidad Nacional de Rio Negro

Traiman (2021), en su estudio manifiesta que los componentes de un Data Warehouse son: sistemas fuente, área intermedia, área presentación, herramienta acceso y repositorio de metadatos. Los mismos que serán desarrollados a continuación:

Sistemas fuentes. Mecanismo mediante el que se recopila la información de múltiples orígenes para resguardarlos en el Data Warehouse. Estas fuentes pueden ser internas, como las bases de datos de sistemas operativos, incluso sistemas desactualizados para datos históricos, así como también alimentarse de información de fuentes externas.

Área intermedia. Radica en ser un espacio intermedio entre las fuentes de información y el Data Warehouse, buscan resguardar la información de la organización y mantenerla disponible su requerimiento. Su principal función, es permitir la consecución de información mediante la transmutación de los datos, sin que esto afecte a los sistemas y procesos originales del giro del negocio.

Área de presentación. En este componente las bases de datos son indispensables, debido a que es el lugar donde reposará la información a ser utilizada en los distintos procesos.

Herramientas de acceso. Proceso en el cual a través del Data Warehouse o Data Mart se provee de información sencilla de comprender hacia los usuarios finales. A través de consultas, el usuario accede al área de presentación y visualiza los resultados en un formato práctico para su trabajo.

Repositorio de metadatos. Corresponde a la descripción del contenido de manera clara que se encuentra en un Data Warehouse, incluyendo la ejemplificación de consultas a la base de datos. Este repositorio de metadatos debe ser la única fuente de documentación del Data Warehouse y debe ser accesible desde cualquier componente del sistema.

Data Mart

Un Data Mart se enfoca en específico a una unidad organización para almacenar la información que en esta se genere. Se trata de una clasificación específica de datos extraídos de un sistema más extenso, lo que facilita un análisis eficaz de la información departamental. Cuya finalidad principal es la de ofrecer información concisa que sea de utilidad en la toma de decisiones por los miembros de las organizaciones. (AWS, s.f.)

Para Acosta (2023), manifiesta en su estudio como un repositorio de información enfocado a una determinada unidad de la organización. Buscando ofrecer acceso independiente a datos relevantes para que diferentes departamentos puedan examinar y extraer información específica según sus necesidades. Su objetivo principal es facilitar la toma de decisiones más informada y personalizada, centrándose en una línea de negocio o equipo y extrayendo datos de una fuente determinada.

Un Data Mart radica en ser una parte de un todo, es decir es una parte de un repositorio de datos centralizado o Data Warehouse buscando de esta manera satisfacer las necesidades particulares de las áreas al estar la información clasificada. A continuación, en la tabla 3 se evidencia la clasificación de los Data Mart, según la fuente que lo abastezca.

Tabla 3

Clasificación de los Data Mart, de acuerdo a la fuente de obtención de los datos.

Clasificación del Data Mart	Descripción
Dependiente	Tiene una única fuente de datos, ya que reciben los datos para su uso desde un Data Warehouse corporativo.
Independiente	No depende de un Data Warehouse corporativo centralizado, más bien se alimenta de la información recopilada de los sistemas transaccionales o sistemas operacionales.
Hibrido	Tiene diversas fuentes de datos como su nombre lo dice híbrido, ya que puede recibir datos de sistemas transaccionales o sistemas operaciones y de Data Warehouse centralizados de las organizaciones.

Nota. En esta tabla se muestra los conceptos de cada tipo de Data Mart de acuerdo a la fuente de obtención de la Data.

La clasificación de los diversos tipos de fuentes de información que nutren los Data Mart pueden traer consigo ventajas y desventajas, que se describirán en la siguiente tabla 4 (Santos, s.f.):

Tabla 4

Ventajas y desventajas de la clasificación de Data Mart según la fuente de obtención de la Data.

Tipos	Ventajas	Desventajas
Dependiente	Su actualización continua y su integración constante en la base de datos general, evitando la necesidad de utilizar sistemas de almacenamiento de datos separados.	Dependen de la base de datos principal corren el riesgo de volverse inoperables en caso de fallos o pérdida de conexión. Además, su operatividad generalmente está condicionada a una conexión a internet.
Independiente	Este método de almacenamiento permite tener un manejo completo de la información y es una elección destacada para aquellas empresas que optan por retener sus datos de manera interna.	Al migrar los datos a un repositorio de datos general, es importante que estos mantengan un formato estándar para que nos convierta en un proceso laborioso.
Híbrido	Integrar datos locales y centralizados en un Data Mart híbrido posibilita mejorar el rendimiento en consultas y análisis específicos al reducir la latencia y aumentar la eficiencia.	Garantizar la consistencia entre los Data Mart locales y el almacén de datos central en un enfoque híbrido puede presentar dificultades, lo que afectaría la integridad de informes y análisis.

Nota. En esta tabla se muestra las ventajas y desventajas de los distintos tipos de Data Mart de acuerdo a la obtención de la Data.

Estructura de un Data Mart

Un Data Mart es capaz de almacenar datos de tipo transaccional debido a que es un tipo de base de datos relacional, lo cual facilita el acceso y la interpretación de la información que ahí se almacena. El arquetipo es capaz de conglomerar datos históricos, lo que hace factible la visualización de patrones y tendencias. Los Data Mart pueden ser estructurados mediante la utilización de esquemas multidimensionales como lo son diseños de: estrella y copo de nieve buscando satisfacer las necesidades de los usuarios, para lo cual se detallan a continuación los tipos de diseños (IBM, s.f.):

Tabla 5*Estructura de los Data Mart*

Tipo de estructura	Descripción
Estrella	Representa a un esquema lógico de tablas dentro una base de datos multidimensional. En este esquema, se forma de una tabla central (Hechos) y tablas que lo rodea (dimensiones) La falta de dependencia reduce las consultas, haciendo que los diseños de tipo estrella sean considerados una opción eficaz para el acceso a extensos conjuntos de datos.
Copo de nieve	Es un esquema ampliado al de tipo estrella debido a que aumentan ramificaciones que lo rodean (dimensiones). Estas tablas se normalizan, para que de esta manera prevalezca integridad de los datos y minimizar la redundancia. Cabe señalar que cuyo mantenimiento puede ser complejo.

Nota. En esta tabla se muestra las ventajas y desventajas de los distintos tipos de Data Marts de acuerdo a la obtención de la Data.

Tipos de data Mart

Es relevante destacar que OLAP (Procesamiento Analítico en Línea) y OLTP (Procesamiento de Transacciones en Línea) son dos categorías SBD (Sistemas de Bases de Datos) cuyas finalidades no son las mismas. A continuación, en la siguiente tabla se detallará cada una de estas (Ávila, C. & Chiquito, J.,2022):

Tabla 6*Tipos de Data Mart*

OLAP (Procesamiento Analítico en Línea)	<p>Conceto: Los Data Mart OLAP se caracterizan por la utilización de cubos OLAP, mediante la incorporación de cubos relacionales, satisfaciendo los requerimientos específicos de cada área. El proceso de creación, utilización y mantenimiento de estos cubos varía según la herramienta empleada.</p> <p>Propósito: Es utilizado para generar informes que están basados en datos históricos que están almacenados en un medio OLAP.</p> <p>Características: Proporciona una vista rápida a los usuarios de grandes conjuntos de datos de una manera eficiente. Posee datos resumidos que se extraen de sistemas de tipo OLAP.</p>
------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

	Concepto: Los Data Mart OLTP se constituyen de partes del Data Warehouse, los cuales mediante mecanismos como el filtrado y agregación son optimizados mejorar su eficiencia. Se ajustan a las necesidades específicas de cada organización.
OLTP (Procesamiento de Transacciones en Línea)	Propósito: Esta proyectado para receptor transacciones operativas diarias y realizadas en tiempo real su procesamiento.
	Características: Admite altos niveles de concurrencia, es decir múltiples usuarios simultáneamente. Permitiendo realizar consultas simples y rápidas con la finalidad de recuperar registros específicos.

Nota. En esta tabla se muestra los tipos de Data Mart como OLAP y OLTP, describiendo su concepto, propósito y características.

Para adquirir información valiosa, es necesario seguir un proceso ETL, el cual debe implementarse a través de una metodología detallada en su respectivo capítulo. Comprender teóricamente el ETL es fundamental para asegurar la eficacia de dicho proceso.

Diferencias entre Data Warehouse y Data Mart

Una vez revisados ambos temas, se elaborará una tabla comparativa que resalte las distintas características entre un Data Warehouse y un Data Mart, presentándola de manera clara y concisa, en la siguiente tabla:

Tabla 7

Comparativa entre Data Warehouse y Data Mart

Criterio	Data Warehouse	Data Mart
Alcance	Es enfocado a satisfacer las necesidades de toda la organización.	Es enfocado a satisfacer las necesidades de determinada área de la organización.
Uso	Decisiones estratégicas	Decisiones tácticas
Administración	Toda la organización	Determinada área
Objetivo	Optimizar la administración	Optimizar reportes de soporte para la administración
Fuentes	Generalmente la información es recopilada de varias fuentes	Generalmente la información es recopilada de pocas fuentes
Tamaño	Datos históricos a nivel atómico	Resúmenes y datos totalizados

Criterio	Data Warehouse	Data Mart
Diseño	Complejo	Sencillo
Tiempo implementación	Meses a años	Pocos meses

Nota. Adaptado de "Inesem Bussines Intelligence: Descubre la principal diferencia entre Data Mart y Data Warehouse, por Víctor Acosta, 2023, <https://www.inesem.es/revistadigital/informatica-y-tics/diferencia-entre-data-mart-y-data-warehouse/> (Acosta, 2023)"

Se puede concluir de la comparativa realizada en la tabla 7, que el conjunto de Data Mart forman un Data Warehouse. Por otra parte, los Data Mart abarcan la información de una unidad específica mientras que el Data Warehouse abarca la información de la organización en su totalidad, es valedero mencionar que este conlleva mayor tiempo su implantación, por ende es más compleja.

Para comprender los tópicos necesarios para el desarrollo del trabajo de integración curricular, es de suma importancia conocer acerca del proceso ETL, el cual es crucial en el presente desarrollo. Para lo cual se tratará el tópico a continuación.

ETL

Para IBM (s.f.), menciona que este mecanismo es la integración de información obtenida de diversos orígenes, consolidándolos en un único repositorio de datos mediante las fases de extracción, transformación y carga. Dicho proceso ETL es esencial para el análisis de datos y Machine Learning. Limpia y organiza datos con reglas comerciales para satisfacer necesidades de inteligencia empresarial, como informes, y mejorar procesos internos.

El proceso ETL radica en tres etapas cuyo objetivo principal de este proceso es facilitar el movimiento de la data y su transformación para ser utilizada según las necesidades del giro del negocio. A continuación, se detallará cada una de las etapas que comprende este proceso:

Extracción

La etapa de extract o en español extracción es la inicial del proceso ETL como se puede apreciar en la figura 3, la cual radica en la obtención de la data de distintas fuentes, la data en esta etapa se la puede obtener de una o varias fuentes.

Transformación

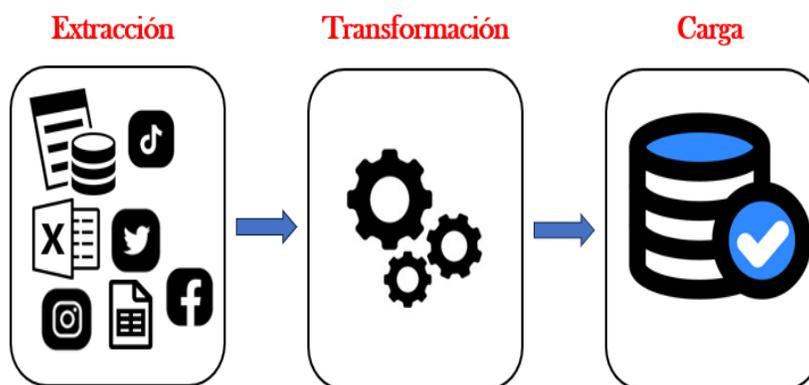
La etapa de transform o en español transformación es la etapa intermedia del proceso ETL como se puede observar en la figura 3. En esta etapa se realiza la transformación de la data mediante procesos como: limpieza de datos, verificación del tipo de datos, revisión de datos duplicados, entre otros; estas transformaciones se las realizan en base a las necesidades que tenga el giro del negocio.

Carga

La etapa de load o en español carga es la final del proceso ETL como se puede apreciar en la figura 3, la cual consiste en el almacenamiento de la data que ha sido transformada en una base de datos, Data Warehouse o Data Mart, la cual servirá para apoyar en la toma de decisiones empresariales.

Figura 3

Esquema de un proceso ETL.



Nota. En la gráfica se visualiza una arquitectura de un mecanismo ETL, desde la obtención de la Data, su transformación y la carga en un almacén de datos.

Características ETL

El proceso ETL constituye la base fundamental de un almacén de datos (Data Warehouse, Data Mart) pero va más allá de un término netamente técnico, siendo también una filosofía de gestión de datos. Un mecanismo ETL es capaz de la extracción de datos de múltiples fuentes para tratarlos buscando mejorar su calidad y sean de utilidad para la toma de decisiones informada.

Este proceso sobrepasa la extracción y carga de información, debido a que, este proceso es encargado de:

- Eliminar inexactitudes y rectificar la ausencia de datos.
- Corregir imprecisiones y solucionar la falta de información.
- Monitorear el tráfico de información generado en las distintas transacciones.
- Estandarizar los datos recolectados mediante proceso de transformación
- Organizar los datos de manera que generen valor.

El proceso de ETL es ajustable y actualizable en relación a los requisitos comerciales de la organización, es decir, a las metas que se pretenden alcanzar para los cuales se busca consolidar la información en el almacén de datos. Dado que un almacén de datos se destina exclusivamente al análisis, es importante considerar el tipo de información que se pretende obtener. (Sepúlveda, 2018)

Capítulo III: Herramientas

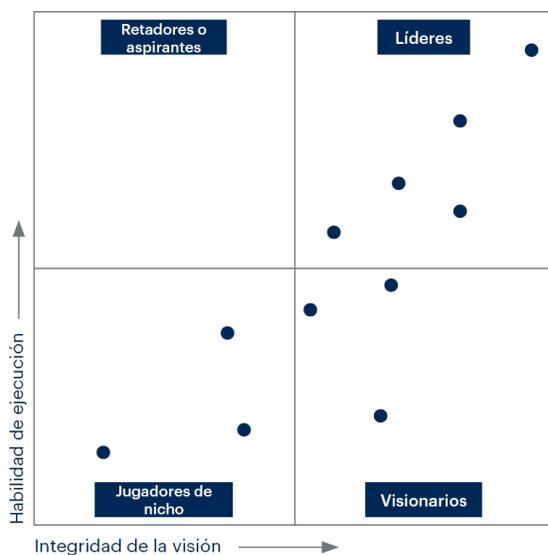
El presente capítulo del trabajo de integración curricular, comprende un análisis exhaustivo de las herramientas que se emplearán tomando como apoyo para la decisión el Cuadrante Mágico de Gartner. Este análisis tiene como propósito generar una comparativa entre ellas, con el objetivo de identificar las herramientas más adecuadas para su implementación en el proyecto.

Cuadrante Mágico de Gartner

El Magic Quadrant de Gartner o en español conocido como Cuadrante Mágico de Gartner es un resumen ejecutivo donde se presenta una investigación realizada en distintos nichos de mercado, mostrando de esta manera las tendencias en los distintos competidores. La cual se da a través de representaciones gráficas y criterios de evaluación consistentes, facilitando la rápida evaluación de los distintos proveedores.

Figura 4

Cuadrante Mágico de Gartner



Nota. En la figura hace referencia a la distribución del "Cuadrante Mágico de Gartner". Tomado de Gartner (s.f.).

Recuperado de <https://www.gartner.es/es/metodologias/magic-quadrant>

Funcionamiento

El Cuadrante Mágico de Gartner es un paso inicial para evaluar proveedores tecnológicos en el contexto de una posible implementación. El Cuadrante Mágico de Gartner emplea un gráfico con ejes vertical y horizontal para expresar el conocimiento de mercado y la capacidad de ejecución, respectivamente. El cuadrante se divide en cuatro categorías como se puede evidenciar en la figura 4:

- Líderes: Actualmente bien posicionados y con una visión robusta para el futuro.
- Visionarios: Comprenden las tendencias del mercado, pero su capacidad de ejecución es limitada.
- Jugadores de nicho: No poseen de manera significativa innovación.
- Retadores o aspirantes: Se caracteriza por su desempeño sólido, pero no acordes a las tendencias del mercado. Gartner (s.f.)

Herramientas para la integración de datos

Gartner en su informe correspondiente al año 2023, respecto a las herramientas de integración de datos presenta un resumen el cual puede ser apreciado en la figura 5.

Figura 5

Cuadrante Mágico Gartner Integración de Datos



Nota. En la gráfica se observa la distribución de las empresas dedicadas a la Integración de Datos en el cuadrante mágico de Gartner. Gartner, 2023.

Para el presente trabajo se realizará el análisis de algunas de las herramientas mostradas en el Cuadrante de Gartner que se visualiza en la figura 5. Sin embargo, es importante tomar en cuenta que el hecho de encontrarse en el segmento de líderes del Cuadrante Mágico de Gartner no deslumbra que sean los mejores. Bajo esta premisa las herramientas a analizar son:

Talend

Esta plataforma de código abierto (Open Source) ofrece una solución integral para la gestión de bases de datos mediante la aplicación del proceso ETL. Brindando soporte en diversos formatos en los que se encuentre la información sean estos: JSON, XML, TXT, PDF, entre otros.

Requisitos

Los requisitos técnicos a tomar en cuenta para el uso de esta herramienta son:

Sistema Operativo

- Apple MacOS en sus versiones Big Sur 11, Catalina 10.15.
- Linux en sus distribuciones Ubuntu 20.04, CentOS 8, Debian 10, entre otros.
- Windows va bien en versiones de Windows 11, 10, Server 2016 y Server 2019.

Memoria

- Indispensable 3 GB – 4 GB

Disco

- Disco para instalación 3 GB
- Disco para uso más de 3 GB recomendado

Nota: Se debe tomar en cuenta que el espacio en disco sea el doble de la información que se desea transaccionar.

Entorno Java

- OpenJDK 11

- Oracle JDK 11

Pros

- Incluye un framework propio que facilita el desarrollo de componentes personalizados.
- Proporciona componentes para casi todas las acciones requeridas.
- Utiliza componentes genéricos que posibilitan la conexión con prácticamente cualquier plataforma, aunque no de manera instantánea.
- Cuenta con documentación y una comunidad extensas y bien organizadas.
- Ofrece una solución integral, eliminando la necesidad de múltiples herramientas y configuraciones adicionales.

Contras

- Su instalación puede ser compleja.
- Debido a las diversas funcionalidades, la curva de aprendizaje es grande.
- Su correcto funcionamiento puede requerir el consumo de amplios recursos del equipo.
- Tiene restricción de compatibilidad con algunos formatos. (Qué es Talend. Pros y Contras, 2022)

IBM InfoSphere DataStage

InfoSphere DataStage es una herramienta que brinda la facilidad de integrar datos mediante el flujo de fuentes diversas. Esta herramienta posibilita la interconexión entre aplicaciones empresariales, lo cual permite resolver problemas a gran escala, mediante el procesamiento de extensos volúmenes de datos, que pueden ser realizados en tiempo real, según las necesidades del proyecto. (IBM, 2021)

Requisitos

Para el uso de esta herramienta es importante tomar en cuenta los siguientes requisitos:

Sistema Operativo

- Compatible con Sistemas Operativos como: Windows, distribuciones Linux y AIX.

Memoria

- Indispensable 4 GB

Disco

- Indispensable 10 GB (variable según las funcionalidades requeridas)

Procesador

- Indispensable 2 CPU. (IBM, s.f.)

Pros

- Tiene la capacidad de manejar extensos volúmenes de datos.
- Brinda la capacidad de conexión con múltiples fuentes de extracción de información, así como también diversos almacenes de datos para la carga de la información.
- Capacidad de trabajo con datos sumamente complejos.
- Gestión de mecanismo ETL centralizado.
- Soporte y mantenimiento continuo por parte del fabricante.

Contras

- Los costos asociados al licenciamiento, mantenimiento y hardware para la implementación pueden ser elevados.
- Presenta una curva de aprendizaje pronunciada debido a la complejidad de la plataforma.
- Existe una dependencia del fabricante en temas de soporte al ser una solución propietaria.
- Para solventar temas de datos masivos, requiere de hardware especializado.
- Puede que no se adapte a los requisitos propios de las organizaciones. (IBM, 2021)

Python

Python es un lenguaje de programación el cual es conocido por su sintaxis de fácil comprensión y su versatilidad. El mismo que se lo puede utilizar para un sin número de propósitos como son: integración de datos, machine learning, entre otros. Mediante el uso de la amplia gama de librerías que tiene a disposición es capaz de formar scripts eficientes y robustos que permiten cumplir con los requisitos planteados en el desarrollo. (Rodríguez F. , s.f.)

Requisitos

Para el uso de la herramienta es indispensable tomar en cuenta los siguientes requerimientos:

Sistema Operativo

- Python es compatible con: MacOS, Windows y las distribuciones Linux.

Memoria

- Indispensable 4 GB
- Recomendable 8 GB

Disco

- Mayor a 20 GB (Escalante, 2023)

Pros

- Python puede ser una opción adecuada para procesos ETL.
- Python tiene una amplia comunidad, con nutrida documentación.
- Python tiene una síntesis de fácil comprensión, facilitando el desarrollo de scripts.
- Python tiene librerías que pueden ser para diversos fines como desarrollo web, tratamiento de datos, entre otros.
- Python es compatible con lenguajes de programación como C y C++.
- Python es multiplataforma, lo cual radica en que el programa desarrollado en Python puede ser ejecutado en múltiples sistemas operativos.

Contras

- Puede presentar inconvenientes en procesos a gran escala.
- Cuando se trabaja con bibliotecas externas puede tener complejidad en cuanto a dependencias.
- Python presenta una desventaja frente a lenguajes de tipo compilados.
- En temas de desarrollo móvil Python no es la solución más óptima. (KeepCoding, 2023)

Herramientas para el almacenamiento de datos

El proceso de subida de datos necesita obligatoriamente un lugar el cual será destinado como un almacén de datos, a continuación, se detallan algunas opciones para proceder con el almacenaje:

MySQL

MySQL es una herramienta de código abierto de propiedad de Oracle Corporation, destinada a la gestión de bases de datos relacionales. La cual es conocida por su escalabilidad y facilidad de uso, la cual la hace muy utilizada debido a la naturaleza de gratuidad, lo cual hace accesible a todo el que la requiera. Como una de sus características se puede mencionar la capacidad de manejar múltiples usuarios concurrentes. (Ibarra, 2023)

Requisitos

Para la utilización de esta herramienta es importante tomar en cuenta los siguientes requisitos:

Sistema Operativo

- Compatible con: MacOS, Windows y distribuciones Linux.

Memoria

- Indispensable 1 GB

Disco

- Indispensable 512 MB. (MySQL)

Pros

- MySQL al ser una herramienta de código abierto permite a los usuarios gestionar su uso de descarga, uso y modificación.
- La curva de aprendizaje de la herramienta es baja debido al intuitivo lenguaje de consultas.
- Tiene la capacidad de manipular extensos volúmenes de datos sin perder la disponibilidad de los mismos.
- Cuenta con una amplia comunidad, donde se obtiene nutrida documentación.

Contras

- Presenta inconvenientes al momento de realizar consultas extremadamente complejas.
- En su versión gratuita puede carecer de algunas funcionalidades avanzadas.
- Requiere una configuración cuidadosa, para obtener un correcto funcionamiento.

(Ibarra, 2023)

PostgreSQL

PostgreSQL, es un recurso utilizado para la gestión de bases de datos, la cual fue creada en la Universidad de California por Michael Stonebraker bajo la premisa Open Source. Siendo es un software gratuito no ha dejado de lado la capacidad de gestión de bases de datos complejas mediante un rendimiento óptimo. (OVHcloud)

Requisitos

Para el uso de la herramienta se debe tomar en cuenta los siguientes requisitos:

Sistema Operativo.

- Es compatible con sistemas operativos como: MacOS, Windows y distribuciones de Linux.

Memoria

- Indispensable 1 GB

Disco

- Mas de 512 MB (Vargas, 2017)

Pros

- Tiene la capacidad de adaptación a las especificaciones de cada equipo, lo cual permite la optimización de los recursos.
- Brinda una interfaz gráfica, lo cual hace que la administración sea intuitiva.
- Se caracteriza por garantizar el cumplimiento el estándar ACID (Atomicidad, Consistencia, Aislamiento, Durabilidad)
- Brinda la capacidad de crear extensiones propias o a su vez usar extensiones desarrolladas por la comunidad.

Contras

- Puede presentar problemas de lentitud en base de datos pequeñas, debido a que se encuentra optimizado para bases de datos extensas.
- Ofrece soporte mediante foros en línea.
- La curva de aprendizaje puede ser altas para quien no tenga noción en lenguaje SQL. (Segovia, 2018)

Oracle Database

Oracle Database recursos desarrollado por la prestigiosa organización Oracle, el cual se caracteriza por ser un gestor de base de datos, el mismo emplea el lenguaje SQL para la administración. Este gestor se enfoca en temas gestión y seguridad de los datos almacenados, para lo cual limita el acceso a agentes no autorizados. (Oracle Database: definición y funcionamiento, 2022)

Requisitos

Para el uso de esta herramienta se deben tomar en cuenta los siguientes requisitos:

Sistema Operativo

- Compatibilidad con sistemas operativos como: MacOS, Windows y distribuciones Linux.

Memoria

- Indispensable 1 GB
- Recomendable 2 GB

Disco

- Indispensable 512 MB
- Recomendable 1 GB (Oracle Corporation)

Pros

- Capacidad de gestión de tareas complejas con grandes conjuntos de datos.
- Mediante mecanismo sólidos busca proteger la integridad de la información que resguarda.
- Ofrece soporte de usuario de alto nivel.
- Tiene una extensa documentación.
- Dispone de funcionalidades avanzadas en cuanto a la gestión de bases de datos se refiere.

Contras

- Al ser de tipo licenciada hace que para implementaciones pequeñas represente desafíos económicos.
- Debido a las extensas y avanzadas funcionalidades la curva de aprendizaje es amplia.
- La configuración inicial de la herramienta puede ser compleja.
- Dependencia del fabricante para el soporte a los usuarios. (Oracle RDBMS, 2022)

Herramientas para la visualización de datos

Gartner en su informe anual correspondiente al año 2023 referente a herramientas tecnológicas para la visualización de datos presenta un resumen el cual se puede observar en la figura 6.

Figura 6

Cuadrante Mágico Gartner Business Intelligence



Nota. En la gráfica se observa la distribución de las distintas empresas dedicadas a ofrecer servicios de análisis de datos correspondiente al año 2023. Gartner, 2023.

Para el presente trabajo se realizará el análisis de algunas de las herramientas presentadas en el Cuadrante de Gartner en su versión 2023, el cual se muestra en la figura 6. Cabe indicar que el hecho de que se encuentre en el segmento de líderes dentro del mencionado Cuadrante no quiere decir que sean los mejores. Las herramientas que se analizarán serán las siguientes:

Power BI

Microsoft Power BI conocida herramienta integral usada para la inteligencia de negocios, mediante el análisis de los datos, provenientes de diversas fuentes. Esta herramienta tiene la capacidad de generar presentar los datos analizados de manera visual a través de informes

avanzados. Es importante acotar que Power BI Desktop es de tipo gratuita, restringiendo el acceso a ciertas funcionalidades.

Requisitos

Para el uso de la herramienta es imprescindible tomar en cuenta los siguientes requisitos:

Sistema Operativo

- Compatible con sistemas operativos como: Windows 8.1, Server 2012 R2.

Memoria

- Indispensable 2 GB
- Recomendable 4 GB

Disco

- Indispensable 1 GB

Complementos

- Net 4.7.2 (Microsoft Learn, s.f.)

Pros

- Presenta una interfaz gráfica intuitiva al usuario.
- Tiene interconexión con una extensa cantidad de herramientas.
- Posee visualizaciones intuitivas, lo que hace que los informes sean de mejor calidad,
- Garantiza que los datos sean precisos mediante su funcionalidad de actualización al instante.

Contras

- Para aprovechar las diversas funcionalidades de la herramienta, se requiere de conocimiento previo.
- Si se maneja volúmenes extensos de información, estos pueden implicar gastos en licenciamiento.
- Es indispensable contar con internet, para aprovechar todas las funcionalidades que ofrece la herramienta, (Gil M. B., s.f.)

Tableau

Tableau es una herramienta de Inteligencia de negocios que posibilita la interpretación de los datos contenidos, de una manera visual e intuitiva, buscando de esta manera que los datos sean aprovechados en los procesos estratégicos de las organizaciones. (Tableau, s.f.)

Requisitos

Los principales requisitos técnicos que deben tener los usuarios de esta herramienta son:

Sistema Operativo

- Apple MacOS es soportado en sus versiones Big Sur 11.4, Catalina 10.15 y versiones posteriores.
- Windows es soportado en versiones de Windows 11, 10, 8 y 8.1.

Memoria

- Indispensable 2 GB

Disco

- Recomendado 1.5 GB

Documentación

- La documentación correspondiente se encuentra disponible en inglés, francés, alemán, italiano, español, portugués, entre otros. (Tableau, s.f.)

Pros

- Posee una interfaz del aplicativo intuitivo, brindado de esta manera a los usuarios la oportunidad de realizar el análisis de datos de manera eficaz.
- Permite conectarse a diversas fuentes (bases de datos, hojas de cálculo y servicios en la nube), lo cual permite consolidar información en un único repositorio.
- Facilita la realización de análisis de datos de forma interactiva.
- Posibilita la colaboración y el intercambio en tiempo real entre usuarios, simplificando el trabajo en equipo.

Contras

- Depende de licenciamiento para obtener todas las funcionalidades, por ende, incurre en gastos que pueden ser elevados.
- Debido a la amplia gama de funciones la curva de aprendizaje es elevada.
- En el análisis de grandes volúmenes de datos puede enfrentar problemas en cuanto al rendimiento y velocidad,
- Para casos complejos, depende del uso de lenguajes de programación u otras herramientas, lo cual hace que se deba tener conocimientos amplios en programación. (Gil M. , s.f.)

Qlik

Qlik es una herramienta idónea para la inteligencia de negocios debido a que tiene la capacidad de tratamiento de los datos convirtiendo a los mismos en información valiosa, que puedan injerir en la toma de decisiones, ofreciendo la información procesada de una manera eficiente. (Marín, 2023)

Requisitos

Para el uso de la herramienta se debe tomar en cuenta los siguientes requisitos:

Sistema Operativo

- Compatible con Windows 10 y 11, Windows Server 2012, Server 2012 R2, Server 2016, Server 2019 y Server 2022.

Memoria

- Indispensable 4 GB

Disco

- Recomendable 900 MB para la instalación. (Qlik, 2023)

Pros

- Las necesidades de hardware para el funcionamiento de la herramienta son mínimos.

- Presenta de una manera intuitiva y rápida la información analizada.
- No es indispensable contar con conocimientos previos.
- Se caracteriza por brindar respuestas instantáneas.

Contras

- Una limitación radica en el análisis en tiempo real.
- En cuanto al uso de memoria RAM del dispositivo puede ser un limitante a la hora del tratamiento de los datos.
- Para el desarrollo de aplicaciones potentes, se debe tener conocimientos técnicos.

(Marín, 2023)

Criterios de selección de las herramientas

Herramienta de integración de datos

Una vez revisado a detalle las distintas herramientas de integración de datos propuestas en el Capítulo III: Herramientas, en la sección Herramientas para la integración de datos, para realizar una correcta elección de la herramienta a utilizar es indispensable realizar una comparativa la cual se muestra en la tabla 8.

Tabla 8

Comparativa Herramientas Integración de Datos

Característica	Python	IBM InfoSphere DataStage	Talend Data Integration
Multiplataforma	Sí	Sí	Sí
Licenciamiento	Open Source	Propietario	Open Source /Propietario
Costo	Bajo	Alto	Variable
Interfaz	No gráfica	Gráfica	Gráfica
Soporte	Amplia comunidad y documentación en línea	Soporte propietario y comunidad	Comunidad activa y soporte de la empresa
Curva de Aprendizaje	Moderada	Alta	Moderada

Característica	Python	IBM InfoSphere DataStage	Talend Data Integration
Capacidad de Personalización	Alta	Limitada	Alta

Nota. En esta tabla evidencia la comparativa entre los aspectos más importantes de las herramientas de Integración de Datos.

Luego de haber realizado y analizado la comparativa de herramientas en la tabla 8, se ha seleccionado la herramienta Python para realizar el proceso ETL, al ser una herramienta Open Source y de requisitos de hardware y software mínimos y entendido la situación del área de Talento Humano se la identificado como la herramienta óptima para el desarrollo del proyecto.

Herramienta de almacenamiento de datos

Una vez analizadas las distintas herramientas de almacenamiento de datos propuestas en el Capítulo III: Herramientas, en la sección Herramientas para el almacenamiento de datos, para realizar una correcta elección de la herramienta a utilizar es indispensable realizar una comparativa la cual se muestra en la tabla 9.

Tabla 9

Comparativa Herramientas Almacenamiento de Datos

Aspectos	MySQL	PostgreSQL	Oracle Database
Curva de aprendizaje	Baja	Moderada	Alta
Tipo de licencia	Código abierto y propietario	Código abierto y propietario	Comercial y licencia propietaria
Lenguajes de Programación Soportados	C, C++, Java, Perl, Python, Ruby, etc.	C, C++, Java, Perl, Python, Ruby, etc.	C, C++, Java, .NET, PL/SQL, etc.
Escalabilidad	Horizontal y vertical	Horizontal y vertical	Horizontal y vertical
Soporte de Clustering	Sí	Sí	Sí
Herramientas de Administración	MySQL Workbench, phpMyAdmin	pgAdmin, psql	Oracle Enterprise Manager, SQL Developer
Integración con lenguajes de programación web	Excelente	Excelente	Excelente

Aspectos	MySQL	PostgreSQL	Oracle Database
Comunidad y Soporte	Gran comunidad de usuarios y soporte de la comunidad	Gran comunidad de usuarios y soporte de la comunidad	Soporte oficial y gran cantidad de recursos de la comunidad
Costo	Gratuito	Gratuito	Comercial

Nota. En esta tabla se muestra una comparativa de los aspectos más importantes de las herramientas de almacenamiento de Datos.

Para el presente trabajo una vez realizado la comparativa de aspectos importantes de las herramientas de almacenamiento como se muestra en la tabla 9, tomando en cuenta la situación del área de Talento Humano se ha elegido a MySQL como la herramienta a utilizar en el almacenamiento de la información del presente proyecto.

Herramienta de visualización de datos

Una vez analizada a profundidad las distintas herramientas de visualización de datos propuestas en el Capítulo III: Herramientas, en la sección Herramientas para la visualización de datos, para realizar una correcta elección de la herramienta a utilizar es indispensable realizar una comparativa la cual se muestra en la tabla 10.

Tabla 10

Comparativa Herramientas de Visualización

Aspecto	Power BI Desktop	Qlik Sense	Tableau
Desarrollador	Microsoft	Qlik	Tableau Software
Licenciamiento	Comercial / Gratis	Comercial / Gratis	Comercial / Gratis
Interfaz de Usuario	Alta	Alta	Alta

Aspecto	Power BI Desktop	Qlik Sense	Tableau
Análisis en Tiempo Real	Sí	Sí	Sí
Soporte	Comunidad extensa	Comunidad extensa	Comunidad extensa
Visualización de datos	Excelente	Excelente	Excelente
Integración con Otras Herramientas	Alta	Alta	Alta
Costo	Variado, incluye versión gratuita con limitaciones	Variado, incluye versión gratuita con limitaciones	Variado, incluye versión gratuita con limitaciones
Curva de aprendizaje	Baja	Baja	Moderado

Nota. La tabla muestra una comparativa de aspectos importantes de las herramientas analizadas con respecto a la Visualización de Datos analizadas.

Para el presente trabajo una vez realizado la comparativa de aspectos importantes de las herramientas de visualización como se muestra en la tabla 10, tomando en cuenta la situación del área de Talento Humano se ha elegido a Power BI Desktop como la herramienta a utilizar en el presente proyecto.

Una vez se hayan seleccionado las herramientas a utilizar en este proceso, se las procederá a instalar de manera correcta como se puede apreciar en el Anexo 1.

Capítulo IV: Metodología

El presente capítulo del trabajo de integración curricular, llevará a cabo un análisis detallado de la metodología a emplear. Se basará en los antecedentes recopilados en el primer capítulo, centrándose específicamente en las metodologías Hefesto y Kimball. El objetivo principal de este análisis es determinar cuál de estas metodologías será seleccionada y utilizada en el desarrollo del presente trabajo.

Metodología de Hefesto

Metodología desarrollada y proyectada por Ricardo Bernabeu a la cual denomino “Metodología de Hefesto”, la misma que puede ser utilizada en la implementación de Data Warehouse o Data Mart, cumpliendo cada una de las fases para su correcta implementación. (Bernabeu, 2010)

Esta metodología propone fases caracterizadas por su fácil comprensión, es decir propone estructuras de fácil adaptación las cuales están basadas en los requerimientos que tiene el usuario. Cuyo objetivo es satisfacer las necesidades de los requerimientos establecidos, mediante fases que no sean extensas y no generen tiempos excesivos. La metodología es independiente de las herramientas que se utilicen dentro del proceso, cabe indicar que el final de cada etapa es el comienzo de la siguiente. (Silva et al., 2021)

La metodología se compone de: análisis de requerimientos, análisis de los OLTP, modelo lógico y la integración de los datos con sus respectivas actividades como se puede apreciar en la figura 7 a continuación:

Figura 7

Fases de la metodología de Hefesto



Nota. La gráfica muestra las fases de la metodología de Hefesto. Tomado de Bernabéu, 2010.

Fases de la metodología de Hefesto

Análisis de requerimientos.

En la implementación de la mencionada metodología aplicada a los Data Warehouse y Data Mart se debe partir por la identificación y análisis de requisitos que hayan sido planteados inicialmente por el usuario o persona que le servirá dicha información, con la finalidad de cumplir los objetivos de la implementación. Para de esta manera identificar los indicadores que serán claves para la implantación de un repositorio de datos. Culminando esta fase en el desarrollo de un modelo conceptual.

Análisis de los OLTP.

Comprende la segunda fase de la metodología, donde se deberá mostrar los cálculos que se realizarán para obtener los indicadores definidos, mediante el análisis de las fuentes OLTP. Además, se podrá establecer los parámetros que se utilizarán para los distintos escenarios que se desarrollarán.

Modelo lógico del Data Mart.

En el desarrollo de esta fase se deberá apalancar del modelo conceptual para el posterior desarrollo del modelo lógico. A lo cual, se considerará seleccionar el tipo de modelo a desarrollar, afectando al desarrollo de las tablas de dimensiones y hechos.

Integración de datos.

Una vez terminada la tercera fase de la metodología en estudio se pondrá en marcha mediante la utilización de procesos ETL. Finalmente se estipulará las reglas y políticas de actualización que se llevarán a cabo, hasta pasar a la fase de visualización donde se podrá mostrar los datos trabajados en paneles de control (dashboard), los cuales serán la culminación para el apoyo en la toma de decisiones.

Características de la metodología de Hefesto

Bernabéu (2010), indica que la metodología que propuso (Hefesto) se caracteriza por:

- Los resultados que se obtiene de cada fase son comprensibles.

- Posee una estructura de fácil adaptación ante los constantes cambios.
- Se enfoca en los requerimientos del usuario.
- Puede aplicarse en Data Warehouse y Data Mart.
- Es independiente y no se condiciona de las herramientas utilizadas para la implementación.
- Tanto el modelo conceptual y lógico es entendible.

Metodología Ralph Kimball

Silva et al. (2019), afirma que la metodología “Ralph Kimball” (Modelo Dimensional) a la cual se la conoce de tal manera en honor a su creador, se fundamenta en el Ciclo de Vida Dimensional del Negocio. Por ende, se refiere a la combinación de tecnologías capaces de construir almacenes de datos. Su principal objetivo es lograr la presentación eficiente de la información mediante la utilización de técnicas para buscar la simplicidad y facilidad de comprensión de la información a ser tratada. El Modelo Dimensional, se establecen estructuras de tablas y relaciones con el fin de mejorar la eficiencia en la toma de decisiones. Esto se logra a través de consultas en una base de datos relacional que se vinculan con la información que es el resultado de los procesos de negocio. Esta metodología se basa en cuatro principios.

Principios de la metodología de Ralph Kimball

El proceso del ciclo de vida dimensional del negocio está constituido por 4 principios, que se los detalla a continuación:

Centrarse en el negocio. Consiste en priorizar la identificación de los requisitos del negocio para obtener un valor agregado en el análisis organizacional. Esta información se utiliza para establecer relaciones sólidas con el personal, mejorando el análisis y la competencia consultiva en la implementación de la metodología.

Construir una infraestructura de información adecuada. Consiste en el diseño de una base de datos, la misma que debe poseer la capacidad de satisfacer las necesidades identificadas de la organización.

Realizar entregas en incrementos significativos. Crear un almacén de datos el cual debe ser reforzado con entregables periódicos, tomando en cuenta plazos no mayores a doce meses, además de los requerimientos de la organización para priorizar los elementos identificados para la entrega.

Ofrecer una solución completa. Comprende en el desarrollo de un almacén de datos el cual tenga componentes sólidos, que sean accesibles para los usuarios de la organización, con la finalidad que sea base para la toma de decisiones.

Para obtener el entregable del desarrollo del almacén de datos bajo la metodología de Ralph Kimball se debe seguir las fases que se muestran en la figura 8:

Figura 8

Fases de la metodología de Ralph Kimball



Nota. En la gráfica se puede observar las fases de la metodología de Ralph Kimball. Tomado de Silva et al, 2019.

Fases de la metodología de Ralph Kimball

Planificación y Definición. Radica en la identificación de los objetivos comerciales de la organización. Haciendo referencia clave el identificar las necesidades del negocio. Para de esta manera establecer un plan de proyecto detallado el cual permita obtener los resultados finales al concluir el proyecto.

Diseño Dimensional. Radica en la realización de modelos dimensionales los cuales mediante los requisitos de la organización deben representar la información que ha sido obtenida de la organización. Mediante el diseño de esquemas de estrella o copo de nieve para facilitar el análisis, definiendo las jerarquías y relaciones dimensionales.

Implementación Física. Consiste en el uso de procesos ETL para la integración de datos tanto de tablas como de estructuras. Lo cual permita el aseguramiento de un rendimiento óptimo del almacén de datos.

Despliegue y Mantenimiento. Se basa en la puesta en marcha satisfactoriamente del almacén de datos implementado, siendo de gran utilidad de los usuarios de la organización, para que sea de apoyo en la toma de decisiones. Llegando a establecer continuos procesos de monitorización y a su vez de mantenimiento. Sin dejar de lado la retroalimentación y ajuste continuo según las necesidades cambiantes de la organización.

Comparación entre las metodologías de Hefesto y Ralph Kimball

Para lograr la eficiencia en el desarrollo de repositorios de datos es importante seguir las fases de ciertas metodologías como lo son las propuestas por Hefesto e Ralph Kimball. Estas metodologías ofrecen una guía estructurada que facilita la consecución de resultados precisos y acordes a las expectativas. Al seguir estas rutas metodológicas, se establece un marco que contribuye a la creación efectiva del almacén de datos y garantiza la obtención de los resultados deseados. Para lo cual nos centraremos en realizar una comparativa de ambas metodologías, como se muestra en la tabla 11:

Tabla 11

Metodología Hefesto vs Metodología Ralph Kimball

Criterio	Metodología Hefesto	Metodología Ralph Kimball
Autor	Ricardo Bernabéu	Ralph Kimball
Modelo	Dimensional	Dimensional
Arquitectura	Híbrida	Top-Down

Criterio	Metodología Hefesto	Metodología Ralph Kimball
Objetivo	Satisfacer las necesidades de los requerimientos establecidos, mediante fases que no sean extensas y no generen tiempos excesivos.	Proporcionar un marco estructurado y práctico para el diseño e implementación de almacenes de datos y Data Mart
Orientación	Creación de Data Warehouse y Data Mart	Creación de Data Warehouse y Data Mart
Tiempos	El desarrollo puede llevar menos de 4 meses	Cada incremento al desarrollo tiene plazos de 6 a 12 meses
Fases	Comprende 4 fases, no extensas	Comprende 4 fases, extensas
Usuario	Final	Final

Nota. La tabla se muestra una comparativa de aspectos importantes de las metodologías de “Hefesto” y “Ralph Kimball”.

Una vez realizado el cuadro comparativo de las metodologías en la tabla 8, servirá para justificar la elección de la metodología propicia a elegir en la siguiente sección.

Justificación de la metodología a seguir

Mediante el análisis realizado de las diferentes metodologías y en la comparativa realizada de las metodologías de Hefesto y Ralph Kimball en la tabla 8 se ha optado por elegir para este trabajo de integración curricular el uso de la metodología de Hefesto, por su fácil implementación, el tiempo de duración y sus características favorables para el desarrollo.

Para respaldar la elección de la metodología se ha tomado como referencia el trabajo realizado por Silva Et al (2019), donde muestra una comparativa de diversos indicadores, como adaptabilidad, tamaño del proyecto, tiempo de implementación, facilidad, documentación, entre otras, dicha comparativa fue realizada a las metodologías de Ralph Kimball, Hefesto y SAS, como se puede observar en la adaptación de la tabla 12.

Tabla 12*Indicadores características metodologías*

Indicadores	Metodologías		
	Ralph Kimball	Hefesto	SAS
Adaptabilidad	3	3	3
Afinidad	3	3	3
Intuitivo	3	3	2
Alcance	2	3	2
Tiempos análisis/diseño	1	3	1
Tiempo implementación	1	3	2
Curva aprendizaje	1	3	1
Documentación	3	3	2
Mas cotidiana	1	3	1
Total	18	27	17

Nota. En esta tabla se muestra un resumen de los indicadores comparados en el trabajo “Análisis de metodologías para desarrollar Data Warehouse aplicado a la toma de decisiones”. Adaptado de Silva et al, 2019.

Focalizados en la necesidad del departamento de Talento Humano del Gobierno Autónomo Descentralizado del Canto Guaranda, se plantea la creación de un Data Mart mediante el uso de la metodología Hefesto. Esta metodología se implementará para la creación del almacén de datos del área el cual servirá para la toma de decisiones.

Entre los beneficios que brindara al área de talento humano la implementación de la herramienta para la toma de decisiones son:

- Permite tomar decisiones basados en información real, evitando de esta manera especulaciones en cuanto a la información que se maneja.
- Permite focalizar la información del área para que de esta manera la información sea relacionada exclusivamente al área que lo necesite.
- Permite tener información procesada, evitando de esta manera información no relevante que puede ser perjudicial en la toma de decisiones.

Capítulo V: Desarrollo

El presente capítulo del trabajo de integración curricular se compone de los siguientes elementos: descripción de la metodología a seguir, caso de estudio, cubos de información, resultados obtenidos, cumplimiento de objetivos y validación de hipótesis.

Descripción de la metodología a seguir

Bernabéu (2010), describe su metodología en los siguientes pasos y actividades:

Fase 1: Análisis de requerimientos

En la primera etapa de la metodología se identifican los requisitos propuestos por el usuario, buscando de esta manera satisfacer los objetivos planteados. Para el desarrollo del Data Mart es indispensable el análisis de las preguntas, de manera que permita la identificación de los indicadores y perspectivas relevantes. Las mismas que serán útiles en el desarrollo de un modelo conceptual el cual contendrá los resultados obtenidos inicialmente.

Identificar preguntas. El proceso comienza con la recopilación de las necesidades de información mediante técnicas como entrevistas y cuestionarios. Buscando como resultado la obtención de los requisitos óptimos, los cuales deben respaldar las metas de la organización. Se hace necesario enfocarse en procesos fundamentales, respaldando la información en sistemas OLTP, los cuales influirán en el desarrollo del Data Mart.

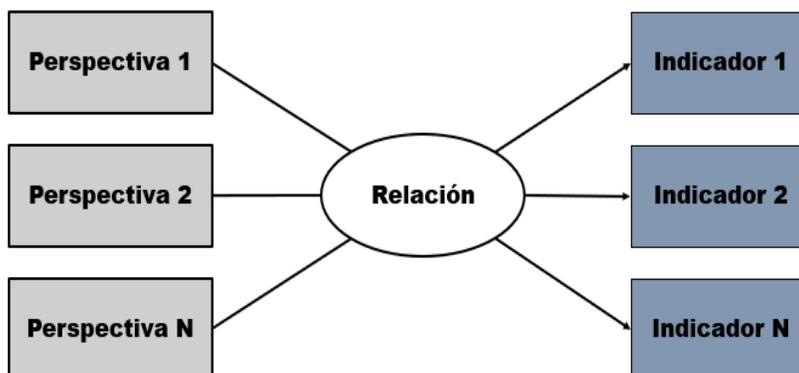
Identificar identificadores y perspectivas. Una vez hayan sido identificadas las preguntas, es importante diferenciar los indicadores y las perspectivas. Cabe indicar que los indicadores corresponden a valores numéricos, por otro lado, las perspectivas corresponden a elementos mediante los cuales se evalúan los indicadores.

Modelo conceptual. Esta etapa es la propicia para el desarrollo del modelo conceptual, en la cual se debe tomar en cuenta los indicadores y perspectivas identificados anteriormente, como se aprecia en la figura 9. Este modelo ofrece una visión clara de los objetivos del proyecto, facilitando su presentación y explicación a los usuarios. En el desarrollo del modelo conceptual se debe tomar en cuenta que las perspectivas se encuentran en la parte izquierda y

los indicadores se encuentran en la parte derecha. Este esquema facilita la comprensión de resultados, variables de análisis e interrelaciones de manera clara y accesible, prescindiendo de conocimientos especializados previos.

Figura 9

Modelo Conceptual



Nota. En la gráfica se representa un modelo conceptual. Adaptado de "HEFESTO: Metodología para la Construcción de un Data Warehouse", (p. 92), por R. D. Bernabéu, 2010, recuperado de <https://www.businessintelligence.info/resources/assets/hefesto-v2.pdf>.

Fase 2: Análisis de los OLTP

Esta fase se centra en el análisis de las fuentes OLTP para relacionar el modelo conceptual preexistente con los resultados obtenidos del cálculo de indicadores. De esta manera se procederá a ampliar el modelo con los campos a agregar en las distintas perspectivas.

Conformar indicadores. Durante este paso, se detalla cómo se calcularán los indicadores, conjuntamente con las fórmulas de cálculo, así como también las funciones de agregación a emplear como SUM, AVG, COUNT, entre otras.

Establecer correspondencias. Se procede a relacionar el modelo conceptual y las fuentes de datos. Garantizando de esta manera la relación entre ambos.

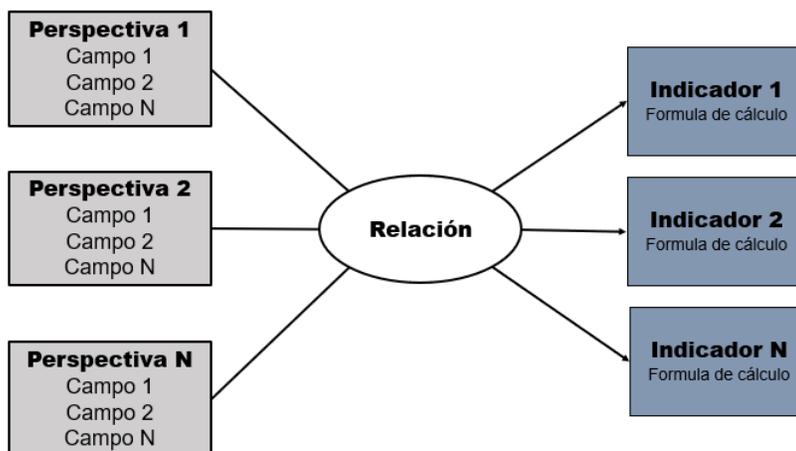
Nivel de granularidad. Se filtrarán los indicadores a utilizar mediante la selección de los campos idóneos de cada perspectiva. Para la realización de este proceso, es indispensable el tener conocimiento de cada campo, para lo cual se desarrollan diccionarios de datos. Una vez

comprendida el significado de cada campo mediante el diccionario de datos, se procederá a seleccionar los campos a utilizar en el análisis. Este proceso de selección de campos a utilizar se denomina granularidad de la información.

Modelo conceptual ampliado. Esta etapa simboliza el proceso de ampliación del modelo, mediante la identificación previa de los factores. Se incorporan los campos seleccionados debajo de cada perspectiva y se añaden las fórmulas de cálculo bajo cada indicador como se puede mostrar en la figura 10. El propósito es generar una representación gráfica de la estructura resultante.

Figura 10

Modelo conceptual Ampliado



En la gráfica se visualiza el esquema de un modelo conceptual ampliado. Adaptado de “HEFESTO: Metodología para la Construcción de un Data Warehouse”, (p. 98), por R. D. Bernabéu, 2010, recuperado de <https://www.businessintelligence.info/resources/assets/hefesto-v2.pdf>.

Fase 3: Modelo lógico del Data Mart o Data Warehouse

Corresponde al desarrollo del modelo lógico tomando como base el modelo conceptual anteriormente diseñado. Debido a que se le puede tomar como punto de partida para el desarrollo de tablas de hechos y de dimensiones. La fase culmina con la realización de las uniones necesarias entre estas tablas para establecer la estructura definitiva del Data Mart.

Tipo de modelo lógico del Data Mart o Data Warehouse. En la selección del esquema para la estructura del Data Mart, ya sea en estrella, constelación o copo de nieve, es crucial considerar los requisitos y necesidades de los usuarios. Esta decisión tiene un impacto considerable en la construcción del modelo lógico del Data Mart.

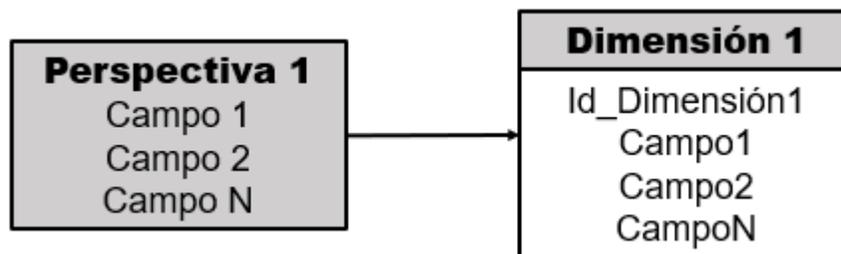
Tablas de dimensiones. Esta fase es propicia para el desarrollo de las tablas de dimensiones que serán parte del Data Mart, donde cada perspectiva será una tabla de dimensiones. El proceso implica:

- Nombrar de manera representativa la tabla de dimensiones.
- Seleccionar un campo principal o clave primaria.
- Ajustar los nombres de cada campo.

Se puede observar gráficamente en la figura 11:

Figura 11

Esquema de una tabla de dimensiones



Nota. En la gráfica se visualiza el esquema de una tabla de dimensiones. Adaptado de “HEFESTO: Metodología para la Construcción de un Data Warehouse”, (p. 99), por R. D. Bernabéu, 2010, recuperado de <https://www.businessintelligence.info/resources/assets/hefesto-v2.pdf>.

Tablas de hechos. Son las cuales tienen la competencia de contener la información de los indicadores de estudio. Lo cual debe seguir los siguientes procedimientos:

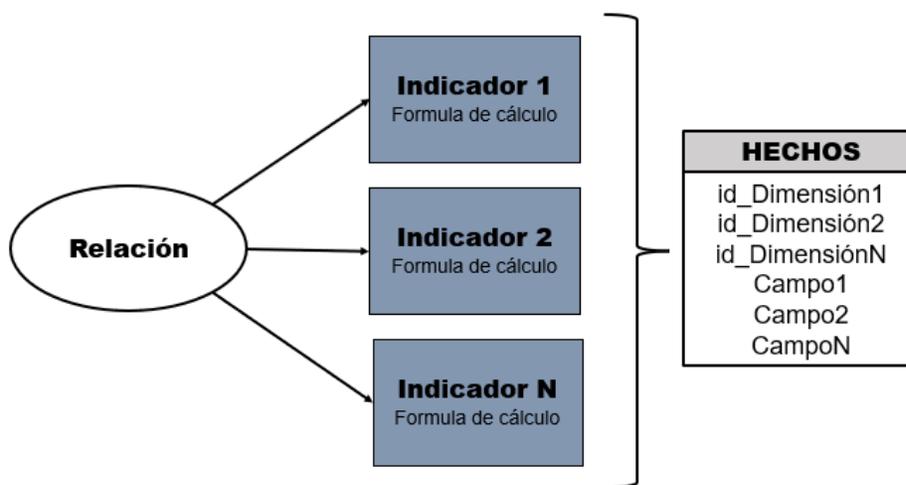
- Nombrar a la tabla de acuerdo a la información a analizar.
- Asignar una clave primaria representativa de las claves primarias de las tablas de dimensiones relacionadas.

- Establecer campos de hechos en las tablas (tablas de hechos) para cada uno de los indicadores presentes en el modelo conceptual.

Se puede observar gráficamente en la figura 12:

Figura 12

Esquema de una tabla de hechos



Nota. En la figura se muestra un esquema de una “tabla de hechos”. Adaptado de “HEFESTO: Metodología para la Construcción de un Data Warehouse” , (p. 102), por R. D. Bernabéu, 2010, recuperado de <https://www.businessintelligence.info/resources/assets/hefesto-v2.pdf>.

Uniones. Para cualquiera de los tres tipos de esquemas sean estrella, copo de nieve o bóveda es indispensable llevar a cabo conexiones apropiadas entre las distintas tablas de dimensiones y hechos.

Fase 4: Integración de datos

Una vez se encuentre desarrollado completamente el modelo lógico, los datos serán cargados con la ayuda de distintas técnicas como la limpieza y mejora de la calidad de datos o en general procesos ETL. Posteriormente, se debe establecer políticas de actualización con la finalidad de que el almacén de datos tenga la información relevante del área a la que se aplica.

Carga inicial. Esta etapa es en la cual se realiza la carga inicial de información al Data Mart, siendo el complemento del modelo de datos que se venía trabajando. Este proceso agrupa

tareas como la limpieza de la data, implementación de procesos ETL, buscando de esta manera obtener datos de mejor calidad.

Este proceso inicia con la carga de las tablas de dimensiones, luego las de hechos garantizando de esta manera la correcta correspondencia entre elementos.

Actualización. Después de crear el modelo lógico, se procederá a llenarlo con datos mediante técnicas que incluyen limpieza y mejora de la calidad de datos, así como procesos ETL. Posteriormente, se establecerán normas y políticas para la actualización del modelo, describiendo los procedimientos responsables de llevar a cabo dichas actualizaciones.

Caso práctico

Una vez analizada las distintas fases y componentes de los mismos se podrá en marcha la implementación del caso práctico.

Fase 1: Análisis de requerimientos

Identificar preguntas. – En este paso se identifican los requisitos, que tiene el usuario entorno a la información proporcionada. Las mismas que se pueden apreciar en el Anexo 2. A continuación un extracto de las mismas:

- Se desea conocer la cantidad de empleados beneficiado por los uniformes de acuerdo al régimen en el que se encuentra. Lo cual consiste en: cantidad de personal de determinado régimen laboral y rol que cumple que se vio beneficiado con la entrega de prendas de vestir (uniformes).
- Se desea conocer la cantidad de prendas entregadas de acuerdo al régimen en el que se encuentra y a su según el rol que cumple. Lo cual consiste en: cantidad de prendas que han sido entregadas al personal de determinado régimen laboral y rol que cumple dentro de la municipalidad.
- Se desea conocer el total del gasto de prendas entregadas a los empleados de acuerdo al régimen en el que se encuentra y al rol que cumple. Lo cual consiste en:

gasto total de prendas entregadas de acuerdo al régimen laboral y rol que cumple dentro de la municipalidad.

Identificar identificadores y perspectivas. – Para tal motivo en base al extracto tomado del Anexo 2, se procederá a la identificación de indicadores y perspectivas, como se muestra a continuación:

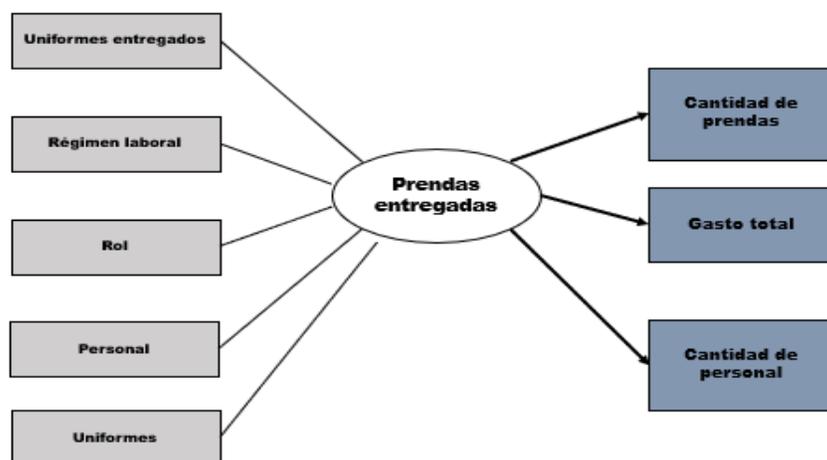
- Cantidad de empleados (personal) de determinado régimen laboral y rol que cumple que se vio beneficiado con la entrega de prendas de vestir (uniformes entregados).
 - **Indicador:** Cantidad de personal
 - **Perspectivas:** personal, régimen laboral, uniformes y rol
- Cantidad de prendas (uniformes entregados) que han sido entregadas al personal de determinado régimen laboral y rol que cumple dentro de la municipalidad.
 - **Indicador:** Cantidad prendas
 - **Perspectivas:** régimen laboral, uniformes entregados y rol
- Gasto total de prendas entregadas (uniformes entregados) de acuerdo al régimen laboral y rol que cumple dentro de la municipalidad.
 - **Indicador:** Gasto total de prendas
 - **Perspectivas:** régimen laboral, uniformes entregados, uniformes y rol

En resumen, se puede manifestar que los indicadores son: cantidad de personal, cantidad de prendas y gasto total de prendas. Por otro lado, las perspectivas son: personal, régimen laboral, rol, uniformes y uniformes entregados.

Modelo conceptual. – Mediante los datos recolectados anteriormente se elabora el siguiente modelo, el cual es la unión de perspectivas e indicadores, como se puede observar en la figura 13:

Figura 13

Modelo Conceptual del caso de estudio



Nota. En la gráfica se muestra el modelo conceptual ampliado, tomando como base el extracto del modelo de base de datos del área de Talento Humano.

Fase 2: Análisis de los OLTP

Conformar indicadores. – Este paso corresponde al modo de cálculo de los indicadores.

Los cuales se calcularán de acuerdo a lo que se indica seguidamente:

- Cantidad de personal:
 - Hechos: Cantidad personal
 - Función: Sumarización

Aclaración: el indicador “Cantidad personal” representa la suma de todo el personal que ha recibido prendas de vestir.

- Cantidad de prendas
 - Hechos: Cantidad de prendas
 - Función: Sumarización

Aclaración: el indicador “Cantidad de prendas” representa la suma de todo todas las prendas entregadas al personal de la municipalidad.

- Gasto total
 - Hechos: $(\text{Cantidad}) \times (\text{valor})$

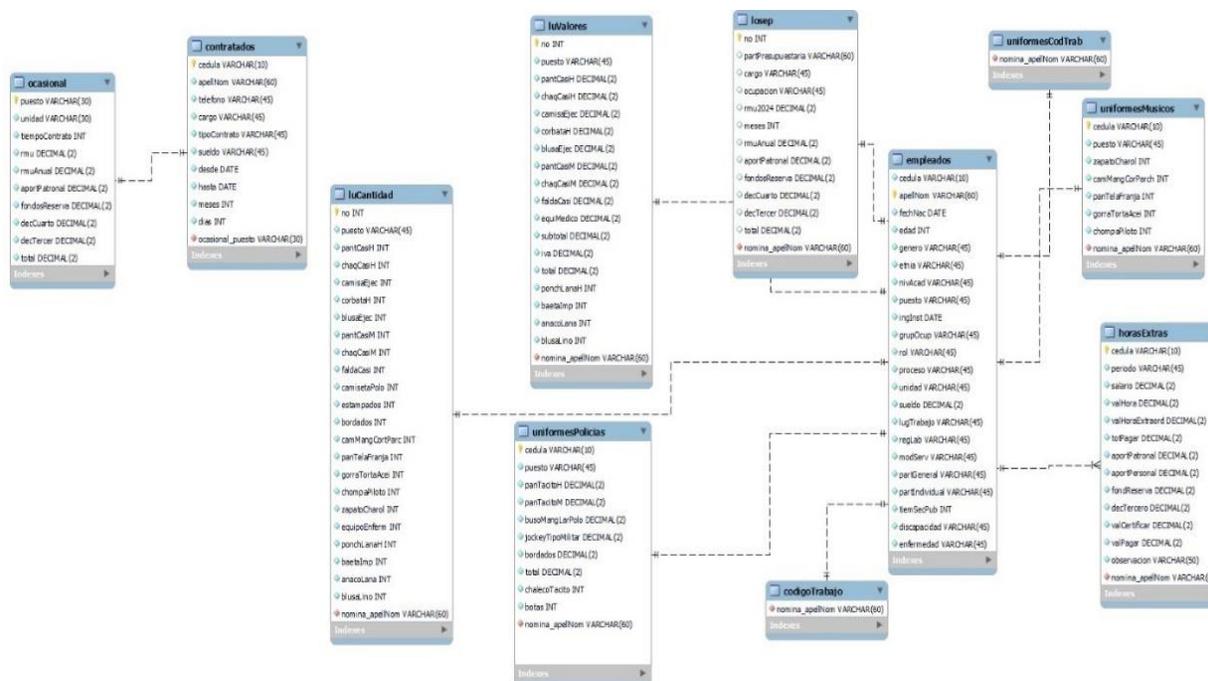
- Función: Sumarización

Aclaración: el indicador “Cantidad personal” representa la suma del producto entre la cantidad de prendas y su valor.

Establecer correspondencias. – Se analiza el diagrama entidad relación del área de talento humano, el cual se muestra a continuación:

Figura 14

Modelo Entidad Relación Encontrado

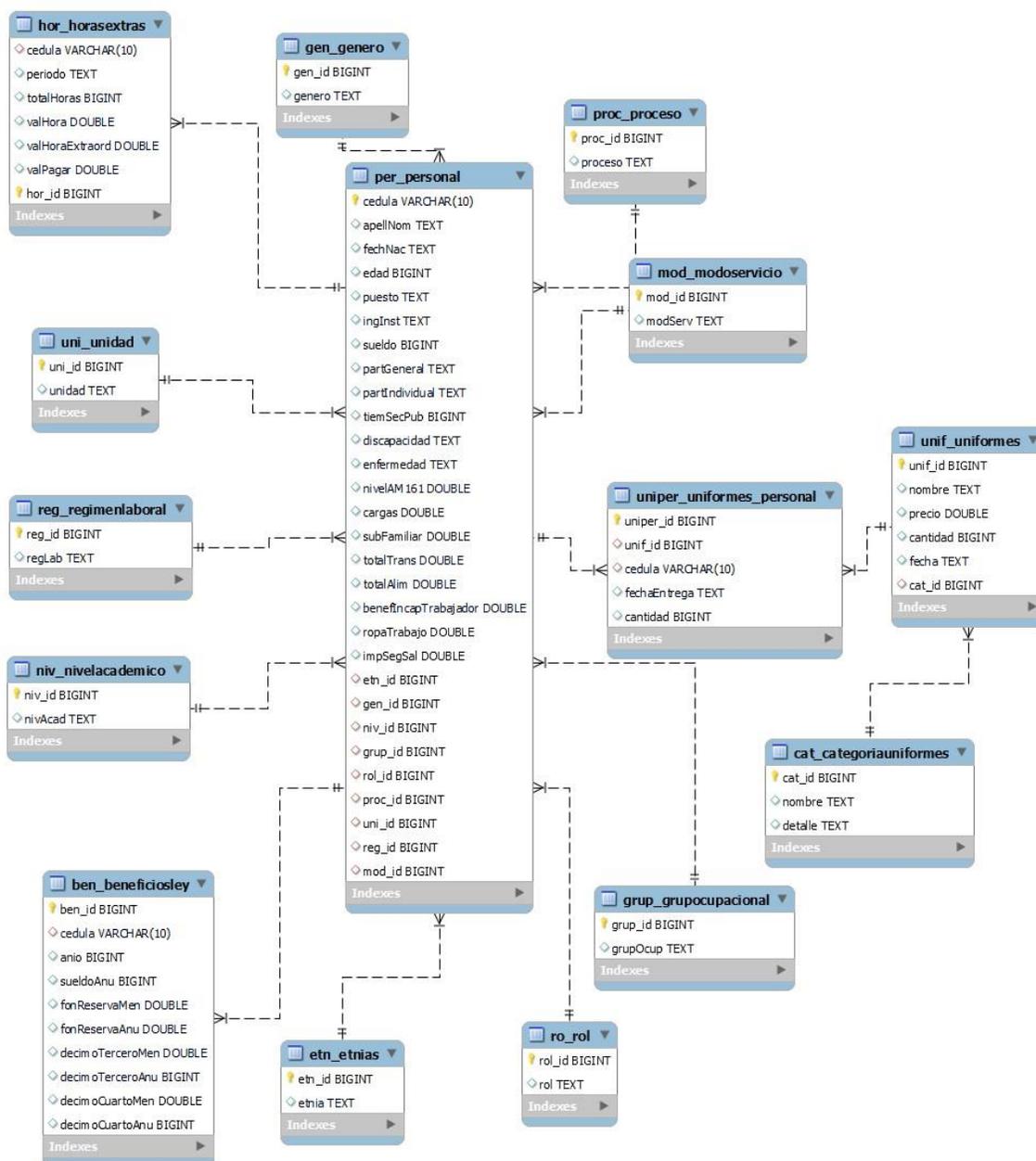


Nota. En la gráfica se muestra el modelo entidad relación inicial del área de Talento Humano.

El modelo presentado en la figura 14 representa el modelo inicial encontrado en base a la data que se tiene. Luego del análisis y tratamiento de la información se logró obtener un modelo más sólido, como se muestra en la figura 15 a continuación:

Figura 15

Modelo Entidad Relación Actualizado

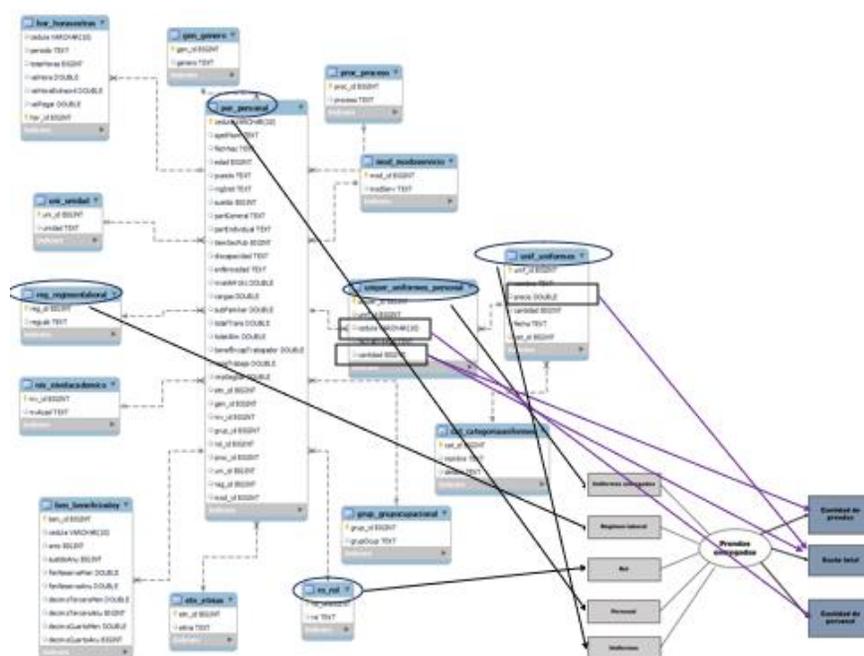


Nota. En la gráfica se muestra el modelo entidad relación actualizado del área de Talento Humano.

Una vez conocido el modelo a trabajar, se realiza la relación entre los dos modelos (conceptual y entidad relación), como se muestra en la figura 16 a continuación:

Figura 16

Relación entre el Modelo Conceptual y Modelo Entidad Relación



Nota. En la gráfica se muestra una relación entre el modelo conceptual y el modelo entidad relación actualizado del área de Talento Humano.

Nivel de granularidad. – Mediante el diccionario de datos que se encuentra en el Anexo 4, se puede conocer con mayor detalle el concepto de cada uno de los campos de las tablas del modelo entidad relación. A continuación, se presenta un extracto del documento en mención, mediante la tabla 13, tabla 14 y tabla 15.

Tabla 13

Diccionario de datos de la tabla uniformes

Tabla: unif_uniformes	
Tabla mediante la cual se almacena un catálogo de prendas de vestir.	
Campos	Descripción
unif_id	Identificador único de registro, de las distintas prendas de vestir.
nombre	Nombre de la prenda.
precio	Precio comprado de la prenda.
cantidad	Cantidad correspondiente al precio de compra.
fecha	Fecha de la compra de la prenda.
cat_id	Identificador de la categoría a la que pertenece la prenda de vestir.

Nota. En la tabla se muestra el diccionario de datos de una tabla en específica.

Tabla 14

Diccionario de datos de la tabla uniformes-personal

Tabla: unif_uniformes_personal	
Tabla mediante la cual se almacena las prendas entregadas al personal.	
Campos	Descripción
uniper_id	Identificador unico de registro, de las prendas entregadas al personal.
unif_id	identificador de la prenda.
cedula	Identificador unico del personal.
fechaEntrega	Fecha de entrega de la prenda de vestir al personal.
cantidad	Cantidad de prendas entregadas al personal.

Nota. En la tabla se muestra el diccionario de datos de una tabla en especifica.

Tabla 15

Diccionario de datos de la tabla rol

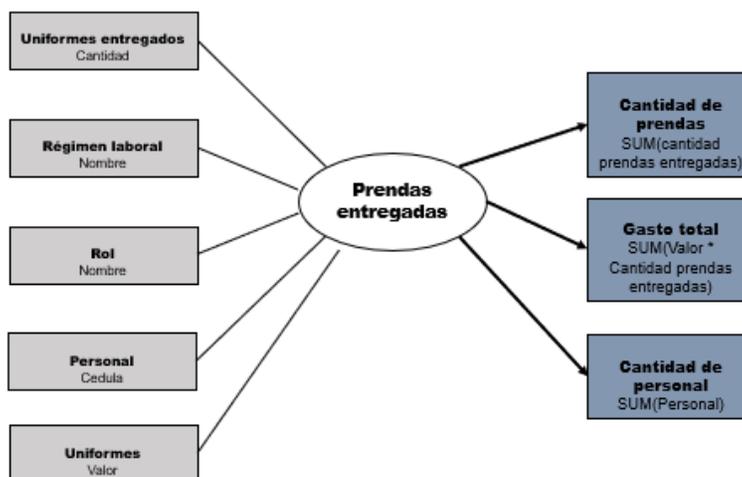
Tabla: ro_rol	
Tabla mediante la cual se almacena el tipo de etnias.	
Campos	Descripción
rol_id	Identificador unico de registro, del tipo de rol.
rol	Nombre del rol.

Nota. En la tabla se muestra el diccionario de datos de una tabla en especifica.

Modelo conceptual ampliado. – Radica en el diseño del modelo conceptual ampliado con los cálculos y campos a utilizar. Como se muestra a continuación en la figura 17:

Figura 17

Modelo Conceptual Ampliado del caso de estudio



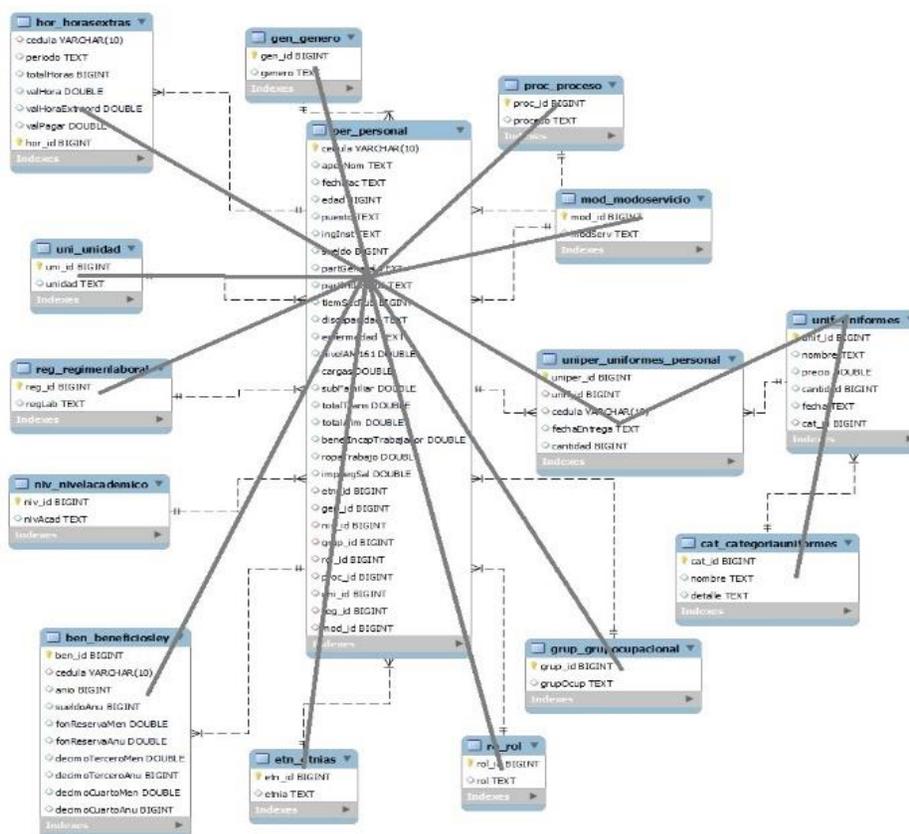
Nota. En la figura se observa el modelo conceptual ampliado del caso de estudio.

Fase 3: Modelo l3gico del Data Mart o Data Warehouse

Tipo de modelo l3gico del Data Mart o Data Warehouse. – El modelo l3gico que se utilizar3 el de tipo Copo de nieve, el cual se revis3 en la Tabla 5 Estructura de los Data Mart y como se puede observar de manera visual en la figura 18.

Figura 18

Modelo Copo Nieve

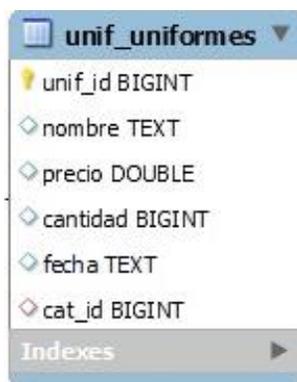


Nota. En la figura se observa el modelo copo de nieve adaptado al modelo entidad relaci3n del caso de estudio.

Tablas de dimensiones. – Es el dise1o y selecci3n de las tablas de dimensiones para el Data Mart. Cabe indicar que cada una de las perspectivas ser3 una tabla de dimensiones, las cuales estar3n compuestas de sus campos relacionados como se puede observar en la figura 19, figura 20 e figura 21, las cuales son un extracto de las tablas de dimensiones, tomadas del modelo entidad – relaci3n.

Figura 19

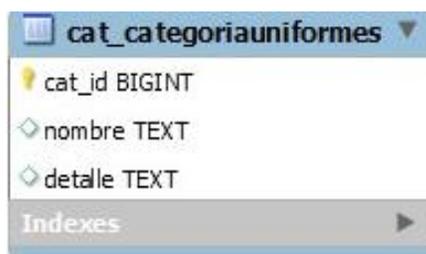
Tabla de dimensiones uniformes



Nota. En la figura se observa la tabla de dimensiones de uniformes.

Figura 20

Tabla de dimensiones Categoría Uniformes



Nota. En la figura se observa la tabla de dimensiones de categoría de uniformes.

Figura 21

Tabla de dimensiones Rol



Nota. En la figura se observa la tabla de dimensiones de rol.

Tablas de hechos. – Este paso es el propicio para la definición de las tablas de hechos, las cuales estarán compuestas por los distintos indicadores, que facilitarán el estudio. Como se puede observar en la figura 22.

Figura 22

Tabla de hechos Uniformes-Personal

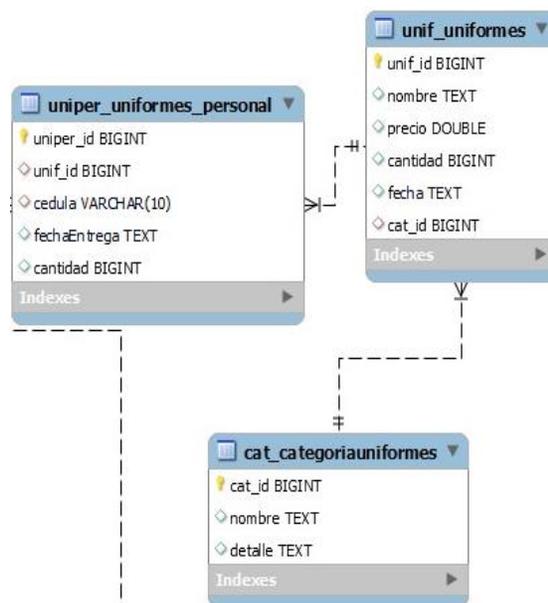
Column Name	Data Type
uniper_id	BIGINT
unif_id	BIGINT
cedula	VARCHAR(10)
fechaEntrega	TEXT
cantidad	BIGINT

Nota. En la figura se observa la tabla de hechos de uniformes entregados al personal.

Uniones. – Se realizan las relaciones (uniones) que existen entre las tablas (hechos, dimensiones). Estas uniones en su mayoría son de uno a muchos y de muchos a uno. Como se puede mostrar en la figura 23 que muestra el extracto del modelo completo.

Figura 23

Uniones realizadas en el modelo



Nota. En la figura se observa las uniones realizadas entre las tablas de hechos y tablas de dimensiones.

Fase 4: Integración de datos

Carga inicial. – Para la carga inicial de la data, la cual se realizó después de realizar algunos procesos, a continuación, se detallan algunos de estos:

- Extracción

- El proceso de extracción de la data se lo realizo de distintos archivos Excel que han sido proporcionado por parte del área de Talento Humano. Algunos de los cuales son horas extras, distributivo, planificación, uniformes, entre otros.
- Transformación
 - Completar datos característicos que se encuentren vacíos.
 - Corrección de formatos de fecha, cedula, teléfono.
 - Estandarización de datos.
 - Eliminación de datos incoherentes.
 - Eliminación de valores nulos.
- Carga
 - Carga de la data al Data Mart en MySQL, identificando las relaciones mediante la asignación de claves primarias y claves foráneas.

Como proceso de carga inicial se puede evidenciar la estructura de tablas cargadas dentro del Data Mart almacenado en MySQL, se muestra la figura 24:

Figura 24

Esquema de tablas en MySQL

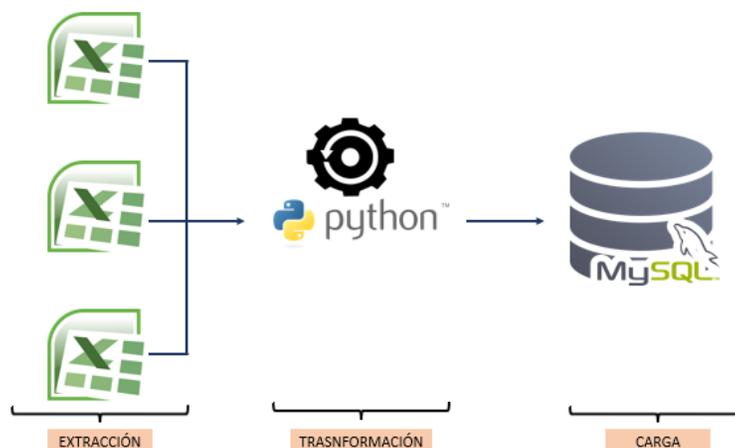


Nota. En la Figura se observa el esquema de tablas cargadas en el Data Mart.

La realización de los procesos antes mencionados se los realizo bajo la arquitectura mostrada en la figura 25:

Figura 25

Arquitectura ETL



Nota. En la figura se observa la arquitectura ETL que se utiliza en este trabajo.

Actualización. – La actualización de la data para el funcionamiento del mecanismo se la realiza en periodos mensuales en los documentos plantillas, los cuales servirán para la carga de la información en el Data Mart del área de Talento Humano. La selección del tiempo de actualización se da debido a que los datos relevantes como horas extras e ingreso del personal se da a inicios de mes.

Cubos de información

En el desarrollo del proyecto, se optó por el uso de Funciones DAX (Data Analysis Expressions) en la herramienta Power BI que previamente en el Capítulo III, fue seleccionada como la herramienta para la visualización de datos, como una alternativa al desarrollo de cubos de información tradicionales, los cuales necesitaban de herramientas dedicadas especialmente al desarrollo de cubos de información.

Las funciones DAX son utilizadas para el cálculo de agregaciones, realizar cálculos avanzados y crear medidas que proporcionen análisis multidimensionales. La decisión de

utilizar funciones DAX para el cumplimiento de los requerimientos que se encuentran en el Anexo 2, parte de las características que ofrecen, que son las siguientes:

- La utilización de funciones DAX nos permite ahorrar tiempo en el proceso de implementación y alcanzar resultados más rápidos en relación con el desarrollo de cubos de información tradicionales.
- Al emplear funciones denominadas DAX, se puede realizar cálculos y manipulaciones de datos complejos de manera sencilla, sin necesidad de habilidades técnicas especializadas.
- Capacidad de realizar cambios en tiempo real, lo cual permite ver los cambios inmediatamente.
- Las funciones DAX pueden ser empleadas para realizar análisis multidimensionales dentro del contexto de un modelo de datos.

Con la finalidad de satisfacer los requerimientos detallados en el Anexo 2, se desarrollaron las siguientes funciones DAX:

Función 1

Empleados por categoría

```
CantidadPersonasPorCategoria =  
CALCULATE(  
    COUNTROWS(  
        SUMMARIZE(  
            uniper_uniformes_personal,  
            uniper_uniformes_personal[cedula],  
            'unif_uniformes'[cat_id],  
            'cat_categoriauniformes'[nombre]  
        )  
    ),  
    FILTER(  
        unif_uniformes,  
        RELATED(cat_categoriauniformes[cat_id]) = 'unif_uniformes'[cat_id]  
    )  
)
```

Nota. En el cuadro se muestra la función DAX realizado en la herramienta Power BI.

Función 2

Prendas entregadas por rol

```
CantidadPrendasPorUnidad =
CALCULATE(
    SUM(uniper_uniformes_personal[cantidad]),
    ALLEXCEPT(ro_rol, ro_rol[rol])
)
```

Nota. En el cuadro se muestra la función DAX realizado en la herramienta Power BI.

Función 3

Gasto empleado en vestimenta

```
GastoTotal =
SUMX(
    uniper_uniformes_personal,
    RELATED(unif_uniformes[precio]) * uniper_uniformes_personal[cantidad]
)
```

Nota. En el cuadro se muestra la función DAX realizado en la herramienta Power BI.

Función 4

Gasto en salarios por rol

```
GastoTotalPorRol =
SUMX(
    SUMMARIZE(
        per_personal,
        ro_rol[rol],
        "GastoTotal",
        SUMX(
            RELATEDTABLE(uniper_uniformes_personal),
            RELATED(unif_uniformes[precio]) *
            uniper_uniformes_personal[cantidad]
        )
    ),
    [GastoTotal]
)
```

Nota. En el cuadro se muestra la función DAX realizado en la herramienta Power BI.

Función 5

Promedio de años de permanencia en la institución de los empleados

```
MedidaAñosEnEmpresa =
    AVERAGEX(
        ALL('per_personal'),
        DATEDIFF('per_personal'[ingInst], TODAY(), YEAR)
    )
```

Nota. En el cuadro se muestra la función DAX realizado en la herramienta Power BI.

Función 6

Contabilización del personal según su nivel académico

```
PersonasPorNivelAcademico =
    COUNTROWS(
        SUMMARIZE(
            per_personal,
            'niv_nivelacademico'[nivAcad],
            "Numero de Personas", DISTINCTCOUNT(per_personal[cedula])
        )
    )
```

Nota. En el cuadro se muestra la función DAX realizado en la herramienta Power BI.

Función 7

Promedio de antigüedad del personal por el rol que ocupa

```
PromedioAntigüedadPorRol =
    AVERAGEX(
        GROUPBY(
            per_personal,
            per_personal[rol_id],
            "AntigüedadPromedio",
            AVERAGEX(CURRENTGROUP(), DATEDIFF([ingInst], TODAY(), YEAR))
        ),
        [AntigüedadPromedio]
    )
```

Nota. En el cuadro se muestra la función DAX realizado en la herramienta Power BI.

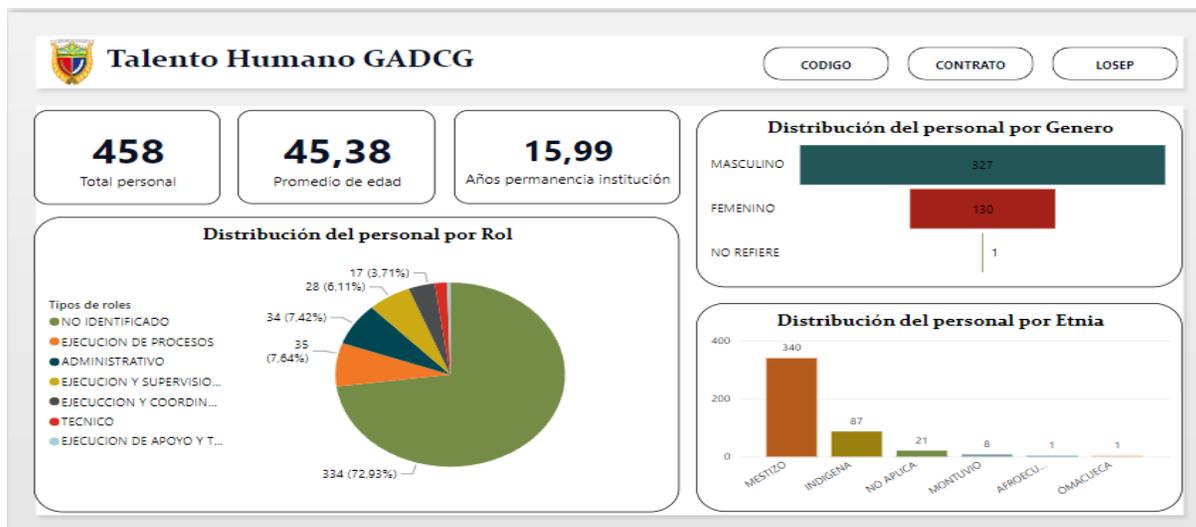
Resultados obtenidos

Una vez realizado el desarrollo del presente trabajo de integración curricular, mediante el tratamiento de las diversas bases de datos y el desarrollo de las funciones DAX presentados en

la sección anterior se puede llegar a mostrar la información tratada mediante paneles de control, como se puede observar en las figuras a continuación, donde se observa el cumplimiento de los requisitos establecidos en el Anexo 2.

Figura 26

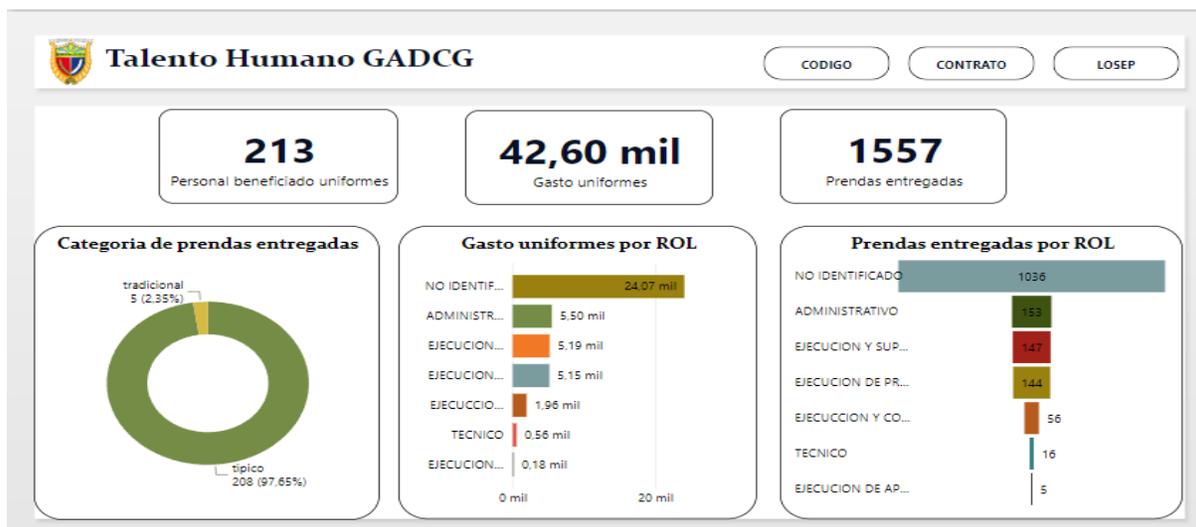
Panel de control del Personal



Nota. En la figura se observa un panel de control desarrollado en Power BI.

Figura 27

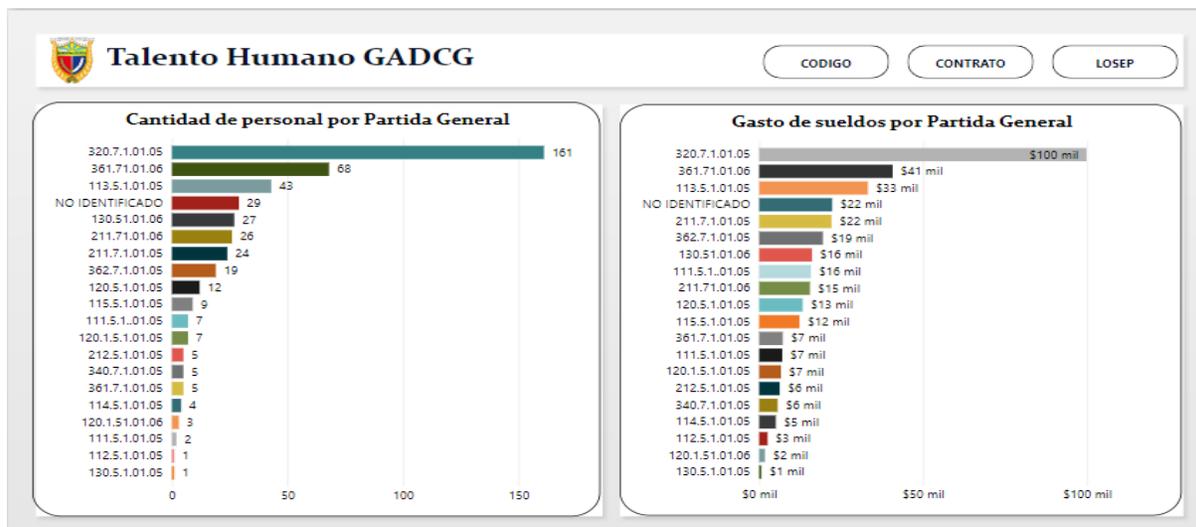
Panel de control de Uniformes



Nota. En la figura se observa un panel de control desarrollado en Power BI.

Figura 28

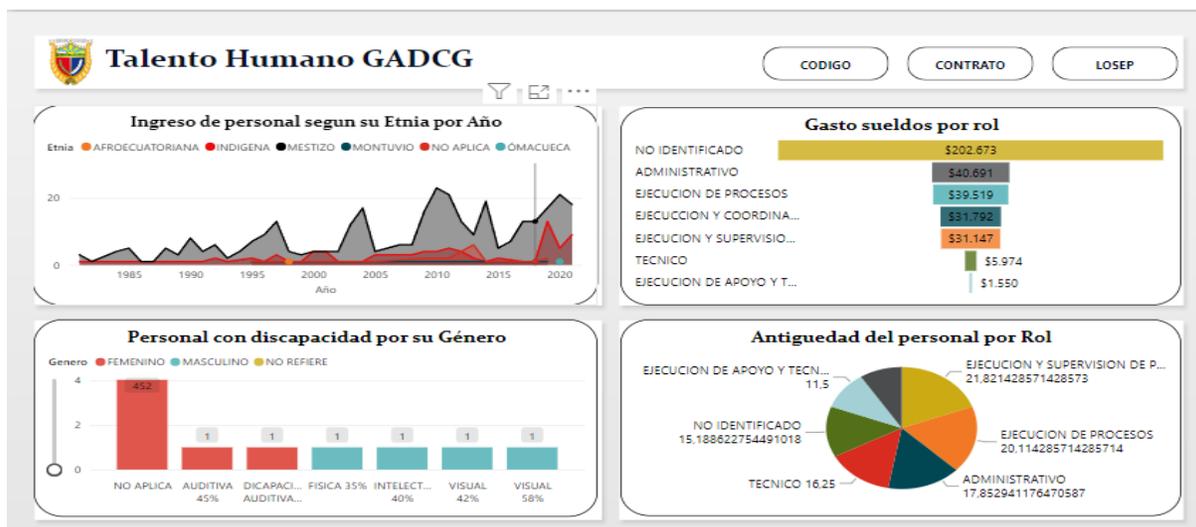
Panel de control de Partidas Presupuestarias



Nota. En la figura se observa un panel de control desarrollado en Power BI.

Figura 29

Panel de control de Roles de los empleados



Nota. En la figura se observa un panel de control desarrollado en Power BI.

Figura 30

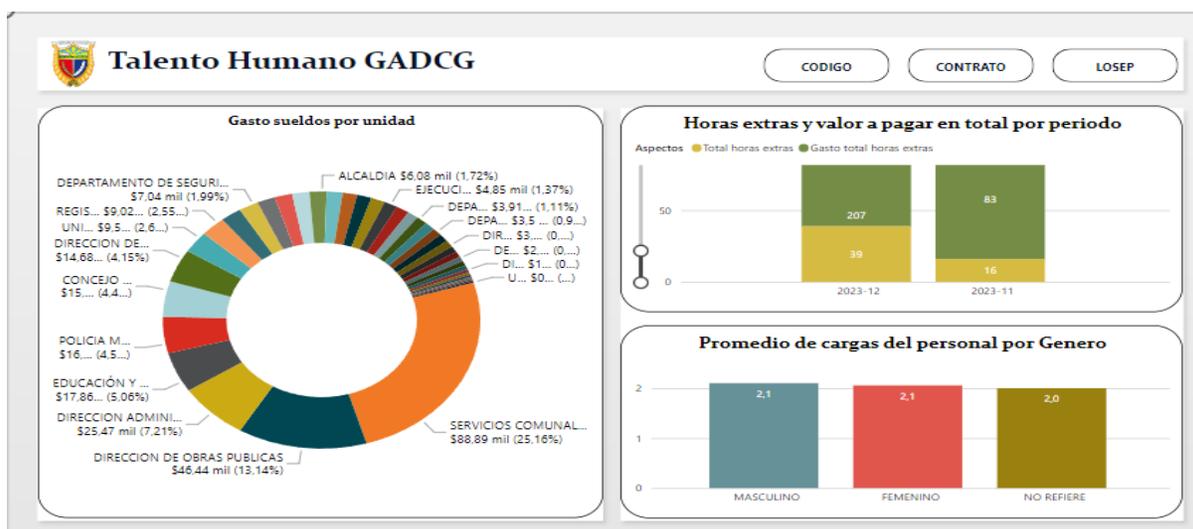
Panel de control de Beneficios de Ley



Nota. En la figura se observa un panel de control desarrollado en Power BI.

Figura 31

Panel de control Horas Extras



Nota. En la figura se observa un panel de control desarrollado en Power BI.

Cumplimiento de objetivos

Revisión de Objetivos:

- Consolidar la información de la Dirección de Talento Humano.
- Desarrollar cubos de información requeridos.

- Generar información para la dirección

Evaluación del Cumplimiento de Objetivos:

Objetivo 1: Consolidar la información de la Dirección de Talento Humano:

Cumplimiento: Sí.

Evidencia: Se llevo a cabo un mecanismo para recopiló, transformación y almacenamiento de manera efectiva la información relevante del Departamento de Talento Humano del Gobierno Autónomo Descentralizado del Cantón Guaranda. Dicha información sometida al instrumento abarca información como: datos del personal, horas extras, entrega de uniformes y beneficios de ley (décimo tercero, décimo cuarto, fondos reserva). Para tal efecto, se alimentó de la información que ha sido procesada al Data Mart que previamente se obtuvo de diversas fuentes. A continuación, se muestra un resumen del número de datos que han sido cargados en el Data Mart del área producto del proceso de integración de las bases de datos iniciales desarrollado en la Fase 4: Integración de datos.

Tabla 16

Resumen de tablas desarrolladas

Tabla	Registros
ben_beneficiosley	458
cat_categoriauniformes	2
etn_etnias	6
gen_genero	3
grup_grupocupacional	23
hor_horasextras	3
mod_modoservicio	2
niv_nivelacademico	5
per_personal	458

Tabla	Registros
proc_proceso	7
reg_regimenlaboral	3
ro_rol	7
uni_unidad	39
unif_uniformes	47
uniper_uniformes_personal	1433

Nota. En la tabla se muestra un resumen de las tablas que se han cargado en el Data Mart.

Por otro lado, se establecieron relaciones, mediante la asignación de claves primarias y claves foráneas. Las cuales se pueden observar a continuación:

Tabla 17

Resumen de relaciones generadas

Tabla	Clave Primaria	Clave Foránea
ben_beneficiosley	ben_id	cedula
cat_categoriauniformes	cat_id	
etn_etnias	etn_id	
gen_genero	gen_id	
grup_grupocupacional	grup_id	
hor_horasextras	hor_id	cedula
mod_modoservicio	mod_id	
niv_nivelacademico	niv_id	
uni_unidad	uni_id	
proc_proceso	proc_id	
reg_regimenlaboral	reg_id	
ro_rol	rol_id	

Tabla	Clave Primaria	Clave Foránea
unif_uniformes	unif_id	cat_id
		etn_id
		gen_id
		niv_id
		grup_id
per_personal	cedula	rol_id
		proc_id
		uni_id
		reg_id
		mod_id
uniper_uniformes_personal	uniper_id	unif_id
		cedula

Nota. En la tabla se muestra un resumen de las relaciones generadas entre las tablas que se han cargado en el Data Mart.

Explicación: Se implemento un mecanismo que tiene la capacidad de integrar datos que previamente son recolectados de múltiples fuentes, lo cual asegurara la disponibilidad de la información y a su vez su accesibilidad. Como se puede observar en la tabla, se consolido la información en múltiples tablas y se identificó las claves foráneas y claves primarias de cada una de las tablas que han sido cargadas como se muestra en la tabla.

Objetivo 2: Desarrollar cubos de información requeridos:

Cumplimiento: Sí.

Evidencia: Mediante el empleo de DAX (Data Analysis Expressions) como una alternativa al manejo de cubos de información tradicionales. Se empleo el uso de funciones DAX, para realizar cálculos y manipulaciones de datos directamente dentro de los modelos de datos. Permitiendo de esta manera crear medidas personalizadas, definir relaciones entre tablas y

llevar a cabo análisis complejos sin la necesidad de construir cubos multidimensionales convencionales, como se muestra en la sección Cubos de información.

Explicación: La decisión de utilizar para este proyecto funciones DAX en lugar del desarrollo de cubos de datos tradicionales se debe a la flexibilidad y eficiencia que estas ofrecen, ya que posibilitan la integración de lógica de negocio personalizada de manera rápida y directa, agilizando el desarrollo y disminuyendo la complejidad del modelo de datos. Además, el empleo de funciones DAX simplificó la realización de análisis detallados y la extracción de información relevante directamente desde el modelo de datos, eliminando la necesidad de configurar un cubo multidimensional independiente. Esto simplificó el proceso de análisis y proporcionó a la dirección una solución más dinámica y adaptable para la toma de decisiones.

Objetivo 3: Generar información para la dirección

Cumplimiento: Sí.

Evidencia: Se desarrollaron paneles de control que permiten visualizar de manera gráfica los requerimientos especificados en el Anexo 2, haciendo uso de la información que previamente ha sido tratada y la incorporación de funciones DAX. Estos paneles de control se presentan de una manera intuitiva, mediante la incorporación de elementos visuales como gráficos, tablas y métricas clave. Los resultados obtenidos en referencia a los paneles de control, se encuentran disponibles en el Capítulo V en la sección “Resultados obtenidos”.

Explicación: Se aplicaron diversas técnicas especializadas en la creación de paneles de control, para lo cual, se utilizó la herramienta de Power BI, para la representación de la data de manera dinámica y personalizada, facilitando de esta manera la comprensión y la toma de decisiones por parte de la dirección de Talento Humano del Gobierno Autónomo Descentralizado del Cantón Guaranda.

La consecución satisfactoria de cada uno de los objetivos específicos antes mencionados permitió el cumplimiento del objetivo general, que se encuentra planteado en el Capítulo I, sección Objetivos.

Validación de la hipótesis

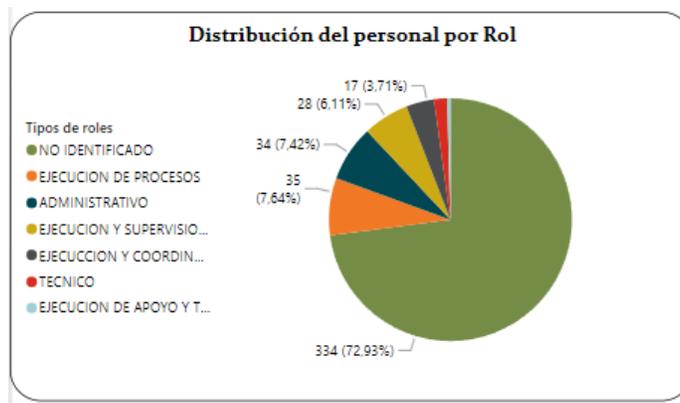
Para la validación de la hipótesis planteada:

“Los Dashboard o paneles de control que se desarrollaran para representar de manera visual los datos almacenados dentro del Data Mart, mejorarán sustancialmente el proceso relacionado a la toma de decisiones, debido a que el análisis de la información presentada en el mecanismo posibilite el desarrollo de estrategias permitiendo aportar de manera positiva al desarrollo institucional. Además, permitirá presentar los datos de manera intuitiva y de igual forma permitirá que los gráficos presentados se ajusten dinámicamente conforme incremente la data del Data Mart.”

Se realizó la consolidación de la información procesada en un Data Mart del Departamento de Talento Humano, el cual será beneficioso para la disponibilidad de la información. Una vez la información se encontró albergada en el Data Mart se procedió con su análisis mediante la utilización de funciones DAX, cuya finalidad es emular el funcionamiento de los cubos de información tradicional, las mismas que fueron utilizadas para consultas y análisis de datos complejos. Permitiendo de esta manera obtener los resultados acordes a los requerimientos que se encuentran en el Anexo 2, a continuación, se presenta un extracto de los resultados obtenidos en los paneles de control:

Figura 32

Distribución del personal por Rol



Nota. En la figura se observa un gráfico estadístico desarrollado en Power BI.

Mediante el desarrollo análisis de los datos se puede agrupar la información obtenida y cálculo de datos, de manera que la información se presente de manera visual y sea intuitiva y eficiente en el proceso de toma de decisiones, como se puede observar en la figura 32, donde se presenta la información relacionada a la distribución de la nómina según el rol que ocupan.

Figura 33

Gasto de vestimenta por Rol

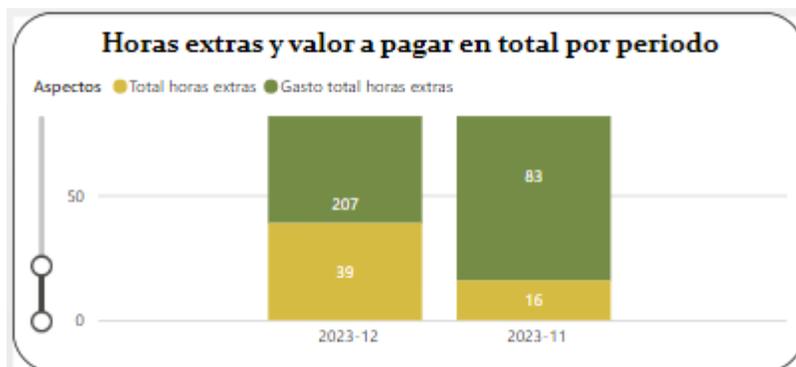


Nota. En la figura se observa un gráfico estadístico desarrollado en Power BI.

En la figura 33, se puede observar la gráfica donde se agrupa los distintos tipos de roles que cumple el personal dentro de la municipalidad y el gasto realizado en las prendas de vestir entregadas. De esta manera facilitando la presentación de la información relacionada a los roles de los empleados.

Figura 34

Valor a pagar por las horas extras



Nota. En la figura se observa un gráfico estadístico desarrollado en Power BI.

En la figura 34, se planteó el gráfico estadístico el cual tiene la capacidad de combinar el total de horas extras realizadas en un periodo y el valor que se ha empleado en las mismas.

Estos gráficos situados en los paneles de control permiten presentar la información almacenada dentro del Data Mart, mostrándose de manera intuitiva, dinámica y clara a los encargados de los procesos de toma de decisiones, permitiendo de esta manera cumplir con los requisitos proporcionado por los usuarios, los mismos que se encuentran en el Anexo 2, permitiendo de esta manera generar un aporte positivo a la institución.

Por otro lado, se planteó una encuesta la misma que se encuentra en el Anexo 3, la cual consta de siete preguntas, las cuales fueron aplicadas a las personas beneficiarias del mecanismo dentro del área de Talento Humano del Gobierno Autónomo Descentralizado del Cantón Guaranda. En base a los resultados obtenidos de la percepción del mecanismo desarrollado, se obtuvieron las siguientes observaciones:

- Mejora sustancial en el proceso de Toma de Decisiones dentro del área de Talento Humano del Gobierno Autónomo Descentralizado del Cantón Guaranda, debido a que los resultados obtenidos los encuestados manifiestan obtener un impacto positivo, obteniendo eficacia en la Toma de Decisiones.

- Los encuestados manifestaron que los paneles de control desarrollados son intuitivos, presentando información útil que es empleada en las funciones que realizan y al respecto ninguno de los encuestados demostró su descontento. Demostrando de tal manera que el mecanismo desarrollado cumple con las expectativas de los encuestados.
- Los participantes de la encuesta coinciden en que el resultado del proyecto cumple con los requisitos establecidos inicialmente, haciendo hincapié a que el mencionado mecanismo puede ser expandido a las demás áreas del Gobierno Autónomo Descentralizado del Cantón Guaranda, debido al impacto positivo y estratégico que genera la implementación de los paneles de control.

Mediante los resultados obtenidos en la encuesta aplicada se puede apreciar una percepción mayoritariamente positiva, en cuanto al desarrollo de los paneles de control implementados.

Los resultados obtenidos tanto en el desarrollo de los requerimientos y en base a la encuesta realizada deslumbran el cumplimiento satisfactorio de la hipótesis planteada inicialmente. Sin embargo, es importante tener en cuenta la necesidad de realizar una evaluación más amplia y a largo plazo para confirmar estos hallazgos.

Capítulo VI: Conclusiones y recomendaciones

Conclusiones

Una vez concluido el presente trabajo de integración curricular se puede determinar que:

- Mediante procesos de integración de la información, el cual se lo realizó de manera cuidadosa se pudo almacenar la información relevante del Departamento de Talento Humano del Gobierno Autónomo Descentralizado del Cantón Guaranda en el Data Mart asegurando de esta manera la disponibilidad de la misma, para los distintos procesos que se deseen realizar con la Data.
- Mediante el uso de funciones DAX se puede emular la funcionalidad de los cubos de datos en base a la información previamente almacenada en el Data Mart, facilitando de esta manera las consultas y análisis complejos realizados para cumplir con los requerimientos identificados por el Departamento de Talento Humano del Gobierno Autónomo Descentralizado del Cantón Guaranda.
- Con el análisis de información y mediante el uso de herramientas de Inteligencia de negocios como Power BI se ha obtenido información relevante para el Departamento de Talento Humano del Gobierno Autónomo Descentralizado del Cantón Guaranda, debido a que la misma se muestra de manera visual mediante el uso de paneles de control, presentándose de una manera clara y eficiente, con la finalidad de respaldar el proceso de toma de decisiones informada.
- Mediante la encuesta de satisfacción que se encuentra en el Anexo 3, la misma que fue realizada a las personas beneficiarias del mecanismo desarrollado, se puede establecer que es un mecanismo de gran utilidad para la toma de decisiones informada dentro de los procesos manejados por el Departamento de Talento Humano del Gobierno Autónomo Descentralizado del Cantón Guaranda.

Recomendaciones

Finalmente, una vez finalizado el trabajo de integración curricular se puede recomendar:

- Es indispensable que los procesos de estratégicos como lo son los de Inteligencia de negocios vayan a la par de los procesos cotidianos de la organización, debido a que, esto garantizará la obtención de la información relevante, que será utilizada como respaldo en los procesos de toma de decisiones.
- Se sugiere mejor el tratamiento y mantener un formato adecuado de las bases de datos, de manera que se pueda facilitar y optimizar el análisis en proyectos futuros. Lo cual implica seguir una estandarización de la data que se recopila en los distintos procesos propios del área.
- Se recomienda la implantación de un Data Warehouse en la Municipalidad de Guaranda debido a que, proporcionara el acceso a la información relevante, permitiendo de esta manera mejorar los distintos procesos que se realizan en el Gobierno Autónomo Descentralizado del Cantón Guaranda.
- Se recomienda tener data histórica, lo cual pueda ser beneficioso para la obtención de patrones que permitan tomar decisiones oportunas avaladas con la información relevante.

Referencias

- Acosta, V. (17 de Abril de 2023). *Descubre la principal diferencia entre data mart y data warehouse*. Obtenido de Inesem Business School:
<https://www.inesem.es/revistadigital/informatica-y-tics/diferencia-entre-data-mart-y-data-warehouse/>
- Ávila , C., & Chiquito, J. (2022). La integración de datamart con datawarehouse. . *UNESUM-Ciencias. Revista Científica Multidisciplinaria*, 8. Obtenido de
<https://revistas.unesum.edu.ec/index.php/unesumciencias/article/view/470/501>
- AWS. (s.f.). *¿Qué es la minería de datos?* Obtenido de <https://aws.amazon.com/es/what-is/data-mining/>
- AWS. (s.f.). *¿Qué es un data mart?* Obtenido de <https://aws.amazon.com/es/what-is/data-mart/>
- Ayala, K. (2021). Diseño e implementación de un Datamart para el área de análisis de una instancia técnica del sector educativo. (*Trabajo de Titulación*). Universidad Agraria la Molina, La Molina. Obtenido de
<https://repositorio.lamolina.edu.pe/bitstream/handle/20.500.12996/5205/ayala-flores-karen-angela.pdf?sequence=1&isAllowed=y>
- Beltran, B. (2018). Minería de datos. (*Trabajo de titulación*). Benemérita Universidad Autónoma de Puebla, Puebla. Obtenido de <http://bbeltran.cs.buap.mx/NotasMD.pdf>
- Benítez de Vendrell, B. (2009). Información para la toma de decisiones., (pág. 17).
- Bernabeu, R. (19 de 07 de 2010). *HEFESTO DATA WAREHOUSING: Investigación y Sistematización HEFESTO: Metodología para la Construcción de un*. Obtenido de
<https://www.businessintelligence.info/resources/assets/hefesto-v2.pdf>
- Carrión, J. (s.f.). *UDGVIRTUAL*. Obtenido de
<http://148.202.167.116:8080/jspui/handle/123456789/869>
- Castro, Y. (2013). Metodologías para desarrollar Almacén de Datos. . *Revista Arquitectura e Ingeniería.*, 12.

- Edith, G. (04 de Diciembre de 2020). *Las 5 V del Big Data: cómo pueden beneficiar a las empresas*. Obtenido de <https://empresas.blogthinkbig.com/5-v-big-data/>
- Escalante, M. (29 de 07 de 2023). *Por dónde comenzar para instalar Python*. Obtenido de abcxperts: <https://abcperts.com/por-donde-comenzar-para-instalar-python/#:~:text=RAM%3A%20M%C3%ADnimo%204GB%20de%20RAM,incluyendo%20Windows%2C%20MacOS%20y%20Linux>
- Etecé, E. (05 de Agosot de 2021). *Concepto*. Obtenido de <https://concepto.de/dato-en-informatica/>
- Gil, M. (s.f.). *Análisis de las Ventajas y Desventajas de Tableau: ¿ Vale la pena utilizar esta herramienta?* Obtenido de Ventajas/Desventajas: <https://ventajasydesventajastop.com/tableau-ventajas-y-desventajas/>
- Gil, M. B. (s.f.). *Power BI: Análisis de las Ventajas y Desventajas que Debes Conocer*. Obtenido de Ventajas/Desventajas: <https://ventajasydesventajastop.com/power-bi-ventajas-y-desventajas/>
- Giraldo, L. (09 de Marzo de 2022). *La importancia de la inteligencia de negocios*. Obtenido de Universidad Espiritu Santo: <https://uees.edu.ec/la-importancia-de-la-inteligencia-de-negocios/>
- Hernández, S. (2017). *Comparativo de metodologías y herramientas para el desarrollo de un Data Warehouse. . (Trabajo de grado)*. Universidad de Manizales, Manizales. Obtenido de https://ridum.umanizales.edu.co/bitstream/handle/20.500.12746/3684/Hern%C3%A1ndez_Mej%C3%ADa_Santiago_2017.pdf?sequence=1&isAllowed=y
- Ibarra, D. (29 de 01 de 2023). *Ventajas y desventajas de MySQL*. Obtenido de CUNOTICIAS.COM: <https://www.cunoticias.com/internet/ventajas-y-desventajas-de-mysql.php>
- IBM. (s.f.). *¿ Qué es un data mart?* Obtenido de <https://www.ibm.com/mx-es/topics/data-mart>

- IBM. (28 de 02 de 2021). *Documentacion*. Obtenido de Visión general de InfoSphere DataStage: <https://www.ibm.com/docs/es/iis/11.5?topic=qualitystage-overview-infosphere-datastage>
- IBM. (28 de 02 de 2021). *InfoSphere DataStage*. Obtenido de InfoSphere Information Server: <https://www.ibm.com/docs/es/iis/11.5?topic=suite-infosphere-datastage>
- IBM. (15 de 11 de 2023). *IBM Data Studio*. Obtenido de https://www.ibm.com/docs/es/db2/11.5?topic=SSEPGG_11.5.0/com.ibm.db2.luw.idm.tools.doc/doc/c0057028.htm
- IBM. (s.f.). *ETL (extraer, transformar, cargar)*. Obtenido de <https://www.ibm.com/mx-es/topics/etl>
- IBM. (s.f.). *Operating systems for a specific product*. Obtenido de Software Product Compatibility Reports : <https://www.ibm.com/software/reports/compatibility/clarity/osForProduct.html>
- KeepCoding. (27 de 12 de 2023). *Ventajas y desventajas de Python*. Obtenido de KeepCoding: <https://keepcoding.io/blog/ventajas-y-desventajas-de-python/>
- Magic Quadrant. (s.f.). *Gartner*. Obtenido de <https://www.gartner.es/es/metodologias/magic-quadrant>
- Marín, R. (22 de 04 de 2023). *Qlikview, una herramienta eficiente para visualización de datos*. Obtenido de INESEM BUSINESS SCHOOL: <https://www.inesem.es/revistadigital/informatica-y-tics/qlikview/>
- Mejias, M. (08 de Marzo de 2023). *Ciclo de vida de los datos: qué es y cuáles son sus fases*. Obtenido de Cyberclick: <https://www.cyberclick.es/numerical-blog/ciclo-de-vida-de-los-datos-que-es-y-cuales-son-sus-fases>
- Microsoft Learn. (s.f.). *Power BI*. Obtenido de <https://learn.microsoft.com/es-es/power-bi/fundamentals/desktop-get-the-desktop>

MySQL. (s.f.). 3.2.1 *System Requirements*. Obtenido de MySQL:

<https://dev.mysql.com/doc/mysql-monitor/8.0/en/system-prereqs-reference.html>

Narvaez, M. (s.f.). *Ciclo de vida de los datos: Qué es y qué etapas tiene*. Obtenido de

QuestionPro: <https://www.questionpro.com/blog/es/ciclo-de-vida-de-los-datos/>

Oracle Corporation. (s.f.). *Requisitos de hardware y software*. Obtenido de Oracle Corporation:

<https://docs.oracle.com/cd/E19879-01/821-1040/abpaj/index.html>

Oracle Database: definición y funcionamiento. (16 de 02 de 2022). Obtenido de Digital Guide

IONOS: <https://www.ionos.es/digitalguide/hosting/cuestiones-tecnicas/oracle-database/>

Oracle RDBMS. (18 de 02 de 2022). Obtenido de DbExperts.Tech:

<https://dbaexperts.tech/wp/database/oracle-rdbms/>

Ortiz, J. (2021). *Diseño y generación del almacén de datos aplicando etl para instituciones de salud. (Trabajo de titulación)*. Universidad Nacional de Chimborazo, Chimborazo.

Obtenido de

http://dspace.unach.edu.ec/bitstream/51000/8159/1/8_Trabajo%20de%20Titulaci%c3

OvHcloud. (s.f.). *¿Qué es PostgreSQL?* Obtenido de OvHcloud:

<https://www.ovhcloud.com/es/lp/postgresql-definition/>

Perez, A. (02 de Diciembre de 2020). *¿Qué tipos de datos se utilizan en el Big Data?* Obtenido

de <https://www.obsbusiness.school/blog/que-tipos-de-datos-se-utilizan-en-el-big-data>

Qlik. (05 de 2023). *Requisitos del sistema para QlikView*. Obtenido de Qlik help:

https://help.qlik.com/es-ES/qlikview/May2023/Content/QV_HelpSites/System-requirements.htm#

Rodriguez, F. (s.f.). *Proceso ETL con Python desde cero y paso a paso*. Obtenido de Data Management Business Intelligence & IT:

<https://datamanagement.es/2019/10/13/proceso-etl-con-python-desde-cero-y-paso-a-paso/>

- Rodriguez, L. (2020). Desarrollo de un sistema de inteligencia de negocio aplicado a la gestion de viajes. . (*Trabajo de grado*). Universidad de Coruña, Coruña. Obtenido de https://ruc.udc.es/dspace/bitstream/handle/2183/26174/L.Rodr%c3%adguez_Silva_2020_Desarrollo_de_un_sistema_de_inteligencia.pdf?sequence=3&isAllowed=y
- Santos, D. (s.f.). *¿Qué es un data mart? Características y ejemplos*. Obtenido de HubSpot: <https://blog.hubspot.es/marketing/que-es-data-mart>
- Segovia, J. (30 de 08 de 2018). *Ventajas y Desventajas de PostgreSQL*. Obtenido de Todo PostgreSQL: <https://www.todopostgresql.com/ventajas-y-desventajas-de-postgresql/>
- Sepúlveda, J. (2018). *Retos de la investigación en ingeniería de sistemas: aplicaciones, herramientas y desarrollos*. Medellín: Corporación Universitaria Americana.
- Silva, G., Córdova, A., Cusco, V., & Estrada, M. (2021). Implementación de un Data Warehouse mediante la metodologí-a Hefestos para la toma de decisiones en el Instituto Nacional de Patrimonio Cultural Regional 3. *Dominio De Las Ciencias*, 135.
doi:<https://doi.org/10.23857/dc.v7i3.2044>
- Silva, G., Zapata , V., Morales , K., & Toaquiza , L. (2019). Análisis de metodologías para desarrollar Data Warehouse aplicado a la toma de decisiones. *Ciencia Digital*, 418.
doi:<https://doi.org/10.33262/cienciadigital.v3i3.4.922>
- Tableau. (s.f.). *¿Qué es Business Intelligence o inteligencia de negocios?* Obtenido de <https://www.tableau.com/es-es/learn/articles/business-intelligence>
- Tableau. (s.f.). *¿Qué es el Business Intelligence? Tu guía para la inteligencia de negocios y por qué es importante*. Obtenido de <https://www.tableau.com/es-mx/learn/articles/business-intelligence>
- Tableau. (s.f.). *¿Qué es Tableau?* Obtenido de <https://www.tableau.com/es-es/why-tableau/what-is-tableau>
- Tableau. (s.f.). *Especificaciones técnicas*. Obtenido de <https://www.tableau.com/es-es/products/techspecs>

- Traiman, R. (2021). Implementación de un Data Mart para la ayuda en la toma de decisiones de la Gerencia de Recursos Humanos. . (*Trabajo de titulación*). Universidad Nacional de Río Negro, Río Negro. Obtenido de https://rid.unrn.edu.ar/bitstream/20.500.12049/8147/1/Traiman%20Schroh_Rossana-2021.pdf
- Trujillo, O. (2018). Construcción de un data mart para el cálculo de indicadores de calidad del servicio en el área de gestión operativa de la CHEC S.A. E.S.P. . (*Trabajo de Maestría*). Universidad Autónoma de Manizales, Manizales. Obtenido de https://repositorio.autonoma.edu.co/bitstream/11182/719/1/Construcci%c3%b3n_data_
- Vargas, H. (09 de 02 de 2017). *Requisitos para instalación de postgres en windows*. Obtenido de Postgresql: https://www.postgresql.org/message-id/CAN3Qy4qpgG3Co%3DWkRqFifo-tW_ppS%3DxN7aobcQ_h%2BUp2qLhG3w%40mail.gmail.com
- Vega, J. (2020). Datos, Ciencia e Ingeniería. *Ingeniare. Revista chilena de ingeniería*, 4. doi:<https://dx.doi.org/10.4067/S0718-33052020000100002>
- Velasco, F. (2020). Datamart para el proceso de Administración Tributaria en el área de Rentas y Finanzas de la Municipalidad de Puente Piedra. . (*Trabajo de titulación*). Universidad Cesar Vallejo, Puente Piedra. Obtenido de <https://repositorio.ucv.edu.pe/handle/20.500.12692/51928>
- Vento, A. (2020). Datamart para evaluar la cobranza en la Municipalidad Distrital de Ate. (*Trabajo de Titulación*). Universidad Cesar Vallejo, Ate. Obtenido de https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/52475/Vento_RAD-SD.pdf?sequence=1&isAllowed=y
- Zas, A. (2018). Construcción y explotación de un Data Warehouse para el análisis de información de una empresa de telecomunicaciones. . (*Trabajo de titulación*).

Universidad de Coruña, Coruña. Obtenido de

https://ruc.udc.es/dspace/bitstream/handle/2183/32029/ZasVazquez_AliciaAmadora_T

Apéndices