

Práctica Empresarial: Desarrollo de un Data Mart en el Área de Gestión Operativa de
Essa y Guía de Trabajo para su Implementación en Otras Áreas

Paula Andrea Castro Mendoza y Paula Catalina Hernández Ramírez
Trabajo de Grado para Optar al Título de Ingenieras de Sistemas

Director

Luis Carlos Gómez Flórez

Magíster en Ingeniería de Sistemas e Informática

Tutor

Víctor Rangel Álvarez

Ingeniero Industrial e Ingeniero de Mercados

Universidad Industrial de Santander

Facultad de Ingenierías Físico-Mecánicas

Escuela de Ingeniería de Sistemas e Informática

Ingeniería de Sistemas e Informática

Bucaramanga

2023

Dedicatoria

Dedicamos este trabajo a Jehová Dios por habernos acompañado en todo momento de nuestras vidas y guiarnos por el camino correcto. Que los frutos de este trabajo sean usados y moldeados por él como el Gran Alfarero (Jer. 18:6).

Agradecimientos

Nos complace expresar nuestro más sincero agradecimiento a todas las personas que brindaron su apoyo y contribuyeron de manera significativa en la realización de este proyecto de grado, el cual representa el resultado de un arduo trabajo y dedicación.

Yo Paula Andrea, quiero agradecer a mi madre, Angela Patricia, por su constante cuidado y apoyo, que hicieron posible mis estudios y mi crecimiento personal, a mi Padre Carlos Castro, a mi familia por ser mi lugar seguro y a mi amiga Paula.

Yo Paula Catalina, agradezco mucho a Jehová Dios por su amor y guía; a mis queridos papás: Henry, Socorro y a Cami, por su amor, y apoyo; y agradezco a mi amiga y compañera Pau.

Agradecemos de manera especial a nuestro director de proyecto, Luis Carlos Gómez Flórez, quien no solo nos ofreció su tiempo y guía, sino también su calidad humana y magnitud como profesional durante el tiempo como director y profesor.

Agradecemos a nuestros tutores, Víctor Rangel y Román Suárez, ambos profesionales de la Electrificadora de Santander, por su generosidad al invertir su tiempo y brindarnos su valiosa orientación a lo largo de todo el proceso. Su acompañamiento fue fundamental para iniciar en el mundo laboral y aplicar nuestros conocimientos académicos de manera acertada. También extendemos nuestro agradecimiento a la Electrificadora de Santander por proporcionarnos la oportunidad de llevar a cabo nuestra práctica empresarial en sus instalaciones.

Este logro no habría sido posible sin el apoyo de nuestras familias y amigos, cuyo apoyo emocional fue un pilar fundamental para superar desafíos y mantenernos enfocadas en nuestros objetivos. Asimismo, deseamos expresar nuestro agradecimiento a nuestra alma mater por

brindarnos una educación de calidad que nos proporcionó las herramientas y conocimientos necesarios para llevar a cabo este proyecto.

Tabla de Contenido

Introducción 14

1. Presentación 16

1.1 Descripción de la Empresa..... 16

1.2 Objetivos 18

1.2.1 Objetivo general..... 18

1.2.2 Objetivos específicos 18

2. Marco de referencia 19

2.1 Inteligencia de Negocios..... 19

2.2 Data Warehouse 20

2.3 Data Mart 21

3. Investigación y Marco de Trabajo 22

3.1 Revisión de Literatura..... 24

3.2 Metodología Escogida 25

3.2.1 Planificación del proyecto..... 29

3.2.2 Gestión del Proyecto 30

3.2.3 Definición de Requerimientos de Negocio 30

3.2.4 Diseño de la Arquitectura Técnica..... 32

3.2.5 Selección e Instalación de Productos 33

3.2.6 Diseño Físico 33

3.2.7 Diseño y Desarrollo del ETL 33

4. Gestión del Proyecto y Definición de Requerimientos de Negocio. 35

4.1 Planificación del proyecto.....	35
4.2 Gestión del Proyecto	36
4.3 Identificación de Procesos Críticos.....	37
4.3.1 Entrevistas para Requerimientos.....	40
4.3.2 Entrevistas Centradas en Datos.....	42
4.4 Priorización de Requerimientos y Establecimiento del Alcance del Proyecto	43
5. Arquitectura Técnica y Herramientas	51
5.1 Back Room.....	51
5.2 Front Room	55
6. Diseño	56
6.1 Diseño físico de la base de datos	56
6.1.1 Definición de estándares para objetos físicos	56
6.1.2 Diseño de base de datos física	57
6.1.2.1 Diseño lógico a físico para área de Gestión Operativa.	58
6.1.2.1.1 Consulta Eventos.	58
6.1.2.1.2 Consulta Consignaciones.	59
6.1.2.1.3 Consulta Trafos por Vereda.	60
6.1.2.2 Diseño lógico a físico para área de Subgerencia de Conexiones.....	62
6.2 Diseño del subsistema ETL	64
6.2.1 Gestión Operativa	66
6.2.1.1 Consulta Eventos.	66
6.2.1.2 Consulta Consignaciones	67
6.2.1.3 Consulta Trafos por vereda.....	68

6.2.2 Subgerencia de Conexiones	68
6.2.2.1 Consulta Recuperación de Energía Eléctrica	69
6.3 Diseño de la Retención de Datos	70
7. Desarrollo.....	72
7.1 Desarrollo del Subsistema ETL	72
7.2 Pruebas de ETL.....	76
7.2.1 Consulta de Eventos.....	77
7.2.2 Consulta de Consignaciones	77
7.2.3 Consulta de Trafos por Vereda	77
7.2.4 Consulta de Recuperación de Energía	78
7.3 Desafíos.....	78
7.3.1 ETL	78
7.3.2 Automatización	81
7.4 Creación de trabajos para ETL	81
8. Guía de Trabajo para Data Marts.....	83
8.1 Identificación de la necesidad.....	83
8.2 Estructura y contenido de la guía de trabajo.....	84
8.3 Impacto de la Guía de Trabajo.....	85
8.4 Conferencias Informativas	86
9. Conclusiones	87
Referencias Bibliográficas	89
Apéndices.....	91

Lista de Tablas

	Pág.
Tabla 1 Objetivos específicos y su cumplimiento	18
Tabla 2 Especificaciones servidor ESSA-TEST18.....	54
Tabla 3 Lista de atributos de la tabla AGO_consulta_eventos	59
Tabla 4 Lista de atributos de la tabla AGO_consulta_consignaciones.....	60
Tabla 5 Lista de atributos de la tabla AGO_consulta_trafosporvereda	62
Tabla 6 Lista de atributos de la tabla PER_consulta_recuperacion	63
Tabla 7 Tabla de compatibilidad entre SQL Server y el servicio de integración Microsoft	82

Lista de Figuras

	Pág.
Figura 1 Estructura organizacional de la ESSA.....	17
Figura 2 Levantamiento de necesidades de automatización hecho por ESSA	23
Figura 3 Arquitectura Bill Inmon	25
Figura 4 Arquitectura Ralph Kimball	26
Figura 5 Diagrama del ciclo de vida de Kimball	29
Figura 6 Los requerimientos de negocio impactan virtualmente cada aspecto del proyecto.....	31
Figura 7 Funcionalidad ETL.....	34
Figura 8 Flujo de identificación de procesos críticos en el proyecto.....	39
Figura 9 Diagrama jerárquico de entrevistas.	40
Figura 10 Carga de datos a Power BI para el área de Gestión Operativa.	48
Figura 11 Carga de datos a Power BI para el área de Subgerencia de Conexiones.....	49
Figura 12 Carga de datos a Power BI para el área de Subgerencia de Conexiones.....	50
Figura 13 Caja de herramientas de SQL Server Integration Services	52
Figura 14 Arquitectura técnica final del proyecto	55
Figura 15 Modelo lógico para la consulta Eventos del área de Gestión Operativa	58
Figura 16 Modelo lógico para la consulta Consignaciones del área de Gestión Operativa.....	59
Figura 17 Modelo lógico para la consulta Trafos por Vereda del área de Gestión Operativa.....	61
Figura 18 Modelo lógico para la consulta Recuperación SAC del área de Gestión de Pérdidas .	62
Figura 19 Flujo de actividades para mejoras SQL.....	64
Figura 20 Diseño de la BD destino	65

Figura 21 Extracción Oracle BD para Gestión Operativa.	66
Figura 22 Extracción Oracle BD para Subgerencia de Conexiones.	69
Figura 23 Retención de datos consulta Eventos	70
Figura 24 Retención de datos consulta Consignaciones	71
Figura 25 Retención de datos consulta recuperación.....	71
Figura 26 Paquete de integración de servicios para la consulta Eventos del área de Gestión Operativa.....	74
Figura 27 Paquete de integración de datos	75
Figura 28 Contenedor de Secuencias	76
Figura 29 Componentes dentro de una tarea de flujo de datos	76
Figura 30 Configuración del componente de tipo ODBC	80

Lista de Apéndices

	pág.
Apéndice A. Resumen de la guía de trabajo para el desarrollo de data marts	91

Resumen

Título: Práctica Empresarial: Desarrollo de un Data Mart en el Área de Gestión Operativa de ESSA y Guía de Trabajo para su Implementación en otras Áreas.*

Autor: Paula Andrea Castro Mendoza y Paula Catalina Hernández Ramírez.**

Palabras Clave: Data mart, analítica de datos, ingeniería de datos, Metodología Kimball.

Descripción: En la era digital, los datos se han convertido en un recurso vital para las organizaciones. Conscientes de su importancia, la Electricadora de Santander S.A. E.S.P. ESSA, perteneciente al grupo empresarial de Empresas Públicas de Medellín (EPM) se embarcó en un proyecto por medio de una práctica empresarial con el objetivo de fortalecer sus conocimientos y capacidades en el campo de la analítica de datos. El objetivo era implementar soluciones que pusieran a disposición los datos del negocio para su posterior análisis en los tableros de inteligencia de negocios. El proyecto se enfocó en las áreas de Gestión Operativa y Subgerencia de Conexiones, donde se realizaban actividades manuales de extracción, transformación y carga de datos. Se propuso la implementación de un data mart y se llevó a cabo un análisis de procesos, definición de requerimientos, diseño de arquitectura técnica, la implementación del sistema de Extracción, Transformación y Carga y pruebas para verificar la calidad de los datos. Como resultado, se logró agilizar y mejorar los procesos que realizaban antes las áreas, brindando a las estudiantes la oportunidad de trabajar de la mano con un equipo de soluciones digitales, haber adquirido una valiosa experiencia laboral y contribuyendo a su formación integral como ingenieras.

* Trabajo de Grado

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería de Sistemas e Informática. Ingeniería de Sistemas e Informática. Director: Luis Carlos Gómez Flórez. Magíster en Ingeniería de Sistemas e Informática. Tutor: Víctor Rangel Álvarez. Ingeniero Industrial e Ingeniero de Mercados.

Abstract

Title: Internship: Development of a Data Mart in the Operational Management Area of ESSA and Work Guide for its Implementation in other Areas.*

Author(s): Paula Andrea Castro Mendoza and Paula Catalina Hernández Ramírez.**

Key Words: Data mart, data analytics, data engineering, Kimball Methodology

Description: In the digital age, data has become a vital resource for organizations. Recognizing its importance, Electrificadora de Santander, belonging to the business group Empresas Públicas de Medellín (EPM), embarked on a project through an internship program with the aim of strengthening its knowledge and capabilities in the field of data analytics. The objective was to implement solutions that would make business data available for subsequent analysis on business intelligence dashboards. The project focused on the areas of Operations Management and Connections Submanagement, where manual data extraction, transformation, and loading activities were being carried out. The implementation of a data mart was proposed, and a process analysis, requirements definition, technical architecture design, implementation of the Extraction, Transformation, and Loading system, and data quality testing were carried out. As a result, processes previously performed by these areas were streamlined and improved, providing the students with the opportunity to work closely with a digital solutions team, gain valuable work experience, and contribute to their comprehensive education as engineers.

* Degree Work in Internship modality

** Faculty of Physical-Mechanical Engineering. School of Systems Engineering and Informatics
Director: Luis Carlos Gómez Flórez. Master in Engineering and Informatics. Tutor: Víctor Rangel Alvarez. Industrial Engineer and Market Engineer.

Introducción

En la era digital, los datos se han convertido en un recurso vital para las organizaciones. El reconocido matemático y empresario británico en el área de ciencia de datos, Clive Humby declara: "Los datos son el nuevo petróleo", ejemplificando así lo valiosos que son los datos para las empresas. Y así como para sacar provecho del petróleo hay que hacer un proceso de extracción, la situación no es diferente a la hora de poner a disposición los datos en las organizaciones. Es por ello que resulta fundamental crear proyectos que permitan el aprovechamiento de los datos, mediante el implementar procesos y herramientas que los pongan a disposición para su posterior análisis.

La Electrificadora de Santander S.A. E.S.P. ESSA, perteneciente al grupo empresarial de Empresas Públicas de Medellín (EPM) y con sede principal en Bucaramanga, es una empresa centrada en la generación, distribución y comercialización de servicios públicos de electricidad en Colombia, cuya área de licencia comprende 87 municipios de Santander.

En el 2022, ESSA mediante un diagnóstico técnico de habilidades en analítica de datos identificó brechas en este campo, y con el objetivo de contribuir a cerrarlas se llevó a cabo el actual proyecto mediante la modalidad de práctica empresarial. Una de las brechas identificadas en ese diagnóstico corresponde a la falta de una solución que pusiese a disposición los datos de los procesos del negocio para su correspondiente análisis.

Para el desarrollo del proyecto se identificó esa brecha en una necesidad dentro de Gestión Operativa y Subgerencia de Conexiones, áreas de la empresa en la que se llevan a cabo actividades manuales de extracción de datos para su posterior análisis en el tablero de inteligencia de negocios (BI, por sus siglas en inglés, Business Intelligence). Es así como se propuso la implementación de data marts que pudiesen suministrar los datos precisos y de manera oportuna. Para ello se realizó

un proceso exhaustivo de análisis de los procesos de negocio y definición de requerimientos, además se elaboró el diseño de la arquitectura técnica, infraestructura de software, diseño físico del data mart y finalmente la implementación de su sistema de Extracción, Transformación y Carga (ETL, por sus siglas en inglés, Extract, Transform and Load). Además se propuso la elaboración de una guía de trabajo a partir de la experiencia adquirida durante el proyecto, que sirviese como orientación para el posterior desarrollo de proyectos similares en otras áreas de la empresa, y que asegurase la gestión del conocimiento y la estandarización de buenas prácticas en futuras implementaciones, dado que según la exploración hecha en la empresa, al iniciar la práctica no se contaba con una implementación similar cuya experiencia se encontrase documentada o soportada bajo una guía.

Estas actividades brindaron a las estudiantes una valiosa experiencia en el ámbito laboral, les permitió poner en práctica conocimientos adquiridos durante la carrera y finalmente apoyar y mejorar los procesos de inteligencia de negocios de la empresa, lo que contribuyó a su formación integral como ingenieras.

El contenido de este libro abarca aspectos fundamentales para comprender el desarrollo y los logros alcanzados en el proyecto de la práctica empresarial llevada a cabo en la Electrificadora de Santander. Desde un marco teórico que resalta la importancia de los datos en la era digital hasta la implementación de un sistema de Extracción, Transformación y Carga (ETL) y la creación de una guía de trabajo para data marts. Además, se explica la planificación y gestión del proyecto, el diseño de bases de datos y subsistemas ETL que se llevaron a cabo. El libro concluye con un resumen de los logros y aprendizajes del proyecto.

1. Presentación

1.1 Descripción de la Empresa

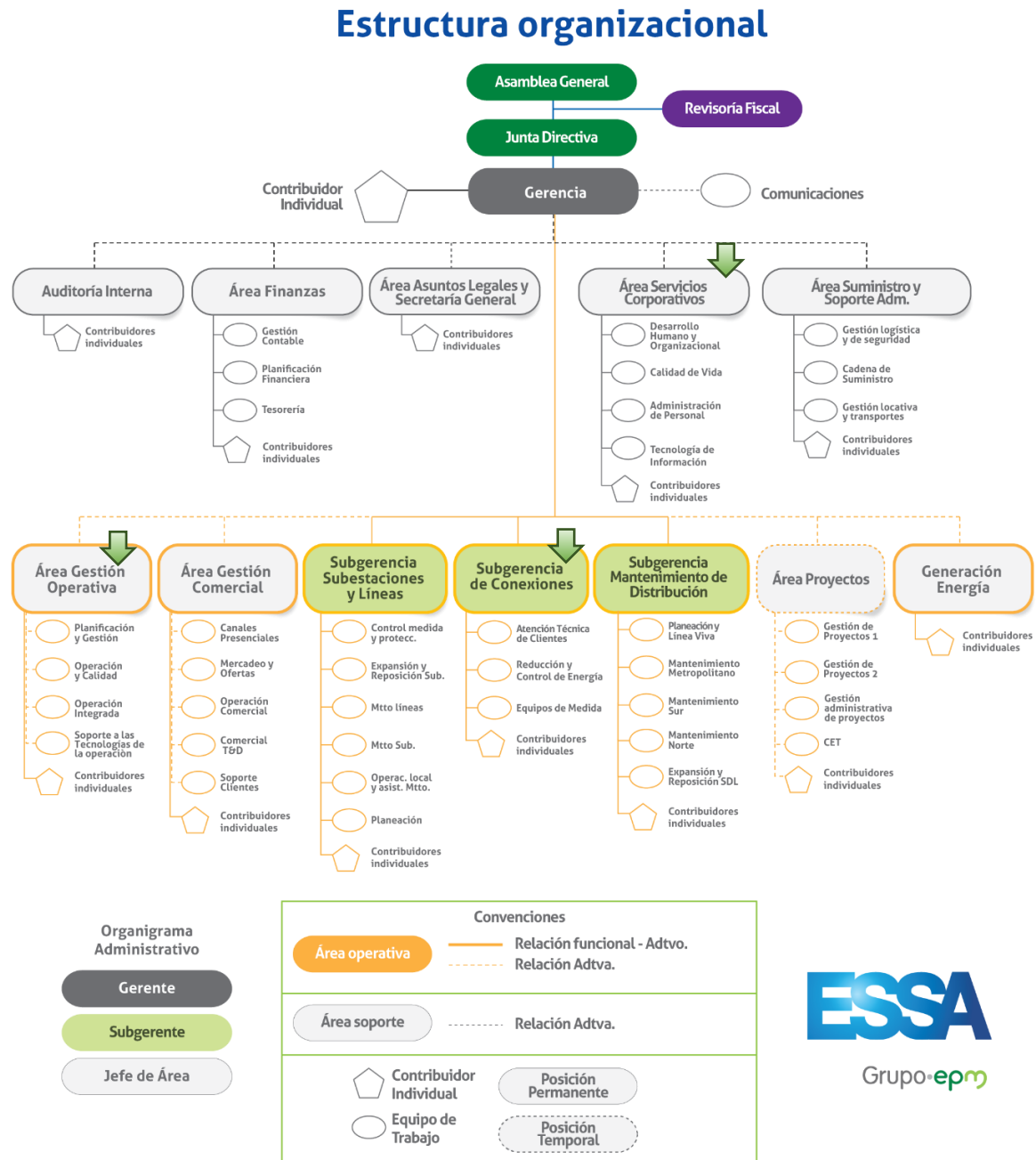
La Electrificadora de Santander S.A. E.S.P. ESSA es una empresa colombiana de servicios públicos mixta dedicada a la transmisión, distribución y comercialización de energía eléctrica. Perteneciente al Grupo Empresarial EPM, esta se ubica en la ciudad de Bucaramanga, Colombia (ESSA, s. f.).

La ESSA es una empresa con una estructura organizacional jerárquica. La compañía cuenta con 13 áreas en total, cada una con sus equipos de trabajo. En la cima de la estructura se encuentra el máximo órgano de administración de la sociedad, la junta directiva, compuesta por tres miembros elegidos por la Asamblea General de Accionistas (ESSA, s. f.). Ver Figura 1, esta representa la jerarquía organizacional en ESSA y permite la identificación de las áreas que fueron beneficiadas por el actual proyecto.

En el contexto de la práctica empresarial, las estudiantes integrantes del proyecto de grado son parte del Área de Servicios Corporativos, y el proyecto data mart se presenta como una solución para el Área de Gestión Operativa, y para la Subgerencia de Conexiones.

Figura 1

Estructura organizacional de la ESSA



Nota. Jerarquía organizacional en ESSA. Tomado de: ESSA. (s.f.). Quiénes somos. Recuperado el 4 de abril de 2023, de <https://www.essa.com.co/site/informacion-corporativa/quienes-somos>

1.2 Objetivos

1.2.1 Objetivo general

Desarrollar un data mart para apoyar la toma de decisiones basadas en datos dentro del área de Gestión Operativa de la Electrificadora de Santander S.A. -ESSA- y proponer una guía de trabajo para la elaboración de data marts en otras áreas de la empresa.

1.2.2 Objetivos específicos

En la Tabla 1 se muestran los objetivos específicos con su respectivo cumplimiento y el capítulo del informe donde se puede encontrar más información sobre cómo se logró cada objetivo.

Tabla 1

Objetivos específicos y su cumplimiento

No.	Objetivo	Cumplimiento	Capítulo
1	Investigar sobre las prácticas y herramientas actuales para la creación de data marts y a partir de la necesidad identificada en la empresa aplicar un marco de trabajo que incorpore las prácticas más idóneas según el punto a desarrollar.	Se investigó sobre las prácticas y herramientas actuales, se evaluaron diversas opciones y a partir de ello se seleccionó un marco de trabajo adecuado para el proyecto.	2. Marco de Referencia, 3. Investigación y Marco de Trabajo
2	Identificar y documentar procesos críticos dentro del área de Gestión Operativa que puedan ser priorizados para el alcance del proyecto.	Sobre los procesos identificados dentro del área de Gestión Operativa se realizó el análisis que como resultado generó la descripción detallada de aquellos procesos. Como punto adicional, se hizo este mismo proceso para el área de Subgerencia de Conexiones.	4. Gestión del Proyecto y Definición de Requerimientos de Negocio

No.	Objetivo	Cumplimiento	Capítulo
3	Elaborar un data mart para el área de Gestión Operativa que permita disponer de los datos para su posterior análisis en inteligencia de negocios.	Se diseñó y desarrolló un data mart para el área de Gestión Operativa. Este permitió poner a disposición los datos para reportes de inteligencia de negocios. Adicionalmente, se hizo el mismo procedimiento para el área de Subgerencia de Conexiones. En total se construyeron dos data marts.	6. Diseño, 7. Desarrollo
4	Desarrollar el proceso de extracción, transformación y carga de la base de datos transaccional al data mart, garantizando calidad y oportunidad.	Se desarrollaron los procesos de extracción, transformación y carga para automatizar la transferencia de datos a los data marts.	6. Diseño, 7. Desarrollo
5	Proponer una guía de trabajo para la elaboración de data marts para orientar y estandarizar futuras implementaciones en otras áreas de la ESSA.	Se elaboró la guía de trabajo para data marts. Esta incluye prácticas, recomendaciones y pautas necesarias para futuras implementaciones en otras áreas de la ESSA. Como entregable, se proporcionó la guía en formato digital y se socializó su contenido mediante conferencias dentro de la empresa.	8. Guía de Trabajo para Data Marts

2. Marco de referencia

En la actualidad, las organizaciones disponen de una gran cantidad de datos como resultado de sus operaciones diarias, lo que se ha convertido en un recurso valioso para la toma de decisiones empresariales, esto debido a que el valor de los datos reside en la información que podemos extraer de ellos, y, por lo tanto, la mejora que produce en la toma de decisiones (Molina y García, 2006). Sin embargo, el volumen de los datos y el nivel de complejidad de ellos hace que su análisis requiera de otras técnicas y herramientas. En este contexto, la inteligencia de negocios, los data warehouse y los data mart se presentan como soluciones fundamentales para el acceso, la gestión y el análisis de los datos empresariales.

2.1 Inteligencia de Negocios

La inteligencia de negocios hace que las organizaciones puedan sacar el mayor provecho de su información poniéndola a disposición de los usuarios y así dar apoyo a la toma de decisiones.

Esto es posible debido a que la inteligencia de negocios implica el convertir datos en conocimiento, y luego aprovecharlo para crear ventaja competitiva del negocio (Medina la Plata, 2012).

Aunque es bien sabido que una de las claves para el éxito en una empresa es la capacidad de tomar decisiones rápidas y precisas, los sistemas de información tradicionales (como los programas de gestión, o aplicaciones a medida) debido a su rigidez no permiten procesos de extracción, integración de datos, almacenamiento de datos históricos y otras funcionalidades más que son necesarias para obtener información (Medina la Plata, 2012). Es por ello que la inteligencia de negocios se apoya en un conjunto de herramientas que faciliten esa obtención de información y por lo tanto de los datos mismos. Pero entonces, ¿qué es a ciencia cierta la inteligencia de negocios?

Curto (2010) define la inteligencia de negocios como “el conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización” (p. 18). Por tanto, las herramientas de inteligencia de negocios incluyen capacidades de análisis multidimensional que facilitan la definición y seguimiento de los indicadores relevantes para el negocio (Medina la Plata, 2012). Esto diferencia los sistemas de inteligencia de negocios de los operacionales. En línea con lo anterior, es necesario en las organizaciones contar con herramientas especializadas que permitan hacer una integración de datos.

2.2 Data Warehouse

En un esfuerzo por potencializar la inteligencia de negocios y facilitar la gestión de datos, surge el concepto de una bodega de datos (data warehouse, en inglés), un repositorio de datos con un enfoque global, común e integrado (Curto, 2010).

Un data warehouse se puede definir como el área donde se recopilan y almacenan datos para su futuro análisis. Su principal característica es su propósito de poner a disposición datos para analítica, y es por ello que está organizado y estructurado para permitir un análisis más complejo y detallado de grandes volúmenes de datos (Ballard et al., 2005), proporcionando así una fuente confiable y oportuna de información para la toma de decisiones empresariales.

Sobre las características de los data warehouse, Curto (2010) menciona las siguientes: orientado a un tema, debido a que organiza datos sobre un tema central; integrado, dado que sus datos tienen múltiples orígenes y aun así sus datos son consistentes; variable en el tiempo, ya que los datos cambian a lo largo de este, pero se mantiene su historial (Yessad y Labiod, 2016); y finalmente, no volátil, puesto que es una base de datos orientada a la consulta de los usuarios finales. Otras características pueden incluir su gran tamaño por la cantidad de datos que almacena y su diseño enfocado y optimizado para analítica (Joyanes, 2019).

Ahora bien, sobre la implementación de un data warehouse surgen diferentes cuestiones: ¿cómo organizarlo?, ¿cuándo incorporarlo?, ¿quién usará estos datos?, ¿cuánto tiempo llevará? Según Ballard et al. (2005), se concluye que esta iniciativa es costosa, ya que requiere una cantidad significativa de recursos y tiempo. Por tanto, motivados por la necesidad imperante de contar con información crítica y aprovechar los beneficios que ofrece un data warehouse, se plantea la implementación de esta solución de inteligencia de negocios de forma eficiente en términos de recursos. Como resultado, surge la idea de implementar data marts.

2.3 Data Mart

Un data mart es una colección de datos organizados para dar apoyo a la toma de decisiones dentro de un área determinada de una empresa. Esta construcción es una evolución del concepto de data warehouse (Ballard et al., 2005), debido a que en las empresas las diferentes áreas del

negocio empezaban a requerir beneficiarse del data warehouse. Así surge el concepto de un data mart, que, en palabras simples, sería un data warehouse que satisfaría las necesidades de un área en particular.

Sobre la estructura de los datos, al igual que en un data warehouse, los datos pueden estar en modelos de estrella o copo de nieve, y también puede que el data mart sea dependiente o independiente del data warehouse (Curto, 2010). Joyanes (2019) refiere que un data mart dependiente es aquel cuyos datos son un subconjunto que proviene del data warehouse, y debido a eso necesita que primero haya alguno construido, y su ventaja es que el usuario final visualiza la misma versión de datos que el resto de los usuarios del data warehouse; diferente del data mart independiente, este es un pequeño data warehouse diseñado para un área o departamento de la empresa y no necesita de la costosa implementación de una data warehouse. Además, menciona que la implementación de una de estas aplicaciones de inteligencia de negocios dependerá del tamaño y estrategia de la empresa, de acuerdo con eso podría determinarse si lo conveniente sería implementar un data mart y cuando el volumen de datos, transacciones y demás lo exija, dar paso a la centralización de estos data marts (Joyanes, 2019).

3. Investigación y Marco de Trabajo

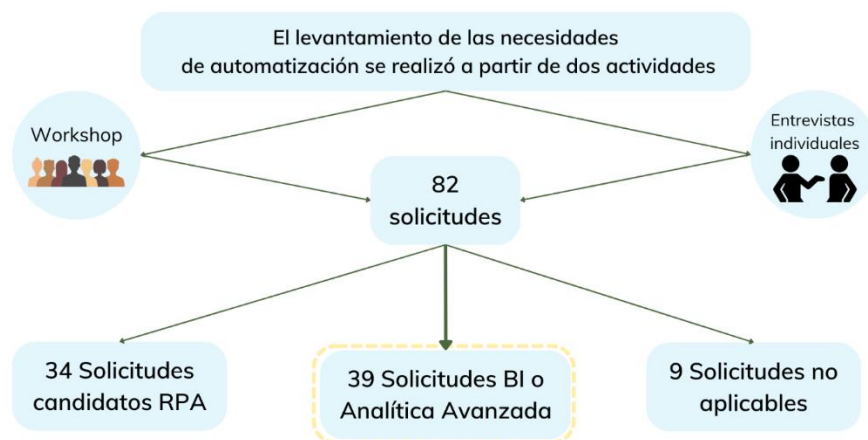
La ESSA actualmente cuenta con un Centro de Excelencia (CoE) de RPA (Robotic Process Automation), el cual se estableció con el propósito de brindar apoyo mediante soluciones digitales a procesos susceptibles de automatización.

A través de un levantamiento de las necesidades de automatización -ver Figura 2- el CoE recibió un total de 82 solicitudes de sus empleados, de las cuales se identificó que 34 correspondían a automatización robótica de procesos, 39 solicitudes para necesidades de analítica avanzada o inteligencia de negocios, y 9 que no eran aplicables para ninguna de las dos áreas. El resultado de

la evaluación de las solicitudes recibidas indicó que la empresa tenía una mayor necesidad de soluciones en el área de inteligencia de negocios y analítica avanzada que en la automatización robótica de procesos. Debido a esto, el CoE tomó la decisión de enfocarse en un proyecto relacionado con la analítica de datos para mejorar aún más la eficiencia y la competitividad de la empresa.

Figura 2

Levantamiento de necesidades de automatización hecho por ESSA



Nota. La figura muestra el proceso del levantamiento de necesidades de automatización en la empresa. Fuente elaborado por los autores.

Para abordar la necesidad identificada se planteó desde la empresa un proyecto que involucraría un data warehouse, y herramientas para procesos ETL (extracción, transformación y carga, por sus siglas en inglés) como parte de una iniciativa de soluciones digitales. Sin embargo, no se contaba con un conocimiento profundo sobre cómo llevar a cabo el proyecto, ni se sabía si su objetivo y justificación estaban alineados con las necesidades actuales de la empresa. Resolver eso implicaba llevar a cabo una revisión de literatura que permitiese definir objetivos y un marco de trabajo para el proyecto, considerando el contexto actual de la empresa y teniendo en cuenta los resultados que arrojaría más adelante un diagnóstico técnico de analítica de datos que se estaba

adelantando para comprender el estado actual de la organización en este tema y determinar los pasos a seguir.

3.1 Revisión de Literatura

La revisión de literatura es un paso que resulta fundamental al querer contar con una base teórica sólida que permita definir de manera clara el problema y los objetivos a alcanzar en el proyecto de práctica empresarial.

Para iniciar esta fase, se identificaron la inteligencia de negocios, el data warehouse y el data mart como conceptos relevantes, para lo cual se hizo uso de herramientas académicas que permitieron realizar una búsqueda de información valiosa y actualizada sobre estos temas. Adicionalmente, se enfocó la revisión de literatura en la búsqueda de una metodología para crear data warehouse y se definió el enfoque a través de la búsqueda de información relevante en bases de datos sobre los temas de interés.

En un principio se usó la Biblioteca Virtual de la UIS, la cual permite el acceso a bases de datos bibliográficas, además del buscador de Google Scholar. Esto se hizo con el fin de encontrar libros electrónicos, artículos y demás, que fuesen de importancia para la práctica empresarial y proporcionaran una guía base para el proyecto. El criterio de selección de las fuentes de información consultadas se basó en la relevancia y actualidad de la información, considerando libros, artículos y proyectos de grado de los últimos 10 años, en idioma inglés y español, o textos que a pesar de su fecha fuesen considerados una obra referencia en el tema, y en la estrategia de búsqueda se utilizaron las palabras clave: inteligencia de negocios, data warehouse, data mart, ETL y SQL Server Integration Services (SSIS).

Para la selección de documentos se tuvo en cuenta la afinidad del título y resumen con el tema, la fecha en que fue publicado, la calidad de la fuente y finalmente, la accesibilidad al

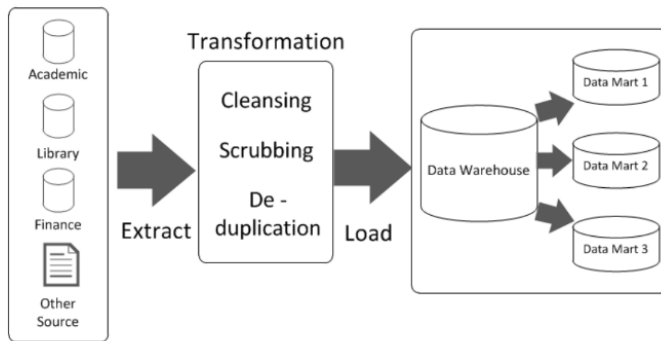
documento completo. Aproximadamente fueron 40 documentos revisados, entre libros, artículos de revistas académicas, y proyectos de grado. De ellos se hizo una revisión de validez, confiabilidad de la información seleccionada y evaluación de la claridad y coherencia en la presentación de las ideas, y según su legibilidad y calidad de ser comprendidos, se seleccionaron 7 libros que se consideró podían aportar una sólida base teórica al proyecto. Finalmente, ello permitió escoger la metodología Kimball como aquella que le daría fundamento al desarrollo del proyecto, y que se explica a continuación.

3.2 Metodología Escogida

Para seleccionar la metodología utilizada en el proyecto, se realizó una comparación entre dos metodologías: la metodología de Bill Inmon y la metodología de Kimball. En el artículo de Yessad y Labiod (2016) que describe la arquitectura del data warehouse de Inmon y el enfoque de Kimball, menciona que la arquitectura de Inmon integra todos los sistemas de información de la empresa en un solo data warehouse, ver Figura 3 , considerando niveles operativo, atómico, departamental e individual. Por otro lado, el enfoque de Kimball se destaca por involucrar fuertemente a los usuarios finales desde las primeras etapas del proyecto, siendo un enfoque impulsado por los requerimientos del usuario y basado en el concepto de data marts compartidos, ver Figura 4.

Figura 3

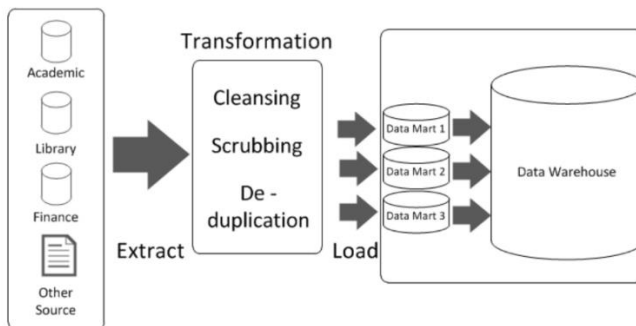
Arquitectura Bill Inmon



Nota. La figura muestra el enfoque de arriba hacia abajo de Bill Inmon para el diseño de un data warehouse. Tomado de Moscoso-Zea, O., & Luján-Mora, S. (2016, September). "Datawarehouse design for educational data mining." En 2016 15th International Conference on Information Technology Based Higher Education and Training (ITHET) (pp. 1-6). IEEE.

Figura 4

Arquitectura Ralph Kimball



Nota. La figura muestra el enfoque ascendente de Ralph Kimball para el diseño de un data warehouse. Tomado de Moscoso-Zea, O., & Luján-Mora, S. (2016, September). "Datawarehouse design for educational data mining." En 2016 15th International Conference on Information Technology Based Higher Education and Training (ITHET) (pp. 1-6). IEEE.

La elección de la metodología se basó en varios factores. En primer lugar, se tuvo en cuenta la necesidad de contar con una arquitectura que permitiera la construcción de data marts y su

posterior integración en un data warehouse, ya que debido a su enfoque en la creación de data marts compartidos, es posible acelerar los tiempos de implementación y tener una mayor capacidad de respuesta a los cambios y requerimientos de la empresa. Además, se consideró la importancia de recolectar los requerimientos del proyecto de manera precisa, y dado que la metodología Kimball se basa en los requerimientos del usuario, esta proporciona muchas ayudas para una buena recolección de requerimientos y se centra en ellos y en las necesidades específicas de los usuarios para el diseño del data warehouse.

La metodología Kimball surge en Metaphor Computer Systems, una empresa innovadora en el campo de la toma de decisiones informadas, en un momento en que no existían prácticas recomendadas ni metodologías formales para brindar el mejor acceso y análisis de datos a los usuarios comerciales. Los autores de esta metodología comenzaron a perfeccionar técnicas y enfoques para abordar los desafíos únicos de proporcionar acceso y permitir el análisis de datos, considerando el crecimiento y la extensibilidad a largo plazo. Como resultado de este esfuerzo, hoy conocemos la Metodología Kimball para el data warehousing y la inteligencia de negocios (Kimball et al., 2008).

Desde ese entonces ha habido una creciente necesidad de soporte a las decisiones en las organizaciones. Esto se debe a que el entorno empresarial ha evolucionado, lo que ha generado un mayor nivel de competencia. Por esta razón, es muy necesario contar con herramientas que permitan a los líderes empresariales tomar decisiones informadas y basadas en datos confiables. Para ello, es necesaria una integración adecuada de numerosas tareas y componentes (Kimball et al., 2008), y es en este sentido que la metodología de Kimball se vuelve una referencia importante en el campo del soporte a la toma de decisiones en las organizaciones. Esta metodología se compone de una serie de pilares importantes para el ciclo de vida de un proyecto data warehouse,

estos se ilustran en la Figura 5. Además de la metodología Kimball, que fue una base importante para el actual proyecto, también se tuvieron en cuenta otras buenas prácticas durante el desarrollo de la práctica empresarial.

La metodología Kimball se conforma de las siguientes etapas:

1. Planificación del proyecto.
2. Definición de requerimientos de negocio.
3. Diseño de la arquitectura técnica.
4. Selección e instalación de herramientas.
5. Especificación de aplicación analítica.
6. Construcción de aplicación analítica.
7. Modelamiento dimensional.
8. Diseño físico.
9. Diseño y desarrollo del ETL.
10. Despliegue.
11. Crecimiento.
12. Mantenimiento y administración del proyecto.

Para el desarrollo del proyecto se abordó la metodología teniendo en cuenta la Figura 5:

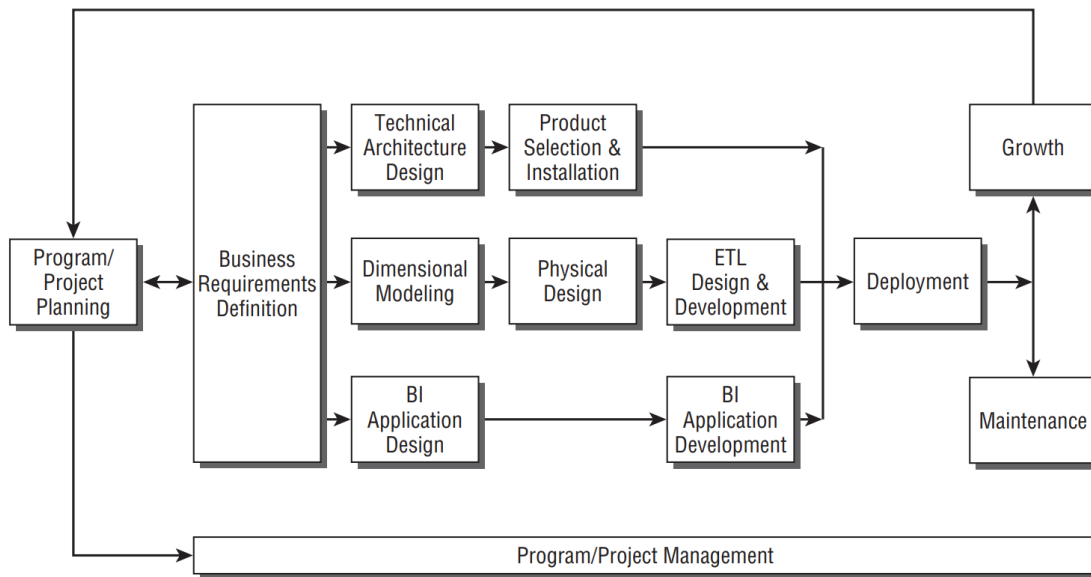
1. Planificación del proyecto.
2. Definición de requerimientos de negocio.
3. Diseño de la arquitectura técnica.
4. Selección e instalación de herramientas.
5. Diseño físico.
6. Diseño y desarrollo del ETL.

7. Pruebas.

Estas serán explicadas a continuación.

Figura 5

Diagrama del ciclo de vida de Kimball



Nota. La figura contiene las etapas del ciclo de vida para un data warehouse usando la metodología Kimball. Tomado de Kimball et al. (3rd ed.), p. 64. John Wiley & Sons, Inc.

3.2.1 Planificación del proyecto

Para asegurar el éxito del proyecto data warehouse, la planificación debe comenzar con la definición clara del alcance del proyecto. Es importante tener en cuenta los requerimientos del negocio a la hora de definir el alcance, lo que se puede observar en la Figura 5 Diagrama del ciclo de vida de Kimball como la dependencia bidireccional entre la planificación del proyecto y los requisitos del negocio.

Según Kimball et al. (2008), es recomendable desarrollar un plan de comunicación integral que incluya tanto a los grupos de negocios como a los de tecnología de la información, puesto que la comunicación efectiva es esencial tanto para administrar las expectativas de todas las partes interesadas, como para garantizar que se alcancen los objetivos del proyecto.

Siendo así, con un plan de comunicación claro y bien definido es posible asegurar que todos los miembros del equipo estén informados sobre el progreso del proyecto, los riesgos y los cambios en el alcance. Como resultado se tiene un plan de proyecto bien estructurado que identifica todas las tareas asociadas con el ciclo de vida de Kimball y las partes responsables, lo que será una parte fundamental para la gestión continua del proyecto de almacenamiento de datos (Kimball et al., 2008).

3.2.2 Gestión del Proyecto

La gestión del proyecto se debe centrar en monitorear el estado del proyecto, debe haber un seguimiento de problemas y una gestión del control de cambios para de esa manera respetar los límites del alcance. También esta fase incluye el desarrollo de planes de contingencia y la gestión de riesgos para minimizar los posibles impactos negativos en el proyecto (Kimball et al., 2008).

La gestión continua del proyecto se debe realizar a lo largo del proyecto como se observa en la Figura 5, Kimball también recalca que la gestión del proyecto es esencial para garantizar que se logren los objetivos del proyecto de almacenamiento de datos de manera oportuna y eficiente (Kimball et al., 2008).

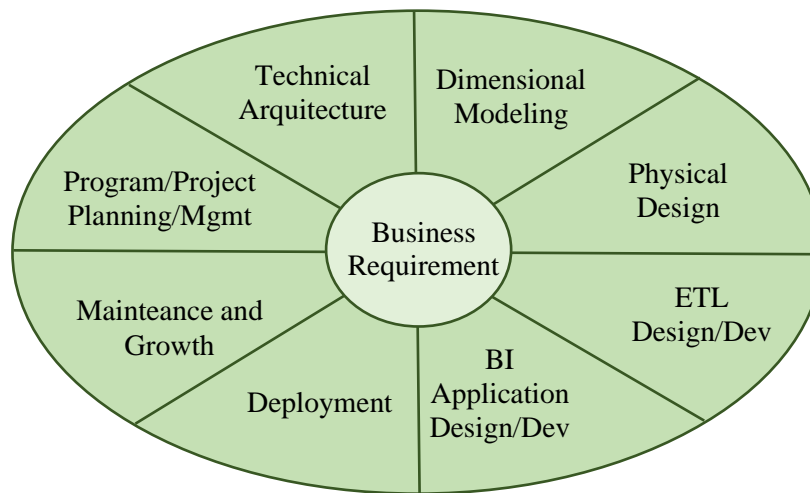
3.2.3 Definición de Requerimientos de Negocio

Kimball et al. (2008) menciona lo fundamental que es la comprensión de la definición de los requerimientos del negocio para el éxito en este tipo de proyectos, ya que una correcta definición de requerimientos influye en todas las actividades del ciclo de vida, desde establecer el

alcance hasta proporcionar soporte de implementación del proyecto. Además, considera que los requisitos son el centro del universo de un proyecto de data warehouse, tal como se muestra en la Figura 6.

Figura 6

Los requerimientos de negocio impactan virtualmente cada aspecto del proyecto



Nota. La figura muestra las etapas del ciclo de vida de Kimball, alrededor de los requerimientos de negocios, visualizando el impacto que tienen en cada aspecto del proyecto. Adaptado de Kimball et al. *The Data Warehouse Lifecycle Toolkit* (2nd ed.), p. 64. John Wiley & Sons, Inc.

Kimball et al. (2008) propone un enfoque de recolección de requerimientos que se diferencia del enfoque tradicional centrado en los datos. Según Kimball, es esencial recopilar los requerimientos de las personas que utilizarán y se beneficiarán del data warehouse, en lugar de simplemente enfocarse en las características técnicas del sistema.

Un ejemplo para comprender ambos tipos de requerimientos, es que un requerimiento tradicional podría establecerse de la siguiente manera: "el sistema debe ser capaz de procesar una

cantidad específica de transacciones en cierta cantidad de tiempo", se puede observar claramente que este tipo de requerimiento se centra en las características técnicas, en cambio un requerimiento orientado al negocio se establece de la siguiente manera: "el sistema debe ser capaz de proporcionar información útil en tiempo real sobre las ventas para mejorar la toma de decisiones empresariales". Este requerimiento se enfoca en cómo el proyecto puede apoyar las necesidades de los usuarios finales.

Una vez que se han definido los requerimientos de negocio, se abordan tres caminos simultáneamente para llevar a cabo el proyecto. El primer camino se enfoca en la tecnología y consta de dos etapas: diseño de la arquitectura técnica y selección de herramientas. El segundo camino se centra en los datos y se divide en tres etapas: diseño del modelo dimensional, diseño físico y diseño y desarrollo del sistema ETL. Por último, el tercer camino se enfoca en las aplicaciones de inteligencia de negocios e incluye el diseño y desarrollo de aplicaciones de inteligencia de negocios (Kimball et al., 2008).

3.2.4 Diseño de la Arquitectura Técnica

Kimball et al. (2008) menciona que el diseño de la arquitectura técnica para un proyecto de data warehouse requiere la integración de numerosas tecnologías y es fundamental establecer un marco y una visión arquitectónica general que permita su integración adecuada.

Él también indica que los requerimientos empresariales responden a la pregunta "¿qué necesitamos hacer?" y la arquitectura técnica responde a la pregunta "¿cómo lo haremos?". Esta última abarca el plan general para lo que se desea que haga el sistema de inteligencia de negocios y se utiliza para representar el flujo de datos desde los sistemas fuente hasta aquellos que toman las decisiones, además describe las transformaciones y el almacenamiento de los datos que recorren todo el sistema de inteligencia de negocios.

3.2.5 Selección e Instalación de Productos

Teniendo el plan de arquitectura técnica definido la siguiente etapa para un proyecto de inteligencia de negocios es realizar una “lista virtual de compras” de los componentes arquitectónicos específicos, como la plataforma de hardware, el sistema de gestión de bases de datos, la herramienta ETL y la herramienta para generación de informes. Una vez seleccionados los productos, se instalan y se prueban para garantizar una integración adecuada de extremo a extremo dentro del entorno del proyecto (Kimball et al., 2008).

3.2.6 Diseño Físico

Sobre el diseño físico Kimball et al. (2008) menciona diferentes actividades que podrían tenerse en cuenta para esta fase. Se resaltan las siguientes: el desarrollo de estándares de nombramiento de tablas, y elementos de ellas, y el desarrollo de un modelo de base de datos físico que supla los requerimientos identificados anteriormente.

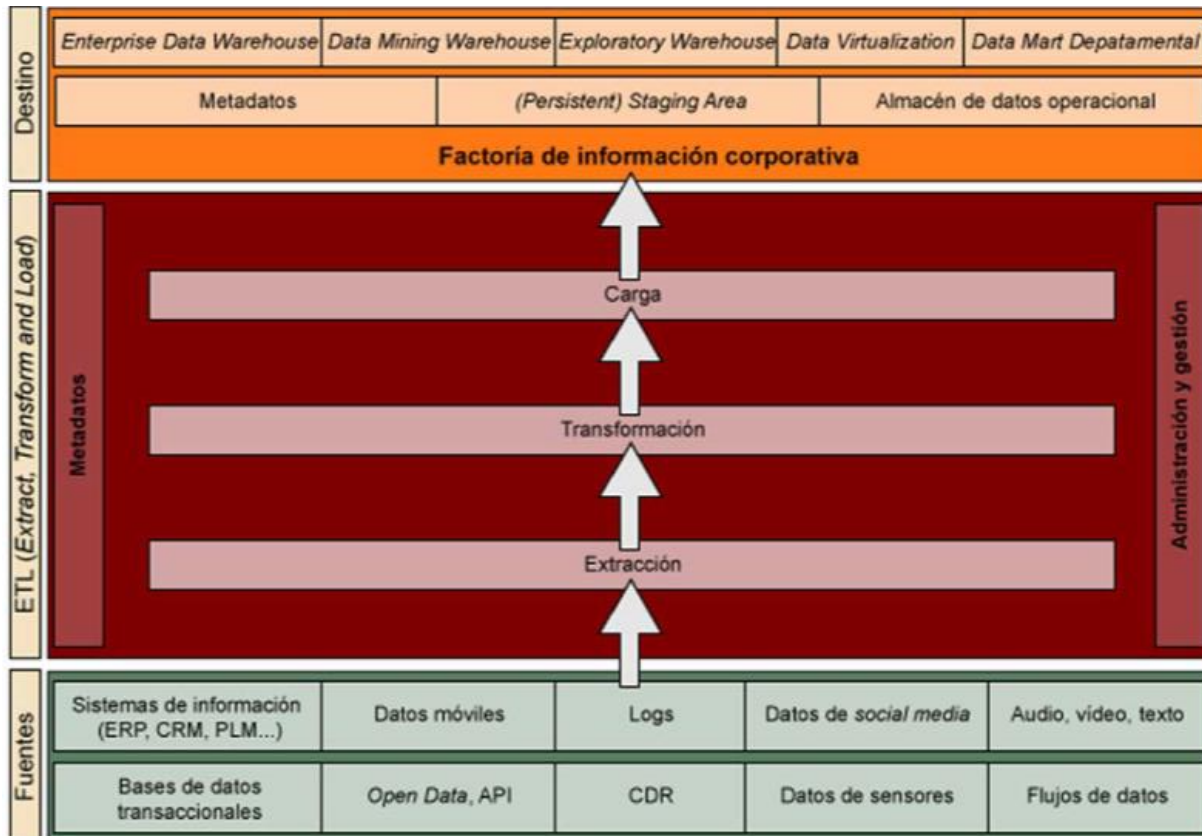
3.2.7 Diseño y Desarrollo del ETL

El sistema de extracción, transformación y carga es una parte crítica en el proceso de implementación de un entorno de inteligencia de negocios, puede consumir hasta el 70% del tiempo de construcción que requiere un proyecto de inteligencia de negocios (Kimball et al., 2008). Esto se debe a que existen diversas restricciones externas que influyen en el diseño del ETL como los son los requerimientos del negocio, los sistemas de origen de los datos además de las habilidades del personal disponible (Kimball et al., 2008).

Curto Díaz (2016) menciona que un sistema ETL: “permite extraer datos del entorno origen, transformarlos según nuestras necesidades de negocio para integración de datos y cargar estos datos en los entornos destino” (p. 83).

Figura 7

Funcionalidad ETL



Nota. La figura contiene los procesos de un ETL para un data warehouse con las distintas fuentes y destinos que puede tener un ETL. Tomado de Curto Díaz, 2016, p. 75.

Un sistema ETL es un conjunto de procesos y herramientas que se encargan de extraer datos de diversas fuentes, transformarlos y cargarlos en un destino.

El proceso de extracción implica la extracción de datos de distintas fuentes, así como se puede ver en la Figura 7. Esa obtención de datos se realiza de distintas fuentes como bases de datos transaccionales, sistemas de información, audio, video entre otros.

La transformación de los datos es la encargada de convertir los datos inconsistentes en un conjunto de datos compatibles y congruentes, para que puedan ser cargados en el data warehouse. Es de vital importancia armonizar un formato de estándares para que todos los datos estén integrados (Hefesto, 2010). Por su parte la carga implica la inserción de los datos que han sido previamente transformados al destino deseado, sea un data mart o un data warehouse.

4. Gestión del Proyecto y Definición de Requerimientos de Negocio.

La clave del éxito de un proyecto radica en una adecuada planificación. Es así como nos da a entender Crispieri, G. P. (2019), quien menciona que una adecuada planificación permitirá anticipar y corregir efectos de la formación de riesgos que afecten la culminación de los objetivos de un proyecto. Además, es necesario durante la planificación establecer objetivos claros, alcanzables e identificar riesgos asociados a cada etapa del proyecto, esto teniendo en cuenta la definición de los requerimientos del negocio.

4.1 Planificación del proyecto

Para esta etapa, se utilizó como punto de partida lo investigado inicialmente sobre la metodología Kimball. Se tomó del sitio web complementario del libro *The Data Warehouse Lifecycle Toolkit*; Kimball Group, una lista de tareas y responsabilidades propuestas que deben llevarse a cabo para la planificación de proyectos data warehouse. Esta lista incluye las actividades, responsabilidades de cada rol y su participación de acuerdo con las actividades de cada fase. Teniendo en cuenta la lista de roles que proponía, se escogieron los más relevantes y posteriormente se identificaron las tareas clave que debían ser completadas en cada etapa del ciclo de vida del proyecto. Como resultado se elaboró el cronograma del proyecto, lo cual facilitó la organización y planificación de las actividades necesarias para su implementación, y se establecieron los objetivos específicos de este.

4.2 Gestión del Proyecto

Establecer una estrategia de reuniones y escoger una herramienta para la supervisión del progreso del proyecto también es esencial para la gestión de este, debido a que permite mantener un seguimiento del progreso y los riesgos durante el desarrollo, mantener cada una de las partes informadas, asegurarse de que las actividades del proyecto estén alineadas con los objetivos de este y finalmente asegurar los entregables en los plazos establecidos.

Para la gestión y planeación se utilizó Microsoft Teams, una aplicación de colaboración creada para mantener los equipos de trabajo informados, organizados y conectados en un mismo lugar (Introducción a Microsoft Teams, s. f.). Esta aplicación, por medio de sus pestañas Planner permite crear un plan, añadir tareas y añadir una descripción, hacer listas de comprobación, añadir fechas de inicio, de entrega y demás. Haciendo uso de esta herramienta, se migró el cronograma existente, permitiendo que cada miembro del equipo pudiera acceder y dar seguimiento al proyecto.

Para crear una estrategia de comunicación, se decidió hacer reuniones diarias para mantener al equipo de trabajo informado sobre el progreso del proyecto, para de esta manera identificar posibles obstáculos y tomar decisiones rápidamente. Por ello, cada día antes de mediodía se hacía una reunión de 15 minutos para hablar de qué tareas se habían hecho el día anterior, cuáles se harían durante el día, y qué obstáculos estaban impidiendo que las tareas se llevaran a cabo. El equipo de trabajo estaba conformado por Paula Castro y Paula Hernández, autoras del presente informe; Victor Rangel, ingeniero tutor de la ESSA y Román Suárez Project Manager de iniciativas Soluciones Digitales en la ESSA.

Adicional a las reuniones diarias, se decidió programar una reunión de sincronización semanal de una hora con el equipo de trabajo para revisar el progreso del proyecto con respecto al

cronograma que se había establecido en un principio, acomodar fechas y tareas según se iba avanzando en cada semana, discutir problemas y tomar medidas tanto de prevención, para evitar retrasos, como correctivas, para ajustar el plan según el caso o crear un plan de contingencia para abordar los problemas presentados y minimizar el impacto que tuviesen estos en las fechas de entrega y calidad del proyecto mismo.

Luego de haber construido un cronograma y haber establecido un plan de seguimiento al progreso del proyecto, se realizó una entrevista con el área usuaria con el objetivo de identificar sus necesidades y determinar qué se esperaba lograr con el proyecto.

Durante la reunión, se presentó al principal usuario que se beneficiaría del proyecto, lo que permitió establecer acuerdos en cuanto a los horarios de encuentro semanal, horarios para llevar a cabo las entrevistas de recolección de requerimientos y centradas en datos. Las reuniones semanales se realizaron con el objetivo de mantener actualizada al área usuaria sobre los avances del proyecto.

Estas reuniones permitieron mostrar los progresos realizados y recibir retroalimentación del área usuaria, lo que contribuyó a asegurar que el proyecto estuviera alineado con sus necesidades y expectativas. La retroalimentación recibida durante estas reuniones fue valiosa para realizar ajustes y mejoras en el proyecto, asegurando así su adecuada planificación y gestión.

Una vez establecida la planificación y gestión del proyecto, y después de haber recopilado información general con el área usuaria, se procedió a la identificación de los procesos críticos del proyecto. En este sentido, se realizaron diversas actividades que se explicarán a continuación.

4.3 Identificación de Procesos Críticos

Para identificar los procesos críticos en el proyecto se siguió un enfoque basado en la metodología propuesta por Kimball en su libro "The Data Warehouse Toolkit". De este enfoque,

se seleccionó la realización de entrevistas como parte del proceso de recopilación de requisitos, con el objetivo de identificar los procesos clave, para ello se realizaron dos conjuntos de entrevistas, cada uno con un enfoque específico. Para ello se realizaron las actividades siguiendo el flujo mostrado en la Figura 8. Se llevó a cabo una etapa de priorización de esos procesos para definir los requerimientos del proyecto y establecer su alcance.

El proyecto del desarrollo del data mart se realizó en el área de Gestión Operativa y el área Subgerencia de Conexiones. Siguiendo la metodología de Kimball se llevaron a cabo las etapas descritas en la metodología escogida. Para simplificar la escritura del informe, se estableció el concepto de "Área Usuaría", que abarca tanto a la Subgerencia de Conexiones como a Gestión Operativa, esto debido a que a pesar de que en un principio se quería realizar el proyecto solo para el área de Gestión Operativa, pero posteriormente se decidió modificar su alcance. Esta simplificación permite referir a ambas áreas de manera conjunta cuando las actividades son similares. Sin embargo, cuando se hace necesario mencionar algún detalle específico de cada una, se utiliza el nombre correspondiente.

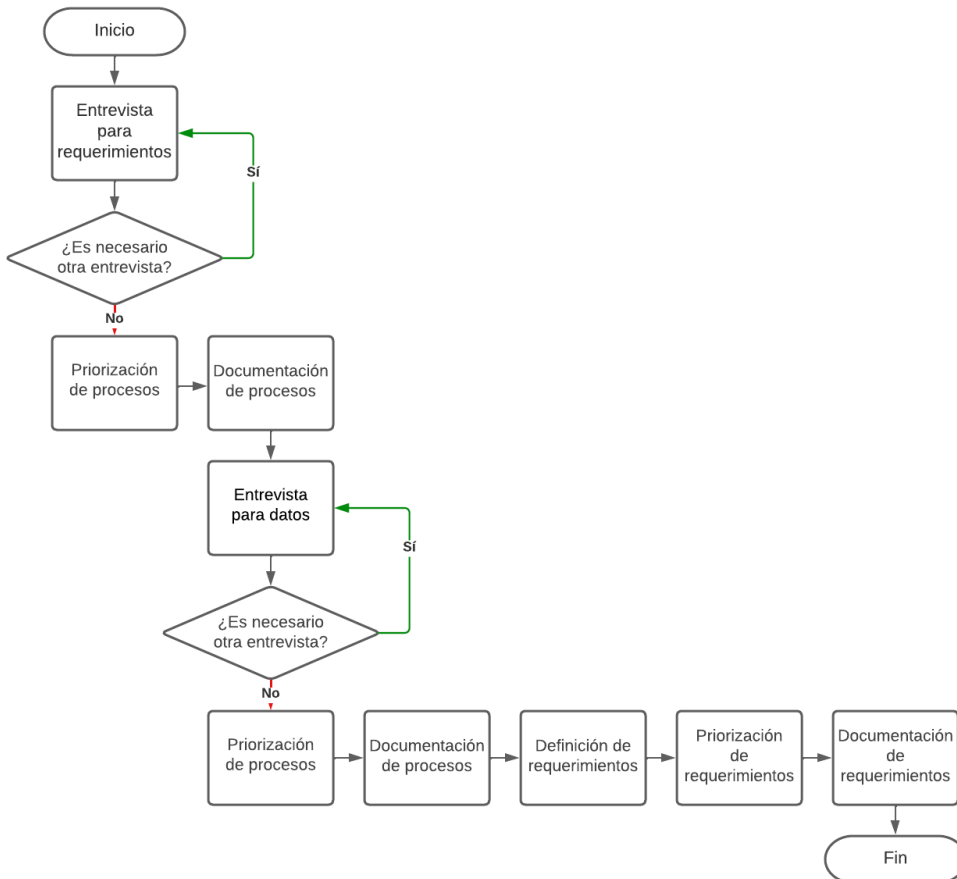
El primer conjunto de entrevistas, entrevistas para requerimientos, se centró en los requerimientos del área usuaria. Se llevaron a cabo dos entrevistas para cada área con el fin de recopilar información sobre los procesos actuales, los desafíos que enfrenta el área usuaria y los objetivos que se esperan alcanzar.

El segundo conjunto de entrevistas, entrevistas centradas en datos, se enfocó en los datos y aspectos técnicos. Se llevaron a cabo dos entrevistas para cada área con el propósito de obtener información específica sobre los datos necesarios para el proyecto, las fuentes de datos disponibles y demás requisitos técnicos y consideraciones. Ver

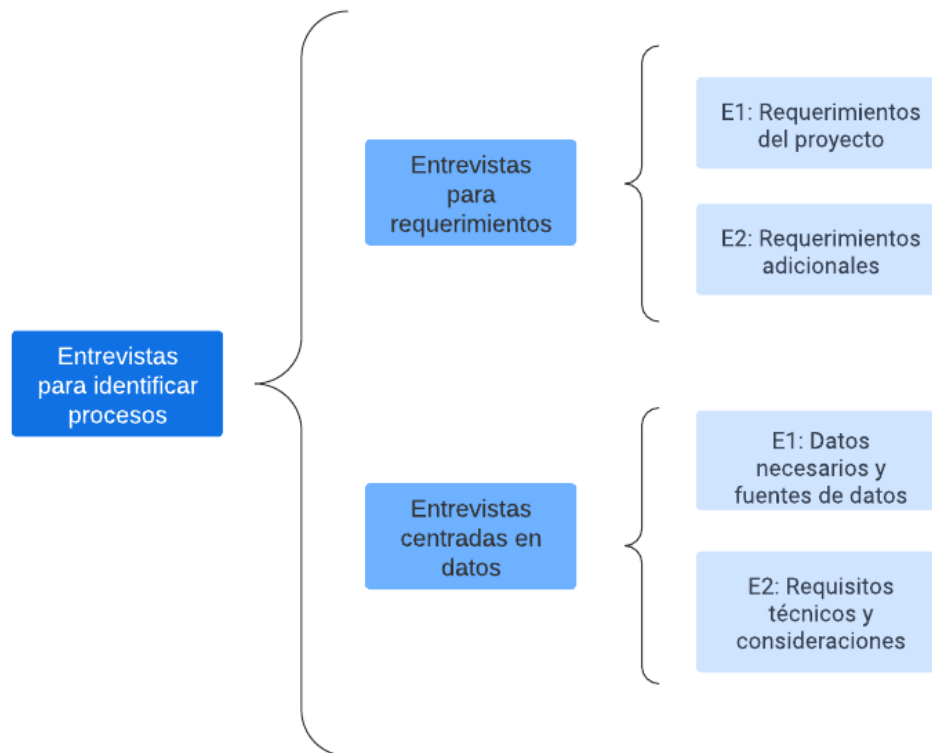
Figura 9.

Figura 8

Flujo de identificación de procesos críticos en el proyecto.



Nota. La figura contiene el flujo de actividades realizado en el proyecto para poder recolectar los requerimientos y definir el alcance de este. Fuente elaborado por los autores.

Figura 9**Diagrama jerárquico de entrevistas.**

Nota. En la figura se presenta un diagrama jerárquico de los dos conjuntos de entrevistas: entrevistas de requerimientos y entrevistas de datos. Fuente elaborado por los autores.

4.3.1 Entrevistas para Requerimientos

No es de extrañar que los requerimientos sean un elemento crucial en el proceso de desarrollo de proyectos. Actúan como un puente entre el equipo de desarrollo y el cliente, la comprensión mutua entre ambas partes es fundamental para el éxito del proyecto. En este sentido, Kimball et al. (2008) señala que los usuarios prácticamente hablan un idioma desconocido para el típico profesional de tecnologías de la información, por esta razón, él menciona que es importante no limitarse a preguntar simplemente qué datos desean los usuarios en el almacén de datos, en su

lugar, se debe iniciar el diálogo hablando sobre sus roles, objetivos y desafíos laborales, para luego adentrarse en los detalles técnicos. Las entrevistas para recopilar los requerimientos en el área usuaria se llevaron a cabo con el propósito de obtener información acerca de qué se hacía en el área, qué necesidades tienen y cómo las querían solucionar. En total, se realizaron dos entrevistas por área en las que se definieron los requerimientos del proyecto.

Durante la primera entrevista, se organizó una reunión con el equipo de cada área y el responsable de Tecnologías de información para presentarles el proyecto y establecer una comunicación inicial. Durante esta reunión, se discutió la iniciativa del proyecto y se identificó el usuario clave que tenía necesidades específicas.

El usuario clave para el área de Gestión Operativa tiene como rol: Administrador CEE (Calidad de Energía Eléctrica), y el usuario clave para el área de Subgerencia de Conexiones tiene como rol: Analista RCE (Reducción y Control de Energía).

El rol del administrador de calidad de energía eléctrica tiene diversas responsabilidades. En primer lugar, se encarga de monitorear y analizar la calidad de la energía eléctrica, debe asegurarse de cumplir con las normas y regulaciones establecidas, con la gestión de eventos y fallas en el sistema eléctrico. Todo ello implica registrar, analizar y tomar acciones para corregir y prevenir incidencias, asegurando un suministro eléctrico confiable en cumplimiento con la resolución CREG 015. El rol de administrador debe elaborar informes y reportes periódicos en tableros de inteligencia de negocios que reflejen la calidad del suministro eléctrico y el cumplimiento de las normas.

El rol de analista RCE (Reducción y control de Energía), tiene responsabilidades alrededor de la recuperación de la energía, y la elaboración de reportes para poder establecer indicadores de pérdidas, y así permitir la toma de decisiones informadas.

En esta primera entrevista se pudo dar a conocer el papel del usuario clave para el proyecto y la importancia de las tareas que desarrolla en su rol. Todo esto sentó las bases para las siguientes interacciones.

En la segunda entrevista, se planteó un enfoque más detallado y se formularon preguntas específicas para comprender a fondo las necesidades y los objetivos de los usuarios. Se exploraron aspectos como los motivos para realizar cambios en las tareas actuales y las actividades que se realizaban con los procesos existentes. Esta entrevista permitió obtener una visión más completa de las actividades que llevaba a cabo el usuario clave.

Estas entrevistas fueron fundamentales para recopilar información precisa y detallada sobre los requerimientos de las áreas usuarias. Con base en esta información, se pudo desarrollar un enfoque más sólido y orientado a las necesidades del usuario para implementar las soluciones adecuadas.

A medida que se adquirió un mayor conocimiento de la situación, la iniciativa y la problemática específica, se tomó la decisión de preparar un cuestionario para llevar a cabo la entrevista centrada en datos. Este cuestionario se diseñó con el objetivo de obtener información precisa y detallada sobre los procesos de extracción, transformación y carga de datos que se llevaban a cabo de manera manual antes de que fueran incorporados al tablero de inteligencia de negocios.

4.3.2 Entrevistas Centradas en Datos

Las entrevistas enfocadas en datos, o como las menciona Kimball et al. (2008), "entrevistas de auditoría de datos" o "data audit", ayudan a comprender si los datos están disponibles para

respaldar las solicitudes de la empresa, si son completos y confiables. Además, estas entrevistas nos permiten descubrir problemas ocultos en los datos.

Para llevar a cabo estas entrevistas, es necesario elaborar un cuestionario antes de comenzar la entrevista. Los cuestionarios para las sesiones de las entrevistas centradas en datos son diferentes de los cuestionarios de las entrevistas de requerimientos.

Para las áreas usuarias, se realizó un cuestionario antes de la entrevista centrada en datos con el objetivo de conocer el manejo que se les da a los datos. Como ya contaban con un tablero de inteligencia de negocios, las preguntas fueron diseñadas para mejorar los procesos de llevar los datos al tablero. Por lo tanto, el cuestionario se diseñó de tal manera que se enfocara en entender los procesos de extracción, transformación y carga que se realizaban a los datos antes de llevarlos al tablero de inteligencia de negocios.

Algunas preguntas que se realizaron fueron: ¿cuáles son los orígenes de los datos?, ¿qué transformaciones se hacen a esos datos?, ¿cuánto tiempo se tarda en ese proceso de extracción?, ¿qué herramientas se utilizan para la extracción y transformación de esos datos?, ¿para qué se utilizan esos datos?, ¿qué decisiones se toman con ellos? y ¿hacia dónde se cargan los datos?

4.4 Priorización de Requerimientos y Establecimiento del Alcance del Proyecto

Una vez hechas las entrevistas se procede a hacer la documentación de estas y de los procesos identificados. Es fundamental hacer esto debido a que la documentación se convierte en la enciclopedia del material de referencia para el proyecto. Si no se documenta desde el principio, se perderá información relevante si miembros del equipo abandonan el proyecto sin documentar sus hallazgos. (Kimball et al., 2008). Así que es a partir de esa documentación de entrevistas y procesos que se logran desarrollar los diagramas de notación de procesos (ver Figura 10,

Figura 11 y

Figura 12) donde se evidencia el proceso que hacían los usuarios antes de llevar a cabo el proyecto, además de que también permitió identificar oportunidades de mejora en los procesos de extracción, transformación y carga de datos, garantizando de esa manera la correcta puesta a disposición de esos datos a los usuarios.

Posterior a esa documentación, se procede a realizar la definición de requerimientos del proyecto, así como la priorización de ellos.

En el área de Gestión Operativa se identificaron tres procesos realizados por el rol de Administrador CEE. Estos procesos tenían actividades comunes para llevar los datos al tablero de inteligencia de negocios, ver Figura 10. Cada proceso se relacionaba con eventos del sistema eléctrico, transformadores por vereda y consignaciones. Aunque los pasos eran similares, cada proceso utilizaba consultas SQL diferentes debido a la naturaleza de los datos. Se seleccionaron estos tres procesos sin necesidad de hacer una priorización, ya que todos ellos contaban con tableros de inteligencia de negocios implementados.

En el área de Subgerencia de Conexiones se identificaron dos procesos con actividades diferentes antes de llevarlos al tablero de inteligencia de negocios. En esta área se realizó una priorización de procesos, ya que el tiempo de entrega del proyecto se vería afectado además de que el segundo proceso (ver

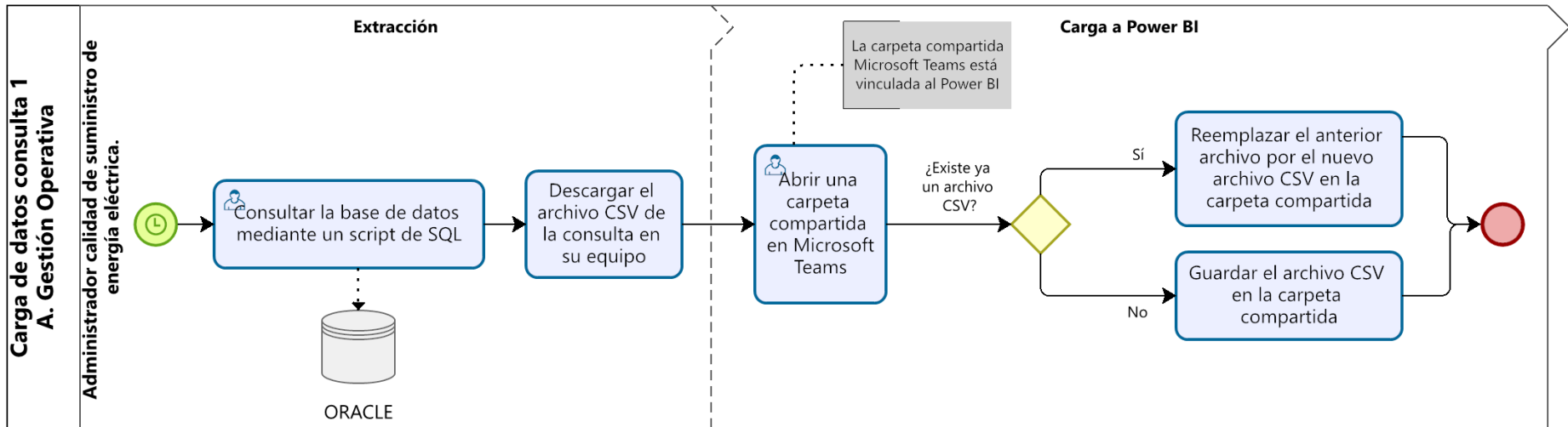
Figura 12) no contaba con un tablero de BI existente. Por lo tanto, se priorizó únicamente el primero (ver

Figura 11), que era de gran importancia para el área al tratar sobre la recuperación de energía, un aspecto crucial para la organización.

Es importante destacar que este análisis de procesos también se incluyó en la documentación del desarrollo como parte de los entregables para el área usuaria.

Figura 10

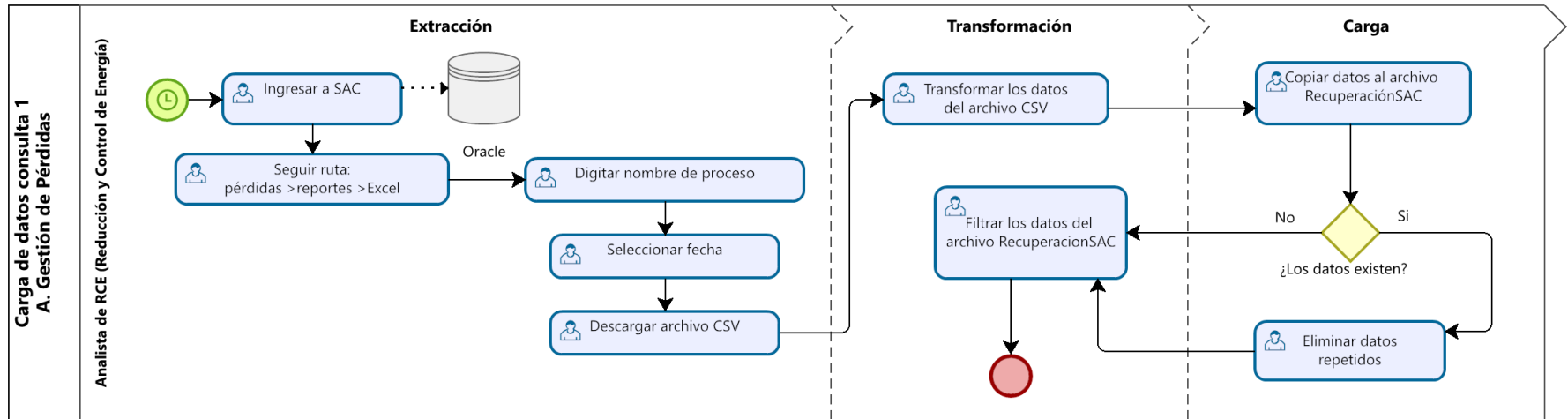
Carga de datos a Power BI para el área de Gestión Operativa.



Nota. La figura contiene los procesos en el área de Gestión Operativa para la carga de datos a la herramienta de inteligencia de negocios Power BI. Fuente elaborado por los autores.

Figura 11

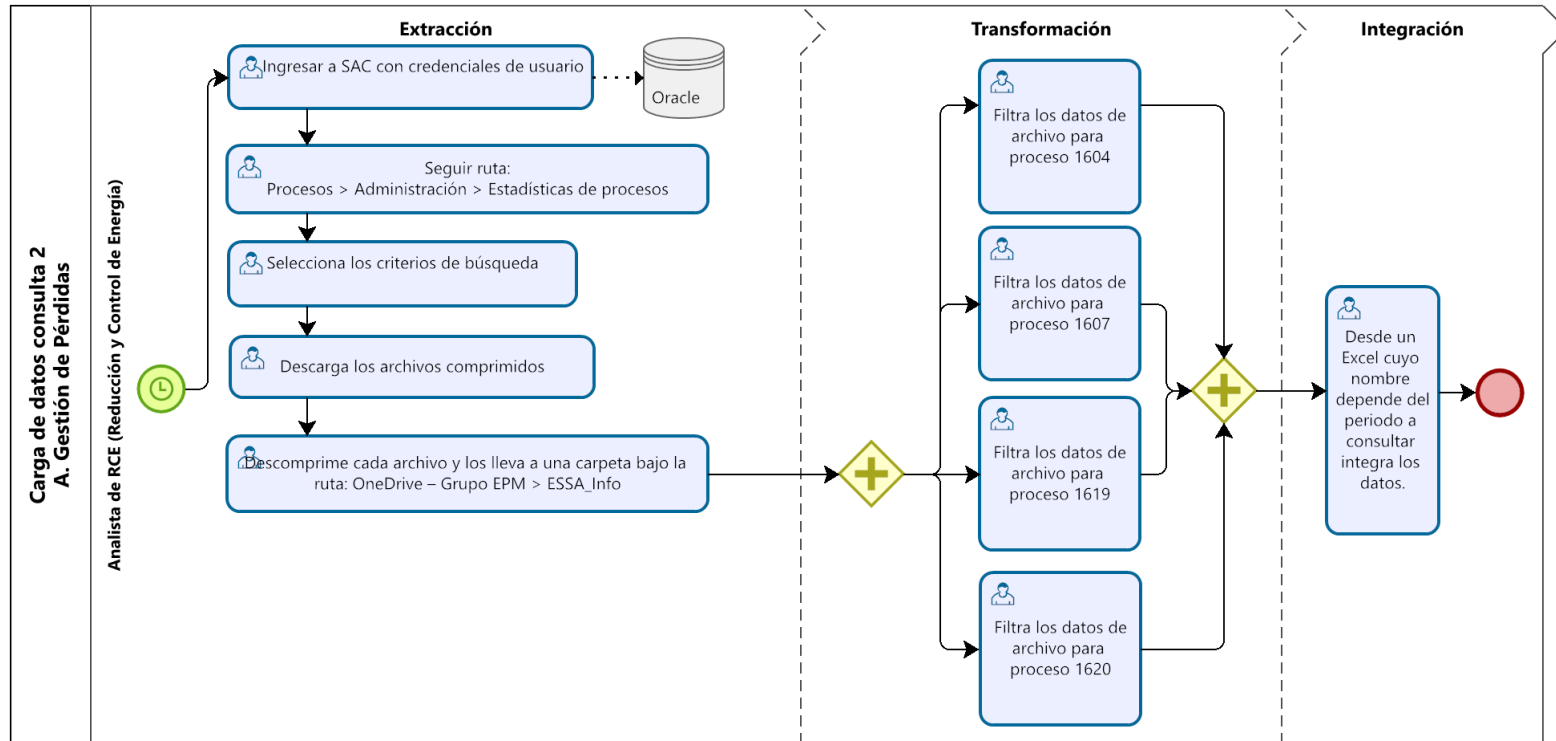
Carga de datos a Power BI para el área de Subgerencia de Conexiones.



Nota. La figura contiene los procesos en el área de Subgerencia de Conexiones para la carga de datos a la herramienta de inteligencia de negocios Power BI. Fuente elaborado por los autores.

Figura 12

Carga de datos a Power BI para el área de Subgerencia de Conexiones.



Nota. La figura contiene los procesos en el área de Subgerencia de Conexiones para la carga de datos a la herramienta de inteligencia de negocios Power BI. Fuente elaborado por los autores.

Gracias a esa documentación se pudo establecer los requerimientos del proyecto por área. El documento de requisitos contiene una descripción detallada de los procesos de minería de datos que realiza el área, historias de usuario, así como la priorización de los requisitos. Esta priorización ayudó a definir el alcance del proyecto y a establecer un tiempo de entrega acorde a las necesidades de la empresa.

5. Arquitectura Técnica y Herramientas

La definición de la arquitectura técnica y la selección de herramientas de un proyecto data mart permite establecer una estructura y guía clara al iniciar un proyecto. Como esto involucra que, a partir de la identificación de requerimientos se definan los componentes, las tecnologías y demás, esto se vio sujeto a una revisión y análisis de las tecnologías y las licencias disponibles dentro de los recursos de la empresa, además de la evaluación de si era necesario adquirir nuevas tecnologías o licencias con el fin de cubrir las necesidades del proyecto, proceso durante el cual el profesional encargado de infraestructura estuvo asignado.

Como resultado, se propuso la arquitectura técnica para el data mart del área de Gestión Operativa y Gestión de Pérdidas. Esta está conformada por dos capas: 1. Back Room. 2. Front Room.

5.1 Back Room

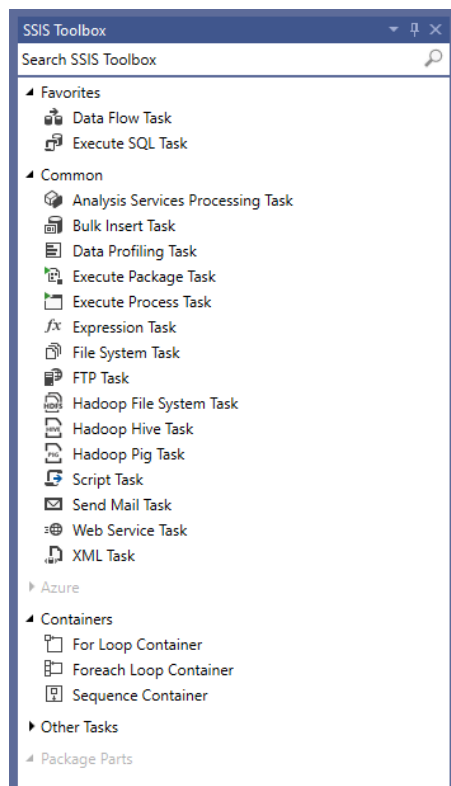
La capa back room, según es descrita por Kimball et al. (2008), es donde el proceso ETL ocurre, lo cual implica que el proceso de llevar datos desde un punto A a un punto B se haga con las transformaciones y en el tiempo adecuado.

Sobre la selección de herramientas: previo al inicio del proyecto Data Mart en la empresa, se había hecho una capacitación a los empleados sobre el uso de la herramienta SQL Server Integration Services (en adelante 'SSIS' por sus siglas en inglés) como extensión de Visual Studio, en la que

se mostraba los beneficios que podría dar ésta a la organización y se les enseñaba a los trabajadores cómo usarla. De acuerdo con el sitio web oficial de Microsoft (2023), SQL Server Integration Services (SSIS) proporciona una amplia gama de capacidades que incluyen, la extracción y transformación de datos desde diversas fuentes hacia uno o más destinos y ofrece un conjunto de herramientas gráficas que facilita la creación de paquetes de manera visual, además de un conjunto de tareas y transformaciones predefinidas que facilitan la manipulación y transformación de datos (ver Figura 13).

Figura 13

Caja de herramientas de SQL Server Integration Services



Nota. La figura contiene la lista de componentes y herramientas de SSIS para realizar una integración de datos. Fuente interfaz Visual Studio.

Y aunque luego de esa capacitación no hay conocimiento de algún proyecto que hubiese usado SSIS, fue en base a ella que se decide que la herramienta de integración a usar que garantizaría la compatibilidad y la integración necesaria para el proyecto fuese SSIS como extensión de Visual Studio 2022.

Posterior a la selección de las herramientas, se comenzó el proceso de autocapacitación de ellas. Esta se hizo con el fin de adquirir las habilidades y conocimientos necesarios para la integración de datos que se quería hacer haciendo uso de SSIS y SQL Server.

Para ello se utilizaron diferentes métodos para la autocapacitación, entre ellos el uso de diferentes recursos en línea, la visualización de videos académicos en plataformas como YouTube, la realización de proyectos académicos a menor escala para la práctica, lectura de la documentación oficial de Microsoft y la creación de resúmenes, cuestionarios y demás materiales para socializar los conocimientos o dudas con el equipo del proyecto. Todo esto con el enfoque de familiarizarse con la herramienta, su interfaz, componentes, funcionalidad, terminología y la aplicación de esos conceptos fundamentales adquiridos en el ambiente práctico. Dado que un incidente cibernético en la empresa obligó a los trabajadores a laborar en modalidad remota, se aprovechó ese tiempo para la autocapacitación dedicando horas específicas cada día para el estudio y la práctica, y se utilizaron recursos propios como los computadores personales, las herramientas en su versión académica y otros como la base de datos Adventure Works, para hacer que la autocapacitación fuese lo más cercana al entorno empresarial, eso permitió que las pruebas y el trabajo una vez se volvió a la empresa se diese sin dificultades significativas.

El servidor puesto a disposición para el proyecto fue el ESSA-TEST18 cuyas especificaciones se ven en la Tabla 2. Este servidor se utilizó para poder desarrollar los ETL haciendo uso de las herramientas Visual Studio y su extensión Integration Services.

Ahora bien, sobre los gestores de bases de datos, se usaron:

- Oracle SQL Developer 22.2.1.234.
- Microsoft SQL Server Management Studio 18.

Adicional, se usó el servidor ESSA-PS06 para alojar la base de datos de los data marts. Para ello, el profesional encargado de infraestructura creó la base de datos y el usuario al que accederíamos desde el gestor en el ESSA-TEST18. Eso significa que el grupo de trabajo no tuvo que manipular directamente aquel servidor.

Tabla 2

Especificaciones servidor ESSA-TEST18

Especificaciones del hardware	
Procesador	Intel Xeon CPU E5-2690 v2, 3.00 GHz
RAM	16 GB
Almacenamiento	130 GB
Especificaciones del software	
Sistema operativo	Windows Server 2019 Standard
Servidor de base de datos	SQL Server 2016

Posterior a esa selección se procede a hacer la instalación de las herramientas escogidas en el servidor ESSA-TEST18, el cual sería el servidor en el que se haría el desarrollo de los paquetes SSIS de cada ETL.

Adicionalmente, fue necesario con el profesional de infraestructura crear una base de datos y configurar un usuario administrador para de esa manera tener acceso desde el servidor ESSA-TEST18 a los data marts.

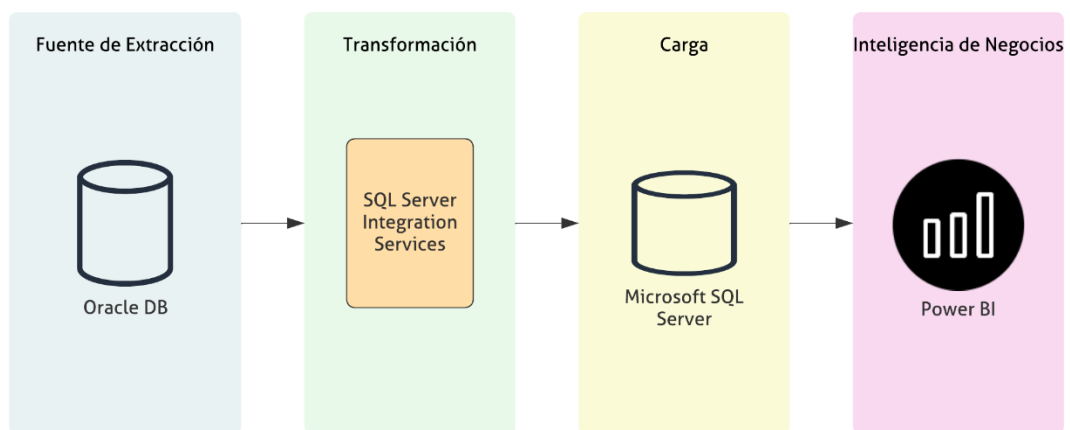
5.2 Front Room

El front room es la cara pública, lo que los usuarios del negocio verán y usarán cada día para encontrar respuestas y soluciones a sus requerimientos (Kimball et al., 2008). En el caso del actual proyecto, a pesar de que se definió desde el comienzo que el alcance del proyecto solo incluiría el backroom, es de conocimiento al equipo que los datos de los data marts serán mostrados en la herramienta Power BI.

Así que la arquitectura final del proyecto se planteó finalmente como se ve en la Figura 14. En la que se muestra que la fuente de extracción de datos sería la base de datos Oracle, la herramienta para hacer los procesos de extracción, transformación y carga sería SQL Server Integration Services como extensión de Visual Studio 2022, la carga de datos en una base de datos Microsoft SQL Server y finalmente de cara al cliente la herramienta de inteligencia de negocios que consumiría los datos es Power BI.

Figura 14

Arquitectura técnica final del proyecto



Nota. La figura contiene la arquitectura del proyecto data mart, en donde se puede evidenciar el sistema ETL, cada proceso con su herramienta respectiva. Fuente elaborado por los autores.

6. Diseño

El diseño en general es de vital importancia en un proyecto data mart. Este permite que se realice una correcta traducción del modelo lógico, que describe la estructura y las relaciones entre los datos, a una base de datos física. Asimismo, asegura que el proceso ETL del data mart garantice la calidad e integridad de los datos que haya en él.

El diseño físico de la base de datos implica diferentes actividades, como la definición de estándares, convenciones de nomenclatura, toma de decisiones sobre el uso de valores nulos, diseño de tablas, columnas, y la debida documentación y revisión final de todo el diseño.

Por otro lado, el diseño del subsistema ETL involucra los procesos de extracción, transformación y carga de datos de acuerdo con las historias de usuario recopiladas durante la fase de recolección de requisitos.

6.1 Diseño físico de la base de datos

6.1.1 Definición de estándares para objetos físicos

El desarrollo y la definición de estándares garantiza la consistencia y calidad en todo el sistema, así cada componente está construido de manera coherente, lo que facilita la navegación del sistema para los usuarios de negocios y desarrolladores. En caso de que el departamento de Tecnologías de la información (en adelante 'TI') ya tenga estándares establecidos, lo más recomendable es aprovechar las prácticas existentes, esto a menos que sea necesario modificar aquellos estándares ya definidos para adaptarlos a las necesidades específicas del proyecto (Kimball et al., 2008).

En ese sentido, se hizo una revisión de las convenciones y estándares existentes en la empresa, para ello se consultó al grupo de TI para determinar si ya contaban con las convenciones establecidas que pudieran ser aplicadas al proyecto. Tras evaluarlas, se tomó la decisión de

utilizarlas tal cual como estaban establecidas, permitiendo así la coherencia y consistencia en todos los sistemas. A continuación, se detallarán las convenciones usadas en el proyecto.

En cuanto a las convenciones usadas para las tablas, se adoptó el uso de siglas que hacen referencia a los módulos o proyectos relacionados, seguidas de un guion bajo. Permitiendo una fácil identificación y organización de tablas.

Con respecto a los esquemas, se siguió una convención similar a la de las tablas. Y finalmente, en cuanto a los atributos, se estableció el uso de mayúsculas y guiones bajos en sus nombres. Se hizo hincapié en que los nombres de los atributos sean descriptivos y relevantes para su función dentro del data mart.

6.1.2 Diseño de base de datos física

De acuerdo con Kimball et al. (2008), el proceso de diseño de tablas es bastante sencillo y de bajo riesgo en el entorno de desarrollo, puesto que los nombres físicos de las tablas y las columnas se derivan casi exactamente del modelo lógico. Es por ello que antes del mismo diseño físico, es necesario tener claro el modelo lógico de los datos que se manejarían en el caso de cada área. Es a partir de las entrevistas y la información recolectada en el punto 4.4 Priorización de Requerimientos y Establecimiento del Alcance del Proyecto, que se hace el gráfico de cada modelo lógico de datos para cada área (Figura 15, Figura 16, Figura 17 y Figura 18).

En el caso de ambas áreas de la empresa atendidas en el proyecto, Gestión Operativa y Subgerencia de Conexiones, cada usuario final entrevistado ya tenía claro qué consulta de datos se realizaría, es decir, cada uno de ellos ya realizaba consultas de manera directa a la base de datos para luego elaborar sus reportes de inteligencia de negocios. Debido a que este proyecto era pionero y representaba un desafío en cuanto a la exploración de herramientas y su integración con la infraestructura de la empresa, se optó por la simplicidad de tener una única tabla en cada data

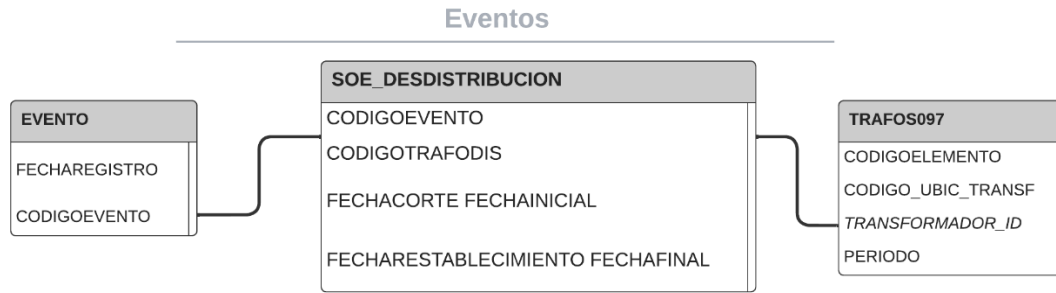
mart para recibir los datos, en otras palabras, habría una base de datos con esquemas específicos para cada área que servirían como data mart. Aunque se elaboró un modelo lógico que mostraba la ubicación de los datos en la base de datos origen, y se realizaron bocetos del modelado dimensional en el data mart, finalmente se optó por la integración de los datos en tablas individuales, cada una ubicada en su respectivo esquema correspondiente al área de la empresa. De esta manera, se aseguraba que cada consulta de cada área tuviera su propia tabla en el data mart, lo que facilitaría las consultas por parte de los usuarios finales hacia el data mart, y se determinó que se crearía la tabla destino con los mismos tipos de datos desde el origen.

6.1.2.1 Diseño lógico a físico para área de Gestión Operativa. En Gestión Operativa, área en la que se identificaron tres diferentes consultas, se hizo la respectiva transformación del diseño lógico de la base de datos a el diseño físico específico para cada una de las consultas.

6.1.2.1.1 Consulta Eventos. Primero, para la consulta de Eventos, se hizo una revisión de los tipos de datos en la base de datos origen y la relación entre ellos, dando como resultado la Figura 15. Siguiendo la decisión mencionada en el punto anterior, se creó una tabla individual que guardaría el resultado de esa consulta, conservando los mismos tipos de datos de la base de datos origen. Esto dio como resultado los atributos descritos en la Tabla 3, los cuales pertenecen a la tabla de nombre AGO_consulta_eventos, del esquema AGO_esquema, y de la base de datos ESSA.BD.DWH.

Figura 15

Modelo lógico para la consulta Eventos del área de Gestión Operativa



Nota. Fuente elaborado por los autores.

Tabla 3

Lista de atributos de la tabla AGO_consulta_eventos

Nombre del atributo	Tipo de dato	Longitud
CODIGO_ELEMENTO	varchar	varchar(50)
CODIGO_EVENTO	float	float
CREG	varchar	varchar(12)
FECHA_INICIAL	datetime	datetime
FECHA_FINAL	datetime	datetime
CAUSA_LAC	varchar	varchar(4000)
FECHA_REGISTRO	datetime	datetime
DURACION	float	float
PERIODO	float	float

6.1.2.1.2 Consulta Consignaciones. Luego de la revisión de los tipos de datos en la base de datos origen y la relación entre ellos (ver Figura 16), se creó una tabla individual que guardaría el resultado de esa consulta, conservando los mismos tipos de datos de la base de datos origen. Esto dio como resultado los atributos descritos en la Tabla 4, los cuales pertenecen a la tabla de nombre AGO_consulta_consignaciones, del esquema AGO_esquema, y de la base de datos ESSA.BD.DWH.

Figura 16

Modelo lógico para la consulta Consignaciones del área de Gestión Operativa

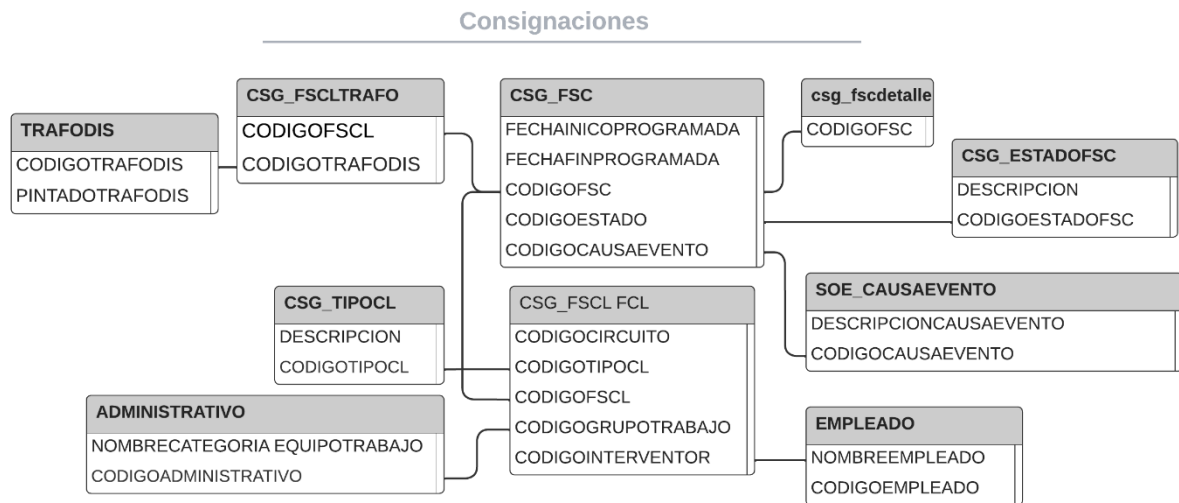


Tabla 4

Lista de atributos de la tabla AGO_consulta_consignaciones

Nombre del atributo	Tipo de dato	Longitud
CODIGO_FSCL	float	float
TIPO_CL	varchar	varchar(50)
ESTADO_CSG	varchar	varchar(50)
EQUIPO_TRABAJO	varchar	varchar(80)
FECHA_INICIO_PROGRAMADA	datetime	datetime
FECHA_FIN_PROGRAMADA	datetime	datetime
CODIGO_CIRCUITO	varchar	varchar(7)
CODIGO_TRAFODIS	float	float
PINTADO_TRAFODIS	varchar	varchar(7)
PROFESIONAL_RESPONSABLE	varchar	varchar(50)
TRABAJOS_REALIZAR	varchar	varchar(4000)
CAUSA_EVENTO	varchar	varchar(80)

6.1.2.1.3 Consulta Trafos por Vereda. Finalmente, también se hizo la revisión y análisis de las relaciones entre los datos traídos en esa consulta. Ver Figura 17. De allí la tabla origen diseñada AGO_consulta_trafosporvereda, fue ubicada en el esquema AGO_esquema y cuyos atributos se especifican en la Tabla 5.

Figura 17

Modelo lógico para la consulta Trafos por Vereda del área de Gestión Operativa

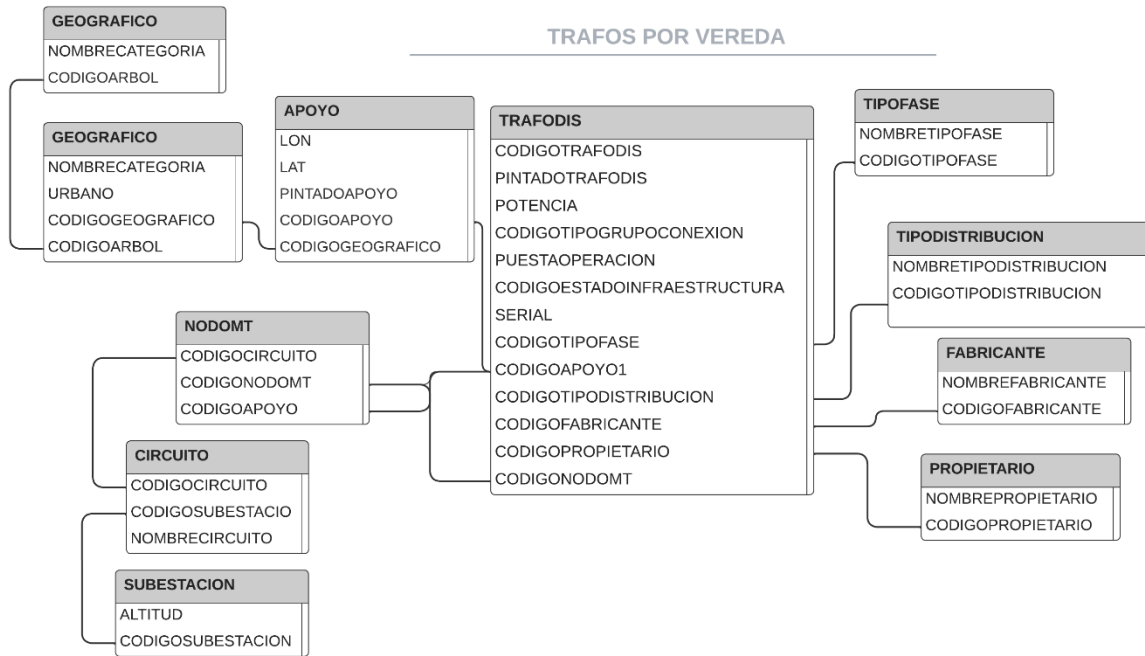


Tabla 5**Lista de atributos de la tabla AGO_consulta_trafosporvereda**

Nombre del atributo	Tipo de dato	Longitud
PINTADO_TRAFOSDIS	varchar	varchar(7)
CODIGO_TRAFODIS	float	float
CODIGO_TIPO_GRUPO_CONEXION	varchar	varchar(2)
POTENCIA	float	float
LON	float	float
LAT	float	float
ALTITUD	float	float
PUESTA_OPERACION	datetime	datetime
CODIGO_ESTADO_INFRAESTRUCTURA	varchar	varchar(1)
PINTADO_APOYO	varchar	varchar(7)
SERIAL	varchar	varchar(20)
NOMBRE_TIPO_FASE	varchar	varchar(30)
CODIGO_TIPO_FASE	varchar	varchar(1)
NOMBRE_PROPIETARIO	varchar	varchar(30)
NOMBRE_FABRICANTE	varchar	varchar(30)
NOMBRE_TIPO_DISTRIBUCION	varchar	varchar(30)
CODIGO_CIRCUITO	varchar	varchar(7)
NOMBRE_CIRCUITO	varchar	varchar(30)
DIRECCION	varchar	varchar(4000)
BARRIO_VEREDA	varchar	varchar(80)
MUNICIPIO	varchar	varchar(80)
URBANO	varchar	varchar(1)

6.1.2.2 Diseño lógico a físico para área de Subgerencia de Conexiones. En Subgerencia de Conexiones, área en la que se identificó para el proceso de Gestión de Pérdidas la consulta Recuperación de Energía Eléctrica, se tuvo como resultado del análisis de las relaciones de los datos el modelo lógico ilustrado en la Figura 18. Finalmente, fue a partir de la revisión de los tipos de datos que se diseñó la tabla origen PER_consulta_recuperacion, ubicada en el esquema PER_esquema y cuyos atributos se especifican en la Tabla 6.

Figura 18**Modelo lógico para la consulta Recuperación SAC del área de Gestión de Pérdidas**



Nota. Fuente elaborado por los autores.

Tabla 6

Lista de atributos de la tabla *PER_consulta_recuperacion*

Nombre del atributo	Tipo de dato	Longitud
NUMERO_PROCESO	float	float
CODIGO_CUENTA	float	float
NOMBRE	varchar	varchar(100)
DIRECCION	varchar	varchar(80)
MUNICIPIO	varchar	varchar(30)
SALDO_ACTUAL	float	float
FECHA_REVISION	datetime	datetime
VALOR_SANCION	float	float
ENERGIA_NO_FACTURADA	float	float
CONSUMO_SANCION	float	float
USER_SISTEMA	varchar	varchar(80)
FECHA_SISTEMA	datetime	datetime
FECHA_LIQUIDACION	datetime	datetime
ESTADO	varchar	varchar(1)

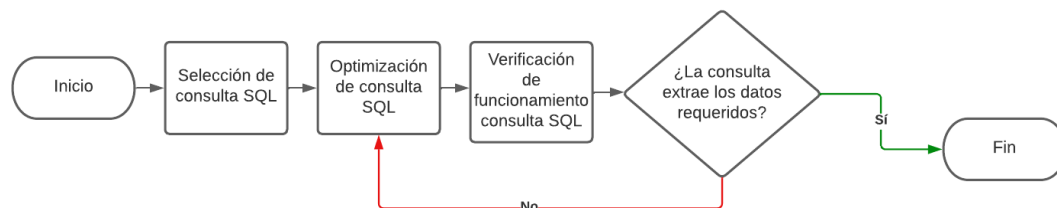
6.2 Diseño del subsistema ETL

En el diseño de los procesos ETL, se hizo una división por subproceso: extracción, transformación y carga, cada uno sobre cada consulta identificada dentro de cada área. A continuación, se explica qué implicó cada subproceso para ambas áreas.

En el diseño de la extracción, se aseguró la selección de las fuentes de datos adecuadas y se realizaron cambios siguiendo el flujo mostrado en la Figura 19 en las consultas SQL para mejorar la eficiencia en la extracción de datos. Una vez que se realizaron estas mejoras, se ejecutaron estas sentencias SQL en la base de datos de origen para verificar que las modificaciones se hubieran implementado correctamente.

Figura 19

Flujo de actividades para mejoras SQL



Nota. La figura contiene el flujo de actividades para mejoras en el código SQL correspondiente a las consultas por cada área. Fuente elaborado por los autores.

Para la transformación, se llevó a cabo un análisis de los datos que serían extraídos, con el objetivo de establecer las reglas de transformación que serían aplicables. Durante este proceso, se tuvieron en cuenta las necesidades identificadas en las entrevistas realizadas a los usuarios para el tratamiento de las cuatro consultas entre las dos áreas. Se identificó que algunos atributos numéricos tenían valores nulos en los datos extraídos. Con el fin de facilitar el análisis de los datos y evitar problemas en los cálculos posteriores en los tableros de inteligencia de negocios, Se

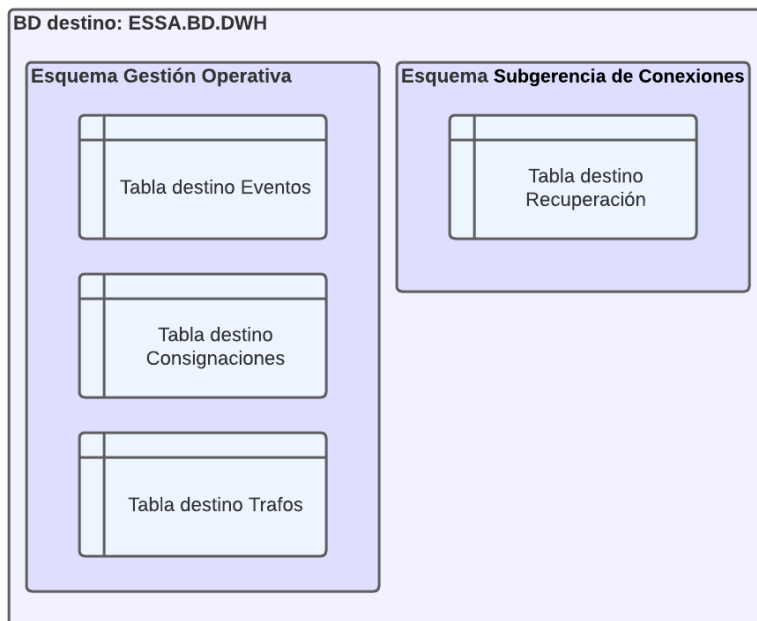
estableció una regla de transformación que asignaba el valor 0 a aquellos registros cuya columna era de tipo flotante, como se puede observar en la tabla seis, en los campos: número de proceso, código de cuenta, saldo actual, valor sanción, energía no facturada y consumo sanción. Cabe mencionar que esta regla se aplicó de manera similar en otras consultas.

Y finalmente, para el diseño de la carga de datos se proyectó hacia la base de datos que utiliza el motor SQL Server, esta BD se diseña en ese motor de acuerdo con las especificaciones proporcionadas por el equipo de TI. Para realizar esta carga de datos, se apunta a cada tabla correspondiente a las consultas SQL definidas previamente.

En la Figura 20, se puede observar cada consulta SQL con su respectiva tabla, y a su vez cada tabla en su correspondiente esquema.

Figura 20

Diseño de la BD destino



Nota. Fuente elaborado por los autores.

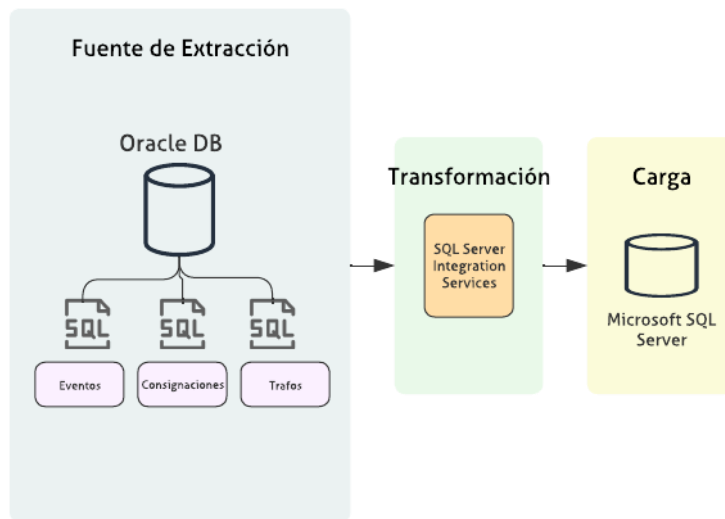
Por otro lado, el subproceso de carga en el ETL conllevó ciertas particularidades que fueron abordadas de manera diferente. A continuación, se explicará cómo se llevó a cabo el proceso de carga en cada una de las áreas.

6.2.1 Gestión Operativa

La fuente de datos utilizada para la extracción es una base de datos transaccional de Oracle que contiene datos relevantes para el área de Gestión Operativa y Subgerencia de Conexiones. Se identificaron tres consultas SQL para extraer los datos necesarios en el área de Gestión Operativa: eventos registrados en el sistema eléctrico, las consignaciones y los transformadores por vereda. Ver Figura 21.

Figura 21

Extracción Oracle BD para Gestión Operativa.



Nota. Fuente elaborado por los autores.

6.2.1.1 Consulta Eventos. La consulta de eventos se encarga de recopilar información detallada sobre los eventos que ocurren en el sistema eléctrico. Cada vez que se produce una falla en el sistema eléctrico de ESSA, se registra como un evento en la base de datos transaccional de

Oracle. Mediante la consulta SQL, se extraen columnas importantes que incluyen el identificador único del activo relacionado al evento, el identificador del evento, el número de transformador, la fecha y hora de inicio y finalización del evento, la causa que dio origen al evento, la fecha y hora de registro, la duración del evento calculada a partir de la diferencia entre la fecha final e inicial, y el período de extracción del registro.

Para asegurar una extracción correcta de los datos, se diseñó una lógica de filtrado basada en fechas. En la primera consulta, anteriormente se seleccionaban los eventos cuya fecha de corte estaba dentro del período especificado manualmente por el administrador CEE. Sin embargo, se realizó una mejora en código de la consulta SQL para obtener el período actual de forma automática, evitando la necesidad de ingresar el período manualmente. Esto se logró utilizando funciones de truncado automático, que obtienen el primer día del mes actual y el primer día del mes siguiente. De esta manera, se asegura la inclusión de los registros correspondientes al mes en curso. Al utilizar la función de truncado automático en lugar de solicitar un período digitado manualmente, se logra una extracción más conveniente y menos propensa a errores humanos.

6.2.1.2 Consulta Consignaciones. La consulta de consignaciones se encarga de recopilar datos relevantes sobre los permisos para trabajar en la infraestructura eléctrica. Estas consignaciones son registros que indican las tareas y trabajos a realizar, así como los equipos y profesionales responsables de su ejecución. Mediante la consulta SQL, se extraen columnas importantes que incluyen el código de la consignación, el tipo de consignación, el estado de la consignación, el equipo encargado del trabajo, las fechas programadas de inicio y finalización de los trabajos, los códigos de circuito y transformador, los detalles del profesional responsable, los trabajos a ejecutar durante la consignación y la causa que originó la misma.

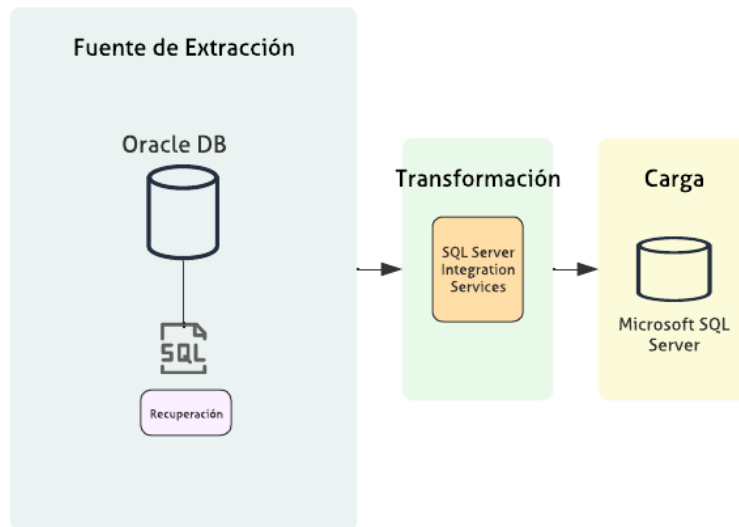
Es importante destacar que, al igual que en la consulta de eventos, se ha implementado una lógica de filtrado basada en fechas para garantizar la extracción de los registros correspondientes al período deseado. En este caso, se establece una condición que verifica si la fecha de inicio programada se encuentra entre la fecha actual truncada y la fecha resultante de agregar 12 meses a la fecha actual truncada. De esta manera, se seleccionan los registros cuya fecha de inicio programada está dentro de los próximos 12 meses a partir de la fecha actual. Esta ventana de tiempo de 12 meses se ha establecido en función de las historias de usuario recopiladas en el área y en base a las necesidades del administrador CEE. Anteriormente, el administrador CEE tenía incertidumbre sobre cuál sería el registro con la fecha más lejana, por lo que solía ingresar una fecha tentativa que cubriera un período considerable, por tal motivo se ha modificado la lógica de filtrado para seleccionar los registros cuya fecha de inicio programada esté dentro de los próximos 12 meses a partir de la fecha actual.

Este enfoque utilizando funciones de truncado automático evita la necesidad de ingresar manualmente el período en cada ejecución de la consulta, asegurando la consistencia de las fechas y simplificando el proceso de extracción de datos.

6.2.1.3 Consulta Trafos por vereda. En el caso de la tercera consulta, Trafos por vereda, no fue necesario realizar cambios significativos en el código SQL de extracción. Solo modificaciones de sintaxis.

6.2.2 Subgerencia de Conexiones

La fuente de datos utilizada para la extracción fue la misma base de datos Oracle, puesto que contiene datos relevantes para el área de Gestión Operativa y para el área de Subgerencia de Conexiones. Para esta última se identificó la consulta de Recuperación de energía eléctrica ver Figura 22.

Figura 22**Extracción Oracle BD para Subgerencia de Conexiones.**

Nota. Fuente elaborado por los autores.

6.2.2.1 Consulta Recuperación de Energía Eléctrica. Esta consulta SQL recopila datos sobre los correctivos relacionados con los procesos del sistema eléctrico cuando ocurre una pérdida de energía. Utilizando la consulta SQL proporcionada, se extraen varias columnas importantes: como el número de proceso, el código de cuenta, el nombre del cliente, la dirección, el municipio, el saldo actual, la fecha de revisión, el valor de la corrección, la cantidad de energía no facturada, el consumo, el usuario del sistema, la fecha del sistema y la fecha de liquidación.

En el diseño de la extracción, se realizaron mejoras en base a las historias de usuario recopiladas. Una de las mejoras consistió en eliminar columnas innecesarias de la consulta, las cuales anteriormente el Analista RCE tenía que eliminar manualmente utilizando herramientas como Excel. Además, se realizaron cambios en la sintaxis del código de extracción. Anteriormente, el Analista RCE no contaba con una consulta SQL base, sino que debía acceder a un sistema donde

se descargaban los datos en formato Excel. Ahora, con la colaboración del equipo de TI, se pudo obtener la consulta base en SQL que cumplía con los requisitos de extracción.

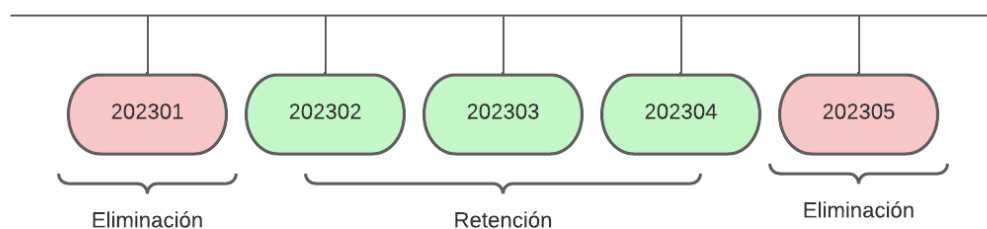
6.3 Diseño de la Retención de Datos

Se diseñó la retención de datos considerando las necesidades específicas de cada área. La retención de datos se aplicó únicamente a las consultas de Eventos, Consignaciones, y Recuperación de Energía.

En el caso de la consulta de Eventos, se diseñó una sentencia SQL que apunta a la tabla destino. Esta sentencia elimina registros de la tabla "AGO_consulta_eventos" que cumplen las siguientes condiciones: las filas que tienen un valor en la columna "periodo" correspondiente al mes y año actual, por ejemplo, si el mes actual es Mayo del 2023 el periodo seria 202305. Y las filas que tienen un valor en la columna "periodo" correspondiente al mes y año de hace cuatro meses, tal y como se observa en la Figura 23.

Figura 23

Retención de datos consulta Eventos

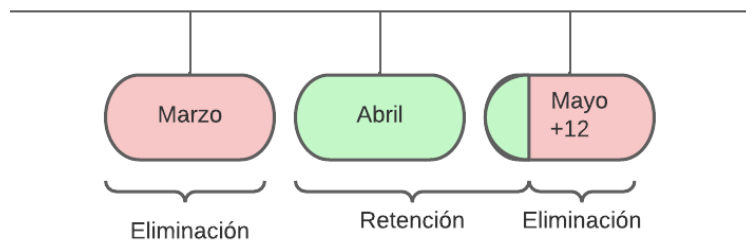


Nota. Fuente elaborado por los autores. Para efectos del ejemplo, se toma la fecha actual como 12 de mayo del 2023.

Para la consulta de Consignaciones, se diseñó una sentencia SQL que elimina registros de la tabla "AGO_consulta_consignaciones" según las siguientes condiciones: la columna fecha inicio programada elimina los registros entre la fecha actual y 12 meses después de la fecha actual. Además, la consulta elimina los registros entre el 1 de hace dos meses y el primer día del mes anterior al mes actual ver Figura 24.

Figura 24

Retención de datos consulta Consignaciones



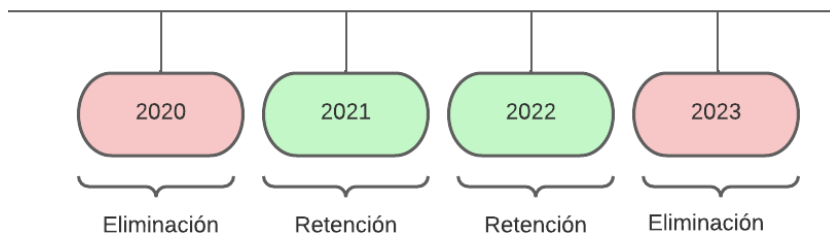
Nota. Fuente elaborado por los autores. Para efectos del ejemplo, se toma la fecha actual como 12 de mayo del 2023.

En el caso de la consulta de recuperación, se diseñó una sentencia SQL que pudiera eliminar los datos de la tabla "PER_consulta_recuperacion" basándose en las siguientes condiciones: la columna fecha liquidación tiene un valor que está entre el 1 de enero del año y la fecha actual, y la columna fecha liquidación tiene un valor que está entre el 1 de enero hace tres años y el 31 de diciembre hace tres años. Estas condiciones permiten mantener los datos relevantes en la tabla de acuerdo con la ventana de tiempo especificada por cada administrador, ver

Figura 25.

Figura 25

Retención de datos consulta recuperación



Nota. Fuente elaborado por los autores.

7. Desarrollo

El desarrollo de la extracción, transformación y carga representa una parte significativa del tiempo y esfuerzo requeridos para construir un proyecto de inteligencia de negocios. Este proceso presenta desafíos debido a las múltiples restricciones externas que influyen en su desarrollo, como los requerimientos, las características de los datos fuente y las habilidades del personal disponible, entre otras. Aunque es difícil comprender completamente la complejidad y los recursos necesarios para el ETL, es crucial reconocer su importancia en el éxito del proyecto (Kimball, R., & Ross, M. 2013). En este capítulo está toda la experiencia vivida en el desarrollo del ETL, las pruebas de retención de datos en la base de datos destino y los desafíos que hubo a la hora de desarrollar el subsistema ETL.

7.1 Desarrollo del Subsistema ETL

Se comenzó creando la base de datos destino, siguiendo el diseño previamente planeado y los atributos para cada tabla según los tipos de atributos presentes en la base de datos origen. Esta base de datos destino, 'ESSA.BD.DWH', fue creada en SQL Server 2016. Después de crear la base de datos, se creó un archivo para el área de Gestión Operativa con tres paquetes de integración de servicios, uno para cada consulta a la base de datos origen, utilizando la herramienta SSIS, ver

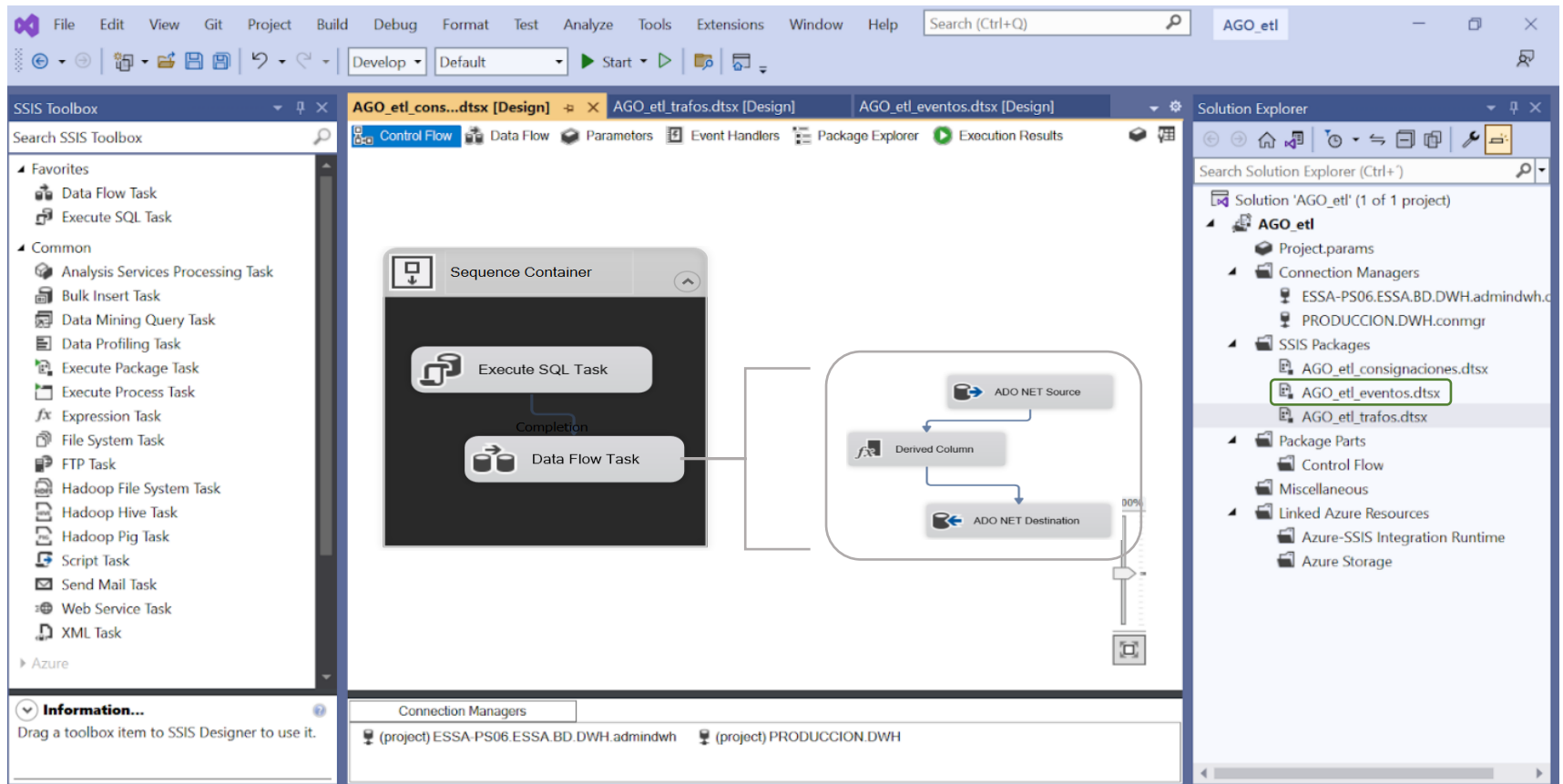
Figura 26. Para el área de subgerencia de conexiones, también se creó un archivo, pero con un solo paquete de integración de servicios, ya que en esta área se priorizó una única consulta para el proceso de ETL.

Según la documentación de Microsoft ("Integration Services (SSIS) Packages", s.f.) un paquete de integración de servicios es una colección de tareas y componentes que se utilizan para realizar diversas operaciones en el proceso de extracción, transformación y carga, ver Figura 27. Estos paquetes se crean utilizando la herramienta SQL Server Integration Services (SSIS), que es parte de la suite de herramientas de Microsoft SQL Server.

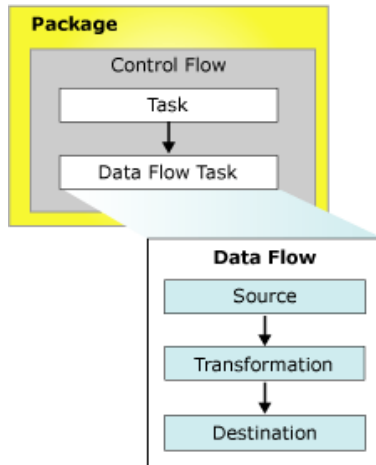
Para las cuatro consultas, se empleó un mismo flujo de componentes de herramientas, a excepción del contenido de cada componente, que incluía consultas SQL específicas para cada escenario de extracción y retención, así como referencias a los esquemas correspondientes. Con el fin de ofrecer una explicación más clara, se describirá el paquete de la consulta de Eventos.

Figura 26

Paquete de integración de servicios para la consulta Eventos del área de Gestión Operativa



Nota. Fuente elaborado por los autores.

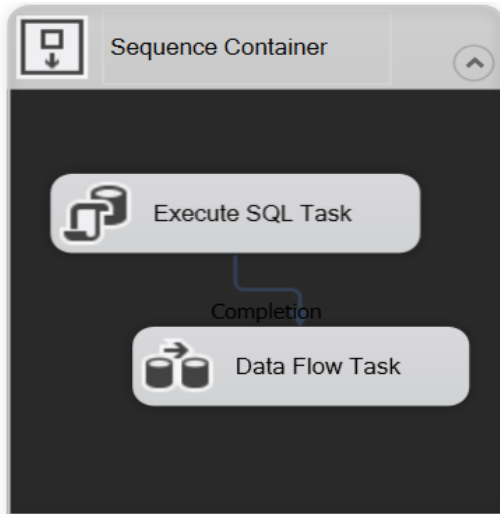
Figura 27**Paquete de integración de datos**

Nota. La figura muestra un paquete simple que contiene un flujo de control con una tarea de flujo de datos, que a su vez contiene un flujo de datos. Fuente Integration Services Packages(s.f.).

El paquete de integración de servicios de la consulta de Eventos consta de un contenedor de secuencias con una tarea de ejecución SQL y una tarea de flujo de datos ver Figura 28. En la tarea de ejecución de SQL se encuentra el código SQL para eliminar los datos según la retención establecida, apuntando a la base de datos destino. La tarea de flujo de datos, como su nombre indica, se encarga de organizar el flujo de datos. Por tanto, dentro de esa tarea existen tres componentes responsables de dicho flujo, ver Figura 29: El primer componente consiste en un origen ADO.NET que se conecta a la base de datos origen de Oracle. El segundo componente es una columna derivada que transforma las columnas numéricas extraídas de la base de datos de origen y reemplaza los valores nulos por ceros. Por último, el tercer componente es un destino ADO.NET que se conecta a la base de datos destino, realiza el mapeo columna por columna y carga los datos en la base de datos destino.

Figura 28

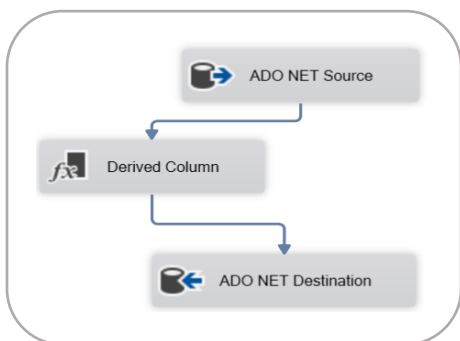
Contenedor de Secuencias



Nota. Fuente interfaz Visual Studio.

Figura 29

Componentes dentro de una tarea de flujo de datos



Nota. Fuente interfaz Visual Studio.

7.2 Pruebas de ETL

El diseño y ejecución de pruebas en el proyecto fue importante debido a que así se garantizaba la precisión del diseño creado anteriormente y la implementación de este ya en la fase de desarrollo. De esta manera, se pudo asegurar que el proceso ETL, y la retención de datos se

hacían de manera correcta, permitiendo que se conservaran los datos especificados en la documentación de requerimientos para cada área. Para ello, fue necesario analizar qué procedimientos se debían hacer para permitir que los datos requeridos estuviesen de manera íntegra y oportuna en el data mart.

7.2.1 Consulta de Eventos

Se realizaron pruebas para la consulta de eventos del área de Gestión Operativa para comprobar que la cantidad de datos transferidos por el proceso ETL fuesen los mismos que descargaba de manera manual y usaba para reportes de BI el Administrador CEE. Adicionalmente, fue importante verificar que el paquete SSIS hiciera la eliminación de datos de tal manera que se permitieran borrar los datos del periodo actual para luego actualizarlos, y los datos de hace cuatro períodos, permitiendo que se conservaran en el data mart datos de los últimos tres periodos y el periodo actual.

7.2.2 Consulta de Consignaciones

Se realizaron pruebas para la consulta de consignaciones del área de Gestión Operativa teniendo en cuenta que se especificó que se requerían los datos desde el día actual hasta 12 meses después, y que se debían retener en el data mart los datos correspondientes al mes anterior y al actual. Para ello se diseñaron y ejecutaron pruebas que verificaban que se hiciera la eliminación de los datos del día actual a doce meses, para posteriormente ejecutar el ETL que actualizaría los datos de ese mismo rango de tiempo, y pruebas para verificar el borrado de los datos de hace dos meses, permitiendo que hubiese una efectiva retención de datos.

7.2.3 Consulta de Trafos por Vereda

La consulta de trafos por vereda del área de Gestión Operativa, dado que no requería retención de datos por la naturaleza de la consulta misma, solo hubo necesidad de probar el

funcionamiento del ETL. Así que se verificó que el paquete SSIS cumpliera con el requerimiento de realizar primero un borrado de datos para posteriormente actualizar el data mart con nuevos datos, y se comprobó que los datos procesados por el ETL fuesen los mismos que usaba el Administrador CEE en sus reportes BI.

7.2.4 Consulta de Recuperación de Energía

La consulta de recuperación de energía del área de Subgerencia de Conexiones requiere realizar el proceso ETL sobre datos desde el primer día del año actual hasta el día actual, y se deben retener los datos de los últimos dos años más el actual. Para probar su funcionamiento se hicieron pruebas de verificación de borrado de los datos del año actual para su posterior actualización, y de borrado de hace tres años, garantizando de esa manera que en el data mart se retengan los datos de los últimos dos años y el actual.

7.3 Desafíos

Durante la etapa del desarrollo de los ETL en paquetes de SSIS y la automatización de su ejecución mediante jobs de SQL Server, surgieron diferentes desafíos que fueron arduos y exigentes en cuanto a que requirieron de mucho tiempo de estudio y experimentación. A continuación, se describirán esas dificultades encontradas y el aprendizaje que resultó de cada una de ellas.

7.3.1 ETL

Durante el desarrollo del subsistema ETL se encontraron diversos desafíos debido al desconocimiento que en un principio se tenía sobre la herramienta. A pesar de haber realizado una fase previa de autocapacitación de herramientas, se encontraron desafíos que no se habían enfrentado antes, esa formación de herramientas se hizo desde un ambiente no empresarial, desde un solo computador y con una cantidad mínima de datos; de modo que cuando se empezó a hacer

el desarrollo en un entorno empresarial, usando servidores, diferentes usuarios y una mayor cantidad de datos, se presentaron errores antes no previstos ni abordados.

Entre los diversos errores que surgieron durante el desarrollo del subsistema ETL, se describirán aquellos que requirieron mayor tiempo de resolución.

Durante los primeros desarrollos para las consultas del área de Gestión Operativa, se notaba la particularidad de que la herramienta SSIS indicaba que la ejecución del paquete había sido completada, pero no indicaba con el símbolo de completado que todos los componentes del paquete se hubiesen ejecutado, independientemente de si la ejecución era exitosa o no. Aquel error no se encontró en ninguno de los foros ni en las documentaciones que tenía Microsoft disponibles en el momento. Por lo tanto, empíricamente se realizaron cambios en el paquete de SSIS, y estos a su vez fueron registrados para ir analizando de qué manera se podía solucionarlo. Para ello se estudió qué significaba cada una de las configuraciones que se le podían hacer al componente (ver Figura 30). Como resultado, cuando se hicieron ciertos cambios en las configuraciones de los componentes ODBC, se logró que ya se completara la ejecución y que todos los componentes indicaran con el símbolo que sí se habían ejecutado y de manera exitosa. Sin embargo, durante una ejecución posterior, habiendo procesado una mayor cantidad de datos, surgió un problema adicional, algunos datos que originalmente no eran nulos se estaban cargando como nulos a la base de datos destino.

La herramienta SSIS presentaba mensajes de error relacionados con datos dañados en una situación de poca memoria. Se trataron de identificar diferentes posibles causas, como la escasez de memoria y la incompatibilidad de los tipos de datos, y aunque desde el diseño se había establecido cómo crear la base de datos destino para que no sucediera dicha incompatibilidad, se probó rediseñándola para ver si se solucionaba el problema. Finalmente, se buscó apoyo externo

y se contactó a un ingeniero de una filial del grupo empresarial EPM que estaba familiarizado con este tipo de proyectos dado que tenía experiencia debido al desarrollo avanzado que tienen en esa empresa con su proyecto data warehouse. Gracias a su apoyo se pudo identificar que el problema realmente surgía desde la clase de componentes usados durante el desarrollo, dado que el tipo de conexión que se había hecho en un principio era de tipo ODBC. Como resultado, se pudo analizar las ventajas que tienen las conexiones de tipo ADO.NET (que fueron recomendadas por el ingeniero de otra filial) sobre las conexiones ODBC, que fueron las que inicialmente se usaron en el proyecto.

Figura 30

Configuración del componente de tipo ODBC

[ODBC Destination ESSABDDWH].[BatchSize]	1000
[ODBC Destination ESSABDDWH].[BindCharColumnsAs]	0
[ODBC Destination ESSABDDWH].[BindNumericAs]	0
[ODBC Destination ESSABDDWH].[DefaultCodePage]	1252
[ODBC Destination ESSABDDWH].[InsertMethod]	1
[ODBC Destination ESSABDDWH].[LobChunkSize]	32768
[ODBC Destination ESSABDDWH].[StatementTimeout]	0
[ODBC Destination ESSABDDWH].[TableName]	
[ODBC Destination ESSABDDWH].[TransactionSize]	1000
[ODBC Source ENERGIS].[AccessMode]	1
[ODBC Source ENERGIS].[BatchSize]	1000
[ODBC Source ENERGIS].[BindCharColumnsAs]	0
[ODBC Source ENERGIS].[BindNumericAs]	0
[ODBC Source ENERGIS].[DefaultCodePage]	1252
[ODBC Source ENERGIS].[ExposeCharColumnsAsUnicode]	False
[ODBC Source ENERGIS].[FetchMethod]	1
[ODBC Source ENERGIS].[LobChunkSize]	32768
[ODBC Source ENERGIS].[SqlCommand]	SELECT D.CODIG
[ODBC Source ENERGIS].[StatementTimeout]	0
[ODBC Source ENERGIS].[TableName]	
AutoAdjustBufferSize	False
BLOBTempStoragePath	
BufferTempStoragePath	
DefaultBufferMaxRows	10000
DefaultBufferSize	10485760
EngineThreads	10
ExecValueVariable	<none>

La propiedad **AutoAdjustBufferSize** indica si el tamaño del búfer se ajustará automáticamente en función del valor de **DefaultBufferMaxRows**. Si está habilitada, el tamaño del búfer se ajustará automáticamente para acomodar el número máximo de filas especificado en **DefaultBufferMaxRows**.

La propiedad **DefaultBufferMaxRows** especifica el número máximo de filas que se pueden almacenar en cada búfer. El valor predeterminado es de **10.000 filas**.

La propiedad **DefaultBufferSize** especifica el tamaño del búfer. Por defecto, el tamaño del búfer es de **10 megabytes**. Si no se establece el tamaño en la propiedad, la tarea utilizará el tamaño predeterminado del búfer especificado en la propiedad **DefaultBufferSize**.
 $10 \text{ megabytes} * 1.048.576 \text{ bytes/megabyte} = 10.485.760 \text{ bytes}$

Nota. Fuente interfaz Visual Studio y notas elaboradas por los autores.

Una vez fueron cambiados los tipos de componentes de conexión ODBC, se encontró que el componente ADO.NET Source no requería de todas las configuraciones que sí necesitaba el componente ODBC, haciendo de su desarrollo algo más automático y fluido, además de que se notó una reducción en el tiempo de ejecución de este.

7.3.2 Automatización

En la programación de Jobs en SQL Server para ejecutar automáticamente los paquetes SSIS era probable enfrentar desafíos debido a que era la primera vez de las autoras utilizando la herramienta SSIS y SQL Server en un entorno empresarial y de producción. Así que esto sumado a que desde infraestructura hacía falta la configuración del componente SSIS en el Microsoft SQL Server, hacía que durante el desarrollo surgieran dificultades que impedían avanzar en el proyecto.

7.4 Creación de trabajos para ETL

Un Job en el Agente SQL Server es una tarea programada que permite ejecutar paquetes de integración de servicios de forma automatizada. Sin embargo, al intentar crear este Job, se encontraron varios errores. Uno de los errores fue la diferencia de versiones. Según la documentación de Microsoft (“Installing Integration Services versions side by side”, s.f.) las versiones de SQL Server pueden tener compatibilidad limitada con SSIS, por eso Microsoft recomienda utilizar las versiones de SSIS incluidas en la versión correspondiente de SQL Server para evitar problemas de compatibilidad con las herramientas y componentes más nuevos.

En el caso de la empresa, se estaba utilizando SQL Server 2016 con el servicio de integración de SSIS 2022. Se tomaron acciones para resolver este problema. Dado que la empresa no tenía acceso a una versión más reciente de SQL Server, se decidió realizar el desarrollo en versiones anteriores. A continuación, se presenta una tabla que muestra la versión de SQL Server junto con su servicio de integración correspondiente, ver Tabla 7. En ese momento, el desarrollo

se migró a la versión de 2016, que se conocía como SSDT (SQL Server Data Tools) para integración de datos.

Tabla 7

Tabla de compatibilidad entre SQL Server y el servicio de integración Microsoft

Versión de SQL Server	Servicio de integración
2016	SSDT
2017	SSDT
2019	SSIS
2022	SSIS

A pesar de que se intentó desarrollar en distintas versiones de igual forma los errores persistían a la hora automatizar el proyecto. Ante esta situación, se solicitó ayuda a una filial del grupo EPM. Sin embargo, debido a restricciones de tiempo, falta de personal capacitado y disponibilidad de soporte a la infraestructura de no fue posible concluir la programación de la ejecución de los Jobs de manera correcta. No obstante, se estableció un puente entre la filial y la Electrificadora de Santander para continuar avanzando en el proyecto. Además, se dejó una guía sobre cómo implementar un paquete de integración de servicios mediante el Agente SQL Server, de manera que, una vez resueltos los problemas de infraestructura, se pueda concluir la programación de los Jobs.

8. Guía de Trabajo para Data Marts

8.1 Identificación de la necesidad

La Electrificadora de Santander, como se ha mencionado en capítulos previos, tiene actualmente necesidades en soluciones digitales, entre ellas aquellas que darían apoyo a procesos de inteligencia de negocios. En este contexto, se reconoció la importancia de documentar toda la experiencia adquirida durante la ejecución de este proyecto, incluyendo recomendaciones, obstáculos encontrados en el desarrollo y acciones aplicables en este ámbito.

La documentación de esta experiencia resultó crucial, ya que permitiría replicar el proyecto en otras áreas de la empresa. Con el fin de lograr este objetivo, las autoras recomendaron la elaboración de una guía de trabajo durante el desarrollo del proyecto, que contiene las estrategias de trabajo utilizadas por ellas y la valiosa experiencia adquirida. Esta guía se convierte en un recurso fundamental para la Electrificadora de Santander al aplicar la gestión del conocimiento en su empresa, ya que como dice Díaz et al. (2005), la empresa que no aprende a sacar el mayor provecho a la información que tiene en su poder y prepara eficazmente el conocimiento que se genera, estará en desventaja frente otras empresas del entorno. Díaz además recalca que si las empresas descansan todo el capital de conocimiento sobre algunos individuos y no realizan acciones para compartirlo corren el riesgo de perderlo, muy por el contrario, la empresa que potencia su capital intelectual puede utilizar este activo para alcanzar los retos y oportunidades de negocios que el entorno brinda.

Al tener documentada esta experiencia y hacer uso posterior de ella, la Electrificadora de Santander demuestra su compromiso con la mejora continua y la aplicación de las mejores prácticas en el ámbito de la gestión del conocimiento. Además, la empresa se beneficia al facilitar su transferencia a futuros proyectos.

8.2 Estructura y contenido de la guía de trabajo

Mediante la elaboración de la guía de trabajo data mart, se permitió establecer una secuencia lógica para una ejecución de un proyecto data mart basada en la experiencia adquirida.

En un principio, en la guía se menciona una lista de roles propuestos por el grupo Kimball para un proyecto data warehouse, y que según las circunstancias del presente proyecto aplicaban. Posteriormente se describe la fase de planificación de proyecto y qué actividades implica. En cuanto a la definición de requerimientos, se proporciona información fundamental para hacer las entrevistas de manera eficiente y se dejaron planteadas estrategias para registrar las historias de usuario y los requerimientos en la documentación del proyecto. Para la priorización de requerimientos y establecimiento del alcance, se explicó qué implicaba y cómo se podía realizar, mencionando el uso de una tabla de priorización de requerimientos, y el método propuesto por Ralph Kimball y Margy Ross. En cuanto a la arquitectura técnica y las herramientas, se sentó la base para el diseño de la arquitectura técnica, se describen sus componentes, y se aborda la elección y documentación tanto de estos como de las herramientas usadas.

Respecto a las fases de diseño del modelado dimensional, diseño físico de la base de datos y diseño y desarrollo del sistema ETL; estas se describieron de manera general y se mencionaron cómo manejar ciertas situaciones dependiendo de las particularidades de cada proyecto.

Y finalmente, en la sección de apéndices se incluyen estrategias usadas durante el proyecto, formatos propuestos, una descripción técnica y detallada sobre los retos afrontados, los errores e incompatibilidades halladas, así como tutoriales para la implementación o despliegue de paquetes de SSIS en un servidor y la configuración del Agente SQL Server para la ejecución y programación de dichos paquetes.

8.3 Impacto de la Guía de Trabajo

Es conocido desde la administración en la empresa que el área de Subgerencia de Conexiones ha experimentado por muchos años una creciente demanda de soluciones digitales para temas de analítica de datos. Sin embargo, fue recientemente, a través de una consultoría llevada a cabo en la empresa, que se reconoció formalmente la importancia y necesidad de implementar un data warehouse como una solución digital en la organización.

Desde el principio, la guía data mart desempeñó un papel fundamental en el avance hacia la solución de esta necesidad, debido a que uno de los data marts desarrollados fue específicamente para el área de Subgerencia de Conexiones, la cual había expresado durante mucho tiempo su necesidad de contar con un data warehouse, pero que no había sido sino hasta ahora que se hacía formalmente un análisis de la necesidad. Este proyecto sirvió como caso de estudio base para mostrar los beneficios que puede tener un data mart a la hora de satisfacer las necesidades de datos para su posterior análisis y proceso de inteligencia de negocios. Dado que había expectativas acumuladas en torno a la implementación del data mart y la elaboración de la guía, fueron los resultados positivos del proyecto los que generaron un mayor entusiasmo e interés en la organización por explorar y tomar acción sobre ampliar las posibilidades de un proyecto data warehouse futuro.

En un principio, debido a la falta de conocimiento sobre cómo llevar a cabo el proyecto data mart, los resultados de las investigaciones y autocapacitaciones hechas quedaron consignados en la guía, de tal manera que resultó ésta siendo un recurso base fundamental, no solo para el proyecto data mart, sino también para el avance hacia la solución de la necesidad desde Subgerencia de Conexiones.

Los profesionales desde el área de Subgerencia de Conexiones que estaban al frente de ese futuro proyecto data warehouse quisieron entender qué avances se habían hecho en esa dirección. Como parte del compromiso y apoyo al área, se participó de manera activa en reuniones, donde se presentaron los progresos logrados en el proyecto y los conocimientos recopilados en la guía data mart. Estas reuniones brindaron la oportunidad de compartir la experiencia y conocimientos adquiridos, los cuales fueron registrados en la guía data mart. Además de que permitieron la resolución de dudas por parte de los profesionales del área. Estas reuniones fueron finalmente un espacio para abordar aquellos conceptos clave, como la arquitectura técnica de un data mart, el diseño, el desarrollo, y los pasos a seguir recomendados para escalar el proyecto.

8.4 Conferencias Informativas

Durante el período de práctica se tuvo la oportunidad de compartir los resultados y hallazgos del proyecto a través de conferencias informativas, estas conferencias informativas se llevaron a cabo en el área de servicios corporativos, específicamente para el equipo de Tecnologías de la Información y el equipo de Desarrollo Humano y organizacional, además del área de subgerencia de conexiones. Estas conferencias informativas presentaron detalles del proyecto, la metodología utilizada y los resultados obtenidos, pasos a seguir y recomendaciones por parte de las autoras.

Esta serie de presentaciones fueron importantes para compartir la experiencia con miembros de la organización y destacar la relevancia del proyecto en el contexto de la empresa, además, las conferencias permitieron interactuar con los asistentes y responder sus preguntas, lo que generó un valioso intercambio de ideas y perspectivas.

9. Conclusiones

En este informe se presenta el trabajo realizado durante la práctica empresarial, enfocado en la mejora de los procesos de extracción, transformación y carga en el área de Gestión Operativa y el área de Subgerencia de Conexiones.

Se logró reducir la cantidad de tareas realizadas por el administrador CEE, pasando de 5 pasos a solo 1; conectarse desde el tablero de inteligencia de negocios al data mart. Además, se observó una notable mejora en el tiempo empleado por el administrador CEE en los procesos que hacía antes. Según las entrevistas realizadas, anteriormente el proceso requería de un aproximado de 4 horas semanales, pero gracias a las mejoras implementadas, este tiempo se redujo a aproximadamente 5 minutos por cada consulta. Esto beneficia al área Gestión Operativa al simplificar y agilizar el proceso, permitiendo al trabajador enfocarse en actividades más estratégicas y de mayor valor agregado.

Para el área de subgerencia de conexiones también hubo una mejora significativa en la cantidad de tareas realizadas por el analista RCE, pasando de 8 pasos a solo 1, también una mejora en el tiempo empleado por el analista ya que gastaba aproximadamente 30 minutos semanales en esas tareas, tiempo que se redujo a 5 minutos, para la consulta de recuperación de energía eléctrica.

Se logró una correcta puesta a disposición de los datos, considerando las características y criterios del usuario, y se acompañó al usuario durante todo el ciclo de vida del proyecto. Estas mejoras contribuyen a cerrar la brecha en términos de analítica de datos que existía en la Electrificadora de Santander. Si bien la organización cuenta con diversas bases de datos, carecía de una base de datos enfocada en el análisis de datos. Se espera que este proyecto sea un punto de partida para cerrar dicha brecha en otras áreas de la organización y fomentar una cultura de análisis de datos en toda la empresa.

Por otra parte, la ejecución de esta práctica empresarial permitió afianzar los conocimientos adquiridos durante el programa de Ingeniería de Sistemas y obtener nuevas habilidades en proyectos de inteligencia de negocios, herramientas la integración de servicios, metodologías para la creación de bodegas de datos, automatización de actividades, análisis de procesos, recolección de requerimientos, priorización de requerimientos y gestión de proyectos, Asimismo, durante esta práctica empresarial conjunta, se adquirió valiosas destrezas profesionales en áreas fundamentales como la comunicación y el desarrollo de proyectos en un entorno empresarial. Se pudo obtener la oportunidad de poner en práctica nuestras habilidades de comunicación tanto oral como escrita, interactuando con diversos stakeholders y presentando y sustentando avances y resultados de manera efectiva. Además, se experimentó el dinamismo y los desafíos propios de un entorno practico de ingeniería, lo que ha permitido desarrollar habilidades de resolución de problemas, adaptabilidad y toma de decisiones en situaciones prácticas. En resumen, la práctica empresarial por parte de las autoras se finalizó cumpliendo las expectativas y los objetivos propuestos al inicio de esta.

Finalmente y a juicio del director del trabajo por parte de la Universidad, el presente proyecto realizado en la modalidad de práctica empresarial, es una muestra de la contribución que desde la academia es posible hacer al entorno empresarial, mediante acciones que asumen problemas y desarrollan soluciones de formas novedosas para las empresas, en este caso particular exigiendo a las autoras profundizar ampliamente en conocimientos y habilidades, incluso poco a abordadas en su formación previa,

Referencias Bibliográficas

Ballard, C. et al. (2005) Data Mart Consolidation Getting Control of Your Enterprise Information. San Jose, CA: IBM.

Bernabeu, R. D. (2010). HEFESTO: Metodología para la Construcción de un Data Warehouse. Data Warehousing: Investigación y Sistematización de Conceptos. Córdoba, Argentina.

Curto Díaz, J. (2010). Introducción al Business Intelligence. Universitat Oberta de Catalunya. <https://www-digitaliapublishing-com.bibliotecavirtual.uis.edu.co/a/20242>

Curto Díaz, J. (2016). Introducción al business intelligence. Editorial UOC. <https://elibro-net.bibliotecavirtual.uis.edu.co/es/lc/uis/titulos/101030>

Díaz, D. G., de Armas, M. P., & Valladares, I. C. (2005). Gestión del conocimiento y su importancia en las organizaciones. *Ingeniería Industrial*, 26(2), 6.

ELECTRIFICADORA DE SANTANDER. (s.f.). Quiénes somos. ESSA. Recuperado el 4 de abril de 2023, de <https://www.essa.com.co/site/informacion-corporativa/quienes-somos>

ELECTRIFICADORA DE SANTANDER. (s.f.). Qué hacemos. ESSA. Recuperado el 5 de abril de 2023, de <https://www.essa.com.co/site/informacion-corporativa/que-hacemos>

Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., & Becker, B. (2008). The Data Warehouse Lifecycle Toolkit (2nd ed.). John Wiley & Sons, Inc.

Kimball, R., & Ross, M. (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling (3rd ed.). John Wiley & Sons, Inc.

Medina la Plata, E. (2012). *Business Intelligence: Una guía práctica*. Universidad Peruana de Ciencias Aplicadas (UPC). <https://www-digitaliapublishing-com.bibliotecavirtual.uis.edu.co/a/45568>

Molina, J. y García, J. (2010). Técnicas de Análisis de Datos.

http://matema.ujaen.es/jnavas/web_recursos/archivos/weka%20master%20recursos%20naturales/apuntesAD.pdf

Moscoso-Zea, O., & Luján-Mora, S. (2016, September). Datawarehouse design for educational data mining. In 2016 15th International Conference on Information Technology Based Higher Education and Training (ITHET) (pp. 1-6). IEEE.

Microsoft. (s.f.). Integration Services (SSIS) Packages. Recuperado el 10 de junio de 2023, de <https://learn.microsoft.com/en-us/sql/integration-services/integration-services-ssis-packages?view=sql-server-ver16>

Microsoft. (s.f.). Installing Integration Services versions side by side. Recuperado el 10 de junio de 2023 de <https://learn.microsoft.com/en-us/sql/integration-services/install-windows/installing-integration-services-versions-side-by-side?view=sql-server-ver16>

Yessad, L., & Labiod, A. (2016, November). Comparative study of data warehouses modeling approaches: Inmon, Kimball and Data Vault. In 2016 International Conference on System Reliability and Science (ICSRS) (pp. 95-99). IEEE.

Apéndices

Apéndice A. Resumen de la guía de trabajo para el desarrollo de data marts

ESSA, consciente de la importancia del capital intelectual en la toma de decisiones, enfrenta problemas comunes en el análisis de datos en sistemas transaccionales. Para abordar estos desafíos, creó una guía para desarrollar data marts, automatizando la consolidación y validación de datos para informes de inteligencia empresarial. Basada en la experiencia de un proyecto piloto, la guía detalla fases clave, desde la planificación hasta el mantenimiento.

Sobre el contenido de la guía, esta ofrece apoyo para la adopción de la metodología Kimball, basada en un ciclo de vida dimensional de negocio, enfoque bottom-up, que permite una implementación ágil de data marts. Además, la guía incluye pasos, procesos y buenas prácticas de otras metodologías.

Roles

Usuario empresarial / Área usuaria: utiliza datos del data mart para tomar decisiones estratégicas sobre la empresa, clientes, procesos y productos. Este usuario incluye gerentes, directores de ventas y técnicos, entre otros.

Patrocinador de Negocio / Business Sponsor: brinda apoyo y garantiza la finalización exitosa del proyecto. También puede ayudar a identificar y resolver problemas, asegurando que se alcancen los objetivos del proyecto.

Gerente de DW/BI / DW/BI Director: supervisa el desarrollo y mantenimiento del data warehouse y los sistemas de inteligencia de negocios. Trabaja en colaboración con la organización para garantizar la eficiencia y efectividad de los proyectos de DW/BI.

Gestor de Proyecto / Project Manager: planifica, ejecuta y evalúa proyectos. Se encarga de gestionar el proyecto de principio a fin, comunicarse con el equipo y partes interesadas, gestionar

riesgos, problemas y recursos, y garantizar que se cumplan los objetivos dentro del presupuesto y el plazo establecido.

Líder de Proyecto de Negocio / Business Project Lead: colabora con los equipos para comprender los requisitos del proyecto, coordina el trabajo del equipo, supervisa el progreso y garantiza el cumplimiento de los objetivos de calidad y tiempo. Se enfoca en la gestión de aspectos relacionados con el negocio del proyecto y presenta los resultados obtenidos.

Analista de Negocios / Business Analyst: evalúa requerimientos, comprende procesos comerciales y contribuye a desarrollar soluciones de negocios.

Data Steward: establece estándares de calidad de datos, gestiona integridad y metadatos, y colabora en políticas de gestión de datos.

Modelador de Datos / Data Modeler: diseña modelos para presentar información de forma comprensible, trabajando con otros roles y documentando.

Arquitecto y Desarrollador ETL: diseña y ejecuta procesos ETL eficientes.

Arquitecto Técnico o Especialista en Soporte Técnico: diseña la infraestructura, monitorea el rendimiento y brinda soporte técnico.

Gerente de Seguridad/Security Manager: asegura los datos en el data mart, establece estándares y herramientas, y brinda asesoramiento sobre seguridad.

Líder de Testeo: diseña y ejecuta pruebas, analiza resultados y sugiere mejoras.

Educador: desarrolla documentación y proporciona entrenamiento al usuario.

Fase: Planificación del proyecto

En esta fase es importante determinar la justificación y el objetivo general del proyecto, establecer un plan de comunicación del equipo y desarrollar estrategias de seguimiento y control.

Bajo un proyecto data mart es posible encontrar dos casos. El primero, cuando no hay ningún

proceso de analítica existente, pero se tienen unos datos que se quieren aprovechar, y aún no están identificadas las preguntas de negocio. Y el segundo, cuando ya hay un proceso de analítica alimentado por un proceso manual y repetitivo de extracción y carga de datos. A lo largo del documento se mencionará cómo proceder según el caso.

Fase: Definición de requerimientos

Entrevistas para requerimientos

Agendar una reunión con todo el equipo del proyecto y el área usuaria con el fin de hacer un primer acercamiento a los requerimientos del usuario. Kimball recomienda evitar hacer preguntas directas sobre aspectos técnicos. Además, es importante hablar con los usuarios sobre sus trabajos, objetivos, y entender cómo toman decisiones hoy y de qué manera las podrían tomar mejor en el futuro.

Asimismo, se podrá preguntar si poseen tableros BI o procesos de análisis de datos que impliquen que algún empleado de manera manual y repetitiva extraiga datos de alguna fuente. La respuesta a esto definirá qué caso de proyecto data mart será.

Entrevista de datos

Estas entrevistas tienen como objetivo obtener información detallada sobre los sistemas y tecnologías ya existentes. Los entrevistados en este caso son profesionales de TI como: arquitectos, desarrolladores, administradores de bases de datos, y demás. Como resultado de estas entrevistas se comprenderá cómo se gestionan e integran los datos en la empresa, qué desafíos técnicos hay y la identificación de fuentes de datos relevantes.

Documentación de los requerimientos revisados

Es importante documentar los requerimientos obtenidos para asegurar que hayan sido comprendidos e interpretados correctamente por todo el equipo de trabajo.

En el primer caso, es importante documentar las historias de usuario, las fuentes de datos y demás hallazgos relevantes para el proyecto. Es importante luego de haber establecido las historias de usuario transformarlas en requerimientos.

En el segundo caso, al construir un data mart para suplir una necesidad en datos ya existente, se recomienda realizar un diagrama de notación bajo estándares BPMN para describir mejor los procesos de carga de datos actual en el área, y al igual que en el primer caso establecer las historias de usuario y requerimientos.

Priorización de requerimientos y establecimiento del alcance del proyecto

Hay diversos métodos de priorización, como la tabla de priorización, análisis costo-beneficio, por votación del equipo, etc. Se debe escoger uno para luego, documentar esta priorización para que, una vez filtrado el primer documento de requerimientos, este pueda ser acordado con el usuario y, así finalmente, determinar el alcance del proyecto en cuanto a cumplimiento de requerimientos.

Aceptación de requerimientos por parte del usuario

Posterior a la documentación de los requerimientos, es importante acordar con el equipo de trabajo y el área usuaria cómo se hará la revisión de los requerimientos y aceptación de ellos por ambas partes, para asegurar que estos han sido correctamente comprendidos y recopilados.

Fase: Arquitectura técnica DM

Para diseñar la arquitectura técnica de un proyecto, es importante conocer las necesidades del negocio y las implicaciones arquitectónicas que este tiene. Se deben hacer reuniones con el equipo de TI para conocer de qué herramientas se dispone actualmente, también es importante aprender de proyectos anteriores y saber si la organización está dispuesta a hacer cambios operativos para lograr los objetivos del proyecto.

Un resultado de esta fase podría ser un diagrama con los elementos de la arquitectura técnica añadiéndole la documentación a cada uno de esos elementos previamente mencionados y que se amplían a continuación.

- Fuentes de extracción: son aquellas fuentes de las que se podrán extraer datos que genera la misma área o que provendrán de fuentes internas o externas a la organización. Estas incluyen: archivos de texto, hojas de cálculo, bases de datos transaccionales, etc.
- Transformación: para documentar esta fase es importante dar una descripción de las herramientas técnicas usadas en el proceso de limpieza, verificación de calidad de datos y la modificación al modelado de datos en caso de haberla.
- Carga: para este proceso se describen las herramientas usadas para la carga de datos transformados en una base de datos data mart.
- Tablero de inteligencia de negocios: se documenta la selección de la herramienta BI y cómo se conectará esta al *data mart*.

Fase: Herramientas

Se deben elegir las herramientas de creación y mantenimiento de un *data mart*. Luego, se deben instalar las herramientas y capacitarse.

Nota: Tenga en cuenta que la selección de estas herramientas se hace en base a los requisitos de cada área usuaria interesada y la disponibilidad de estas en la empresa, por lo cual la selección de estas puede variar por proyecto.

Fase: Gestión de la Metadata

La gestión de la metadata es el proceso de gestionar y documentar los metadatos de aquellos datos que estarán almacenados en el *data mart*. Los metadatos incluyen información adicional sobre los datos, por ejemplo, significado, origen, formato, rango, etc; es por ello por lo

que la documentación y gestión de los metadatos es de sumo valor para el proyecto, pues le permite al usuario del *data mart* conocer qué datos están disponibles, en qué formato, con qué procedencia y demás. Esto es útil para tener un fundamento al tomar decisiones sobre cuáles operaciones son necesarias en procesos de extracción, transformación y limpieza.

Algunos estándares y procedimientos para tener en cuenta en esta fase son:

- Procedimientos de documentación: establecimiento de procedimientos para documentar clara y detalladamente los metadatos. Establecer un formato para dejar registro de campos, tipos de datos, tamaño, valores de ejemplo, cálculos, etc.
- Estándares de calidad de datos: definir estándares para la calidad de los datos en el *data mart*. Entre estos estándares se encuentra la validación de datos, la detección de valores atípicos, la consistencia de datos, y la integridad referencial.
- Interoperabilidad: establecer procedimientos para documentar los procesos de interoperabilidad del *data mart*, especificando qué herramientas de inteligencia de negocios se alimentan de él.
- Control de cambios: determinar qué procedimientos se llevarán a cabo para asegurar la documentación de los cambios realizados en el *data mart*, quiénes aprobarán esos cambios y qué pruebas se harán antes de la implementación en producción.

Fase: Diseño del modelo dimensional

El modelado dimensional es un proceso de modelado bajo el cual se construyen diversos esquemas para hacer las bases de datos simples. Bajo la metodología Kimball el esquema escogido es el de estrella. A continuación, se explican los pasos a seguir durante esta etapa según Kimball:

1. Elegir el proceso de negocio: en este paso se debe elegir un proceso del conjunto de procesos priorizados seleccionados en la actividad de priorización de requerimientos y

establecimiento del alcance del proyecto. Elegir el proceso de negocio depende fundamentalmente del análisis de los requerimientos.

2. Establecer el nivel de granularidad: esto implica especificar el detalle al que se llegarán en cada una de las entidades, es decir, qué tantos campos de cada entidad se llevarán al *data mart*. Esta elección depende de las necesidades del negocio ya identificadas.
3. Elegir las dimensiones: las dimensiones son características específicas de los datos que se utilizan para analizar y entender mejor la información almacenada en un almacén de datos. En este paso se escogen el conjunto de campos que darán una perspectiva sobre la tabla de hechos.
4. Identificar medidas y las tablas de hechos: Después de haber identificado las dimensiones, el siguiente paso es identificar las medidas y las tablas de hechos. Una medida es un campo de una tabla que se quiere analizar mediante diferentes operaciones (sumas, promedios, mínimos, etc.). Estas medidas dependerán de lo escogido en el nivel de granularidad y serán aquellos campos que se guarden dentro de la misma tabla de hechos.
5. Identificar técnicas de *Slowly Changing Dimension Techniques* (SCD): una SCD es una técnica que es parte del proceso *Change Data Capture* (CDC) que es usado para llevar registro de los cambios en las columnas a lo largo del tiempo.
6. Realizar un modelo gráfico: después de identificar las tablas de hechos, se debe realizar un modelo gráfico para identificar de una manera sencilla qué tablas se relacionarán con la tabla de hechos.
7. Identificación de atributos de dimensiones y tablas de hechos: en este punto se añade al diseño de cada tabla, una lista de campos. Se puede graficar una grilla con los atributos de cada tabla y demás características relevantes.

8. Revisión y validación del modelo: el equipo de trabajo debe revisar el modelo, realizando un estudio del producto que se ha obtenido, además de poder validarlo.
9. Documentos finales: de todos los pasos anteriores verificar que en la documentación se hayan incluido los siguientes puntos.
 - Modelado de datos de alto nivel.
 - Atributos de las tablas de hechos y de las dimensiones.
 - Diagrama de las tablas de hechos y de dimensiones (UML de los esquemas estrella).

Fase: Diseño físico de la base de datos

El diseño del modelo dimensional de la fase anterior debe trasladarse a una base de datos física, es por eso por lo que la fase de diseño físico de la base de datos es tan importante. En esta fase, se resolverán tareas que incluyen la definición de estándares para objetos físicos y desarrollar el modelo de base de datos física. A continuación, se explica con base a recomendaciones en Kimball los pasos a seguir en esta fase.

- 1. Definición de estándares para nombres y bases de datos:** el equipo se reúne para establecer un estándar para las nomenclaturas. Es importante definir si se usará un estándar ya definido por la empresa o si el equipo usará su propia nomenclatura (camel case, notación húngara, *under_score*, *snake_case*, etc.).

También es importante aclarar en qué situaciones serán válidos los valores null o qué tratamiento se les hará para poder identificarlos.

- 2. Desarrollar modelo de base de datos física:** este modelo incluye las tablas que serán utilizadas, tales como *staging area*, las tablas del propio *data mart*, las tablas de auditoría, los esquemas en la base de datos, etc. En esta fase también se establecen las relaciones entre las diferentes tablas, así como las restricciones de integridad referencial, tales como

claves primarias y foráneas. También se definen los tipos de datos y las longitudes de los campos, se establecen los índices necesarios para optimizar el rendimiento de la base de datos. Además, se determina la ubicación física de la base de datos.

- 3. Indexación:** definir las estrategias de indexación adecuadas para mejorar el rendimiento y optimizar las consultas en el data mart. Se deben definir los índices necesarios para cada tabla, considerando la frecuencia de acceso y las consultas que se realizarán con mayor frecuencia. También es importante evaluar el impacto de los índices en el rendimiento de carga de datos y en el espacio de almacenamiento necesario para la base de datos.

Fase: Implementación física de la base de datos

La implementación física de la base de datos implica la creación y configuración de la base de datos en los servidores de producción. En esta fase se llevan a cabo las tareas de creación de tablas, índices, procedimientos o funciones almacenadas. Además, se configura la base de datos para su correcto funcionamiento, se configuran los permisos de acceso, esto para posteriormente realizar los procedimientos ETL para cargar de datos al *data mart*, y más adelante hacer pruebas de calidad.

Fase: Diseño del sistema de extracción transformación y carga

Esta fase de diseño implica diferentes actividades: diseño del flujo e integración de los datos, definición de procedimientos de carga y creación de los scripts que extraerán los datos de las fuentes y actualizarán los datos en la base de datos destino data mart.

Fase: Desarrollo del sistema ETL

En esta fase de desarrollo o implementación del sistema ETL se crean los procesos en la infraestructura necesaria para extraer, transformar y cargar los datos al data mart. Se debe definir el flujo de datos, establecer lógicas de transformación, reglas de carga y actualización del data

mart. Esta fase implica la creación de los paquetes de ETL que contendrán el proceso ETL, y la programación de la ejecución de los *Jobs*.

Fase: *Testing* y mantenimiento

En esta fase se llevan a cabo diferentes tipos de pruebas: pruebas unitarias para verificar la funcionalidad de cada componente de manera aislada, pruebas de integración para comprobar la correcta interacción entre los diferentes componentes y pruebas de aceptación para validar el sistema ETL en su conjunto.

Después de que el *data mart* ha sido implementado y puesto en operación, se pueden llevar a cabo pruebas de monitoreo y rendimiento para asegurarse de que está funcionando de manera adecuada y cumpliendo con los requisitos del negocio. Estas pruebas incluyen:

- Pruebas de rendimiento: Estas pruebas se realizan para medir el rendimiento del *data mart* en términos de velocidad, capacidad y escalabilidad.
- Pruebas de seguridad: Se realizan para verificar que los datos del *data mart* están protegidos y asegurar que sólo los usuarios o roles autorizados tienen acceso a la información.
- Pruebas de integridad de datos: Estas pruebas se realizan para asegurarse de que los datos almacenados en el *data mart* son precisos y coherentes.