

**UNIVERSIDAD DE EL SALVADOR
FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA DE INGENIERÍA DE SISTEMAS INFORMÁTICOS**



**PROTOTIPO DE DISEÑO E IMPLEMENTACIÓN DE UN MODELO
DIMENSIONAL PARA EL PROCESO DE NEGOCIO DE INFORMACIÓN
FINANCIERA DE LA EMPRESA “INVERSIONES DIGITALES”**

PRESENTADO POR:

AQUINO ALAS, DORIS ELIZABETH
GÓMEZ EVORA, FRANCISCO EVERARDO

PARA OPTAR AL TÍTULO DE:

INGENIERO(A) DE SISTEMAS INFORMÁTICOS

CIUDAD UNIVERSITARIA, NOVIEMBRE 2025

UNIVERSIDAD DE EL SALVADOR

RECTOR:

M.SC. JUAN ROSA QUINTANILLA

SECRETARIO GENERAL:

LIC. PEDRO ROSALÍO ESCOBAR CASTANEDA

FACULTAD DE INGENIERÍA Y ARQUITECTURA

DECANO:

ING. LUIS SALVADOR BARRERA MANCÍA

SECRETARIO:

ING. JULIO ALBERTO PORTILLO

ESCUELA DE INGENIERÍA DE SISTEMAS INFORMÁTICOS

DIRECTOR:

ING. CÉSAR AUGUSTO GONZÁLEZ

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA DE INGENIERÍA DE SISTEMAS INFORMÁTICOS

Trabajo de Graduación previo a la opción al Grado de:

INGENIERO(A) DE SISTEMAS INFORMÁTICOS

Título:

**PROTOTIPO DE DISEÑO E IMPLEMENTACIÓN DE UN MODELO
DIMENSIONAL PARA EL PROCESO DE NEGOCIO DE
INFORMACIÓN FINANCIERA DE LA EMPRESA “INVERSIONES
DIGITALES”**

Presentado por:

**AQUINO ALAS, DORIS ELIZABETH
GÓMEZ EVORA, FRANCISCO EVERARDO**

Trabajo de Graduación Aprobado por:

Docente Asesor:

ING. RENE FABRICIO QUINTANILLA GÓMEZ

SAN SALVADOR, NOVIEMBRE 2025

Trabajo de Graduación Aprobado por:

Docente Asesor:

ING. RENE FABRICIO QUINTANILLA GÓMEZ

INDICE

1.	Introducción	1
2.	Objetivos	2
a)	Objetivo general.....	2
b)	Objetivos específicos.....	2
3.	Marco teórico.....	3
a)	Empresa.....	3
b)	Big Data	4
c)	OLAP (Online Analytical Processing)	6
d)	Data Warehouse.....	8
e)	Modelo de Inmon.....	10
f)	Modelo de Kimball	11
4.	Capítulo I: Especificación del proyecto	12
a)	Situación actual	12
i)	Antecedentes	12
ii)	Descripción del problema	21
iii)	Planteamiento del problema	22
b)	Alcances.....	24
c)	Justificación	24
d)	Diagrama del data set	25
e)	Diccionario de datos del data set.....	27
f)	Data profiling.....	46
5.	Capítulo II: Análisis y diseño de la propuesta de solución	65
a)	Metodología de trabajo	65
b)	Descripción de la propuesta de solución	66
c)	Descripción de la tecnología a utilizar.....	68
d)	Diagrama arquitectónico de la solución.....	73
e)	Descripción de cada componente de la solución.....	74
6.	Capítulo III: Estrategia de implementación de la propuesta de solución	90
a)	Estrategia de implementación	90
b)	Presupuesto de implementación	104
c)	Análisis de resultados.....	107
7.	Conclusiones.....	112

8.	Recomendaciones	113
9.	Bibliografía	114
10.	Glosario	115
11.	Anexos	116
	Anexo 1: Cronograma de actividades	116
	Anexo 2: Presupuesto	117
	Anexo 3: Script de generación de tablas en RedShift	119

INDICE DE TABLAS

Tabla 1	Descripción de datos del data set. Tabla: account_account	28
Tabla 2	Descripción de datos del data set. Tabla: account_account_tag	29
Tabla 3	Descripción de datos del data set. Tabla: account_account_account_tag	29
Tabla 4	Descripción de datos del data set. Tabla: account_move_line	32
Tabla 5	Descripción de datos del data set. Tabla: dw_account_order	34
Tabla 6	Descripción de datos del data set. Tabla: dw_account_sum	35
Tabla 7	Descripción de datos del data set. Tabla: res_company	45
Tabla 8	Data profiling. Análisis numérico. Tabla: account_account	46
Tabla 9	Data profiling. Análisis String. Tabla: account_account	47
Tabla 10	Data profiling. Análisis fecha/tiempo. Tabla: account_acccount	48
Tabla 11	Data profiling. Análisis booleanos. Tabla: account_account	48
Tabla 12	Data profiling. Análisis numérico. Tabla: account_account_account_tag	49
Tabla 13	Data profiling. Análisis String. Tabla: account_account_account_tag	50
Tabla 14	Data profiling. Análisis fecha/tiempo. Tabla: account_account_account_tag	51
Tabla 15	Data profiling. Análisis booleanos. Tabla: account_account_account_tag	51
Tabla 16	Data profiling. Análisis numérico. Tabla: account_move_line	53
Tabla 17	Data profiling. Análisis String. Tabla: account_move_line	54
Tabla 18	Data profiling. Análisis fecha/tiempo. Tabla: account_move_line	54
Tabla 19	Data profiling. Análisis booleanos. Tabla: account_move_line	54
Tabla 20	Data profiling. Análisis numérico. Tabla: dw_account_order	55

Tabla 21 Data profiling. Análisis String. Tabla: dw_account_order	56
Tabla 22 Data profiling. Análisis fecha/tiempo. Tabla: dw_account_order.....	56
Tabla 23 Data profiling. Análisis booleanos. Tabla: dw_account_order.....	57
Tabla 24 Data profiling. Análisis numérico. Tabla: dw_account_sum	57
Tabla 25 Data profiling. Análisis String. Tabla: dw_account_sum	58
Tabla 26 Data profiling. Análisis fecha/tiempo. Tabla: dw_account_sum.....	59
Tabla 27 Data profiling. Análisis booleanos. Tabla: dw_account_sum.....	59
Tabla 28 Data profiling. Análisis numérico. Tabla: res_company	61
Tabla 29 Data profiling. Análisis String. Tabla: res_company	63
Tabla 30 Data profiling. Análisis fecha/tiempo. Tabla: res_company.....	64
Tabla 31 Matriz de bus entre la tabla de hechos y sus dimensiones.....	67
Tabla 32 Presupuesto de Implementación. Recurso humano	105
Tabla 33 Presupuesto de implementación. Equipo informático, Hardware.....	105
Tabla 34 Presupuesto implementación, Licencias	106
Tabla 35 Presupuesto implementación, servicios AWS	106
Tabla 36 Presupuesto de implementación.....	106
Tabla 37 Presupuesto de implementación solución planteada, Recurso humano.....	117
Tabla 38 Presupuesto de implementación solución planteada, Equipo informático	117
Tabla 39 Presupuesto implementación solución planteada, servicios AWS.....	118
Tabla 40 Presupuesto de implementación solución planteada.	118

INDICE DE FIGURAS

Figura 1 Proceso OLAP	8
Figura 2 Arquitectura de un Data Warehouse	9
Figura 3: Enfoque de sistemas para el planteamiento del problema	22
Figura 4 Diagrama del data set.....	26
Figura 5 Modelo dimensional propuesto	67
Figura 6 Logo de Odoo	68
Figura 7 Logo de PostgreSQL.....	69
Figura 8 Logo de DBeaver.....	70

Figura 9 Logo de DataCleaner	70
Figura 10 Logo de Talend Open Studio	71
Figura 11 Logo de Amazon S3	72
Figura 12 Logo de Amazon RedShift	72
Figura 13 Diagrama arquitectónico de la solución – Arquitectura de Ralph Kimball	73
Figura 14 Conexiones de metadatos dentro de Talend Open Studio	75
Figura 15 Conexión de Talend Open Studio hacia la base de datos de PostgreSQL	76
Figura 16 ETL de extracción de información. res_company	77
Figura 17 tDBInput-Extracción de los datos	77
Figura 18 tFileOutputDelimited -Extracción de los datos	78
Figura 19 ETL limpieza y transformación. account_account_tag.....	78
Figura 20 tMap, proceso de transformación account_account_tag.....	79
Figura 21 ETL Cálculo y preparación de datos. DimCuenta.....	80
Figura 22 tMap DimCuenta	80
Figura 23 tMap ejemplo de Inner Join. DimCuenta	81
Figura 24 Tmap-Ejem. Expresión.....	81
Figura 25 ETL-Descarga de archivo de S3.....	82
Figura 26 tS3Get.....	82
Figura 27 Job para la descarga de archivos de la Zona Process para la DimCuenta	83
Figura 28 ETL de carga de archivo a S3	83
Figura 29 tS3Connection para subir a S3	84
Figura 30 tS3Put para S3	84
Figura 31 ejemplo de Job que sube todos los archivos a la zona RAW en S3	84
Figura 32 Comando Copy en RedShift.....	85
Figura 33 Estructura de carpetas dentro de S3.....	86
Figura 34 Repositorio en S3	86
Figura 35 RAW Zone en S3	87
Figura 36 Stage Zone en S3	87
Figura 37 Access Zone en S3	88
Figura 38 Modelo dimensional en Redshift	88
Figura 39 Grupos de usuarios en S3.....	89
Figura 40 Usuarios en S3	89

Figura 41 Power BI . Conexión a servidor RedShift	90
Figura 42 Power BI, ingreso de credenciales de RedShift para obtener información.	90
Figura 43 Descarga de Talend Open Studio	92
Figura 44 Formulario de registro para descarga de Talend Open Studio	93
Figura 45 Espacio de trabajo de Taslend Open Studio.....	94
Figura 46 Creación de cuenta en AWS	94
Figura 47 Inicio de registro para creación de cuenta AWS	95
Figura 48 Ingreso de contraseña para creación de cuenta AWS	95
Figura 49 Formulario de ingreso de datos de contacto para creación de cuenta AWS.....	96
Figura 50 Tablero de AWS básico.....	97
Figura 51 Botón para creación de bucket en S3.....	97
Figura 52 Creación de bucket en s3, seccion 1	98
Figura 53 Creación de bucket en s3, seccion 2	99
Figura 54 lista de buckets en S3	99
Figura 55 Componente de Amazon RedShift en AWS.....	100
Figura 56 Creación de Cluster de RedShift.....	101
Figura 57 Redshift - Modificar la configuración de acceso público.....	101
Figura 58 Redshift - Habilitar acceso público	102
Figura 59 Seleccionar servicio Amazon IAM	102
Figura 60 Menú para acceder a la descarga de PowerBI	103
Figura 61 Botón para descarga de PowerBI	103
Figura 62 Pantalla de inicio de PowerBI	104
Figura 63 Reporte financiero sobre el beneficio neto	107
Figura 64 Reporte financiero sobre el índice de retorno de inversión (ROI)	108
Figura 65 Reporte financiero sobre el índice de liquidez.....	109
Figura 66 Reporte financiero sobre tasa de crecimiento en utilidades por período	110
Figura 67 Reporte financiero sobre índice de eficiencia de cobros	111
Figura 63 Cronograma de actividades.....	116

1. Introducción

En el dinámico mundo empresarial actual, la toma de decisiones estratégicas basadas en datos precisos y oportunos se ha convertido en un pilar fundamental para el éxito y la competitividad. En este contexto, la empresa Inversiones Digitales, especializada en la fabricación de sillas y escritorios de oficina de alta calidad, reconoce la importancia de maximizar la eficiencia en la gestión financiera. Con el objetivo de potenciar la inteligencia de negocios y facilitar la toma de decisiones informadas, este documento explora el desarrollo de un Data Warehouse Financiero diseñado específicamente para las necesidades únicas de Inversiones Digitales.

A lo largo de las siguientes secciones, exploraremos detalladamente los objetivos del proyecto, desde el objetivo general hasta los específicos, destacando cómo este Data Warehouse financiero abordará las necesidades particulares de Inversiones Digitales. Además, se presentará un marco teórico para contextualizar la importancia y las aplicaciones prácticas de esta solución en el entorno financiero.

El Capítulo I expondrá al lector la situación actual de la empresa, explorando sus antecedentes y describiendo problemas encontrados. Este análisis servirá como base para establecer los objetivos, alcances, justificación, cronograma de actividades y presupuesto del proyecto, los cuales se detallarán exhaustivamente.

En el Capítulo II, se adentrará en el análisis y diseño de la propuesta de solución. Desde la metodología de trabajo hasta la descripción tecnológica y el diagrama arquitectónico, este capítulo proporcionará una visión completa de la estructura propuesta, asegurando una comprensión clara de cómo la solución abordará los desafíos financieros identificados.

El Capítulo III delinearé la estrategia de implementación, presentando un presupuesto específico y un análisis previsto de los resultados. Concluirá con un análisis de las lecciones aprendidas y recomendaciones para futuras implementaciones.

En resumen, este documento representa un paso significativo hacia la mejora de la gestión financiera en Inversiones Digitales, proporcionando una guía detallada para el desarrollo e implementación exitosa de un Data Warehouse Financiero.

2. Objetivos

a) Objetivo general

Diseñar tableros informativos basados en las métricas establecidas para el análisis de datos y toma de decisiones en la empresa Inversiones Digitales por medio de la implementación de un Data Warehouse financiero.

b) Objetivos específicos

1. Analizar la base de datos transaccional identificando las tablas, relaciones asociadas y entidades que sean relevantes ligados a los procesos principales que alimentan la información financiera de la empresa “Inversiones Digitales”.
2. Elaborar el modelo de Data Set tomando un conjunto de tablas de la base de datos asegurando una estructura óptima que sea un insumo fundamental para la construcción del Data Warehouse.
3. Construir un Data Warehouse que brinde una solución efectiva a la problemática presentada por la empresa Inversiones Digitales.
4. Crear por medio del software Talend Open Studio una estructura de carpetas adecuada para generar los ETL para carga de datos, ETL para transformación de datos, ETL para construcción de Data Warehouse y ETL para descarga de datos.
5. Diseñar una estructura de carpetas en Amazon S3 que permita almacenar los datos obtenidos de los diferentes ETL generados.
6. Implementar el modelo dimensional en el servicio de Redshift de AWS.
7. Consumir los datos desde Amazon S3 para que éstos se almacenen en la estructura de modelo dimensional de Redshift.
8. Elaborar a través de la herramienta Power BI los tableros necesarios para el análisis de las métricas establecidas para dar solución a la problemática de la empresa Inversiones Digitales.

3. Marco teórico

a) Empresa

Una empresa es una entidad organizativa con el objetivo de cubrir ciertas necesidades en un sector en concreto y obtener beneficios a partir de ello. El nacimiento de una compañía surge cuando la solución a una necesidad es materializada por medio de un concepto previamente planificado¹.

La esencia de cualquier empresa está condicionada por el objetivo establecido, el cual suele ser objeto de un análisis detenido antes de formalizar el inicio de la actividad empresarial. Aunque los objetivos de un proyecto pueden experimentar modificaciones a medida que se profundiza en el estudio del mercado, ya que la investigación continua puede revelar novedades que justifiquen cambios significativos en la dirección estratégica del desarrollo empresarial.

Características de una empresa:

- Satisfacen una necesidad.
- Buscan un beneficio económico.
- Poseen recursos humanos.
- Disponen de capital.
- Variedad de perfiles profesionales.
- Son un equipo de trabajo.
- Diversidad de departamentos.
- Trabajan para cumplir objetivos.

Una empresa se compone de diversos elementos que colaboran para su funcionamiento como una entidad productiva en la sociedad. Esta estructura se divide en los siguientes componentes:

Personal:

Comprende a los empleados, quienes desempeñan actividades específicas previamente asignadas y para las cuales fueron contratados. Cada individuo cumple un papel crucial dentro de la compañía, contribuyendo al logro de sus objetivos.

Habilidades:

Representa los conocimientos especializados que cada empleado aporta a la organización. Estas habilidades son la base que impulsa la ejecución eficiente de tareas y actividades diarias, contribuyendo al crecimiento constante de la empresa.

Recursos Materiales:

¹ Chavez, J. (2022, 8 julio). Ceupe. Ceupe. <https://www.ceupe.com/blog/que-es-una-empresa.html>

Involucra los bienes tangibles que la empresa posee y pone a disposición para facilitar la realización de actividades específicas. Estos recursos incluyen maquinaria, herramientas, equipos e instalaciones que respaldan las operaciones diarias de la entidad.

Economía Empresarial:

Refiere al recurso monetario disponible para la empresa, destinado a realizar pagos diversos y llevar a cabo inversiones. La gestión eficiente de la economía empresarial implica la planificación cuidadosa de estos recursos en un marco de tiempo definido, con el objetivo de alcanzar los objetivos estratégicos establecidos.

b) Big Data

Historia del Big Data

Aunque el término "big data" es relativamente reciente, sus raíces se remontan a las décadas de 1960 y 1970, cuando los primeros centros de datos y las bases de datos relacionales sentaron las bases del mundo de los datos.

Alrededor de 2005, se comenzó a tomar conciencia de la abrumadora cantidad de datos generados por usuarios en plataformas como Facebook y YouTube. Fue en este mismo año cuando se introdujo Hadoop, un marco de código abierto diseñado específicamente para almacenar y analizar grandes conjuntos de datos. Este período también marcó el ascenso de la popularidad de las bases de datos NoSQL.

La evolución de marcos de código abierto, como Hadoop (y más recientemente, Spark), desempeñó un papel crucial en el crecimiento del big data, simplificando su manejo y reduciendo los costos de almacenamiento. En los años subsiguientes, el volumen de big data experimentó un crecimiento exponencial. Aunque los usuarios continuaron generando cantidades masivas de datos, el panorama se diversificó con la llegada del Internet de las cosas (IoT), donde objetos y dispositivos conectados generan información valiosa sobre patrones de uso y desempeño de productos. La aparición del aprendizaje automático contribuyó aún más a la explosión de datos.

A pesar de los avances significativos, el big data está apenas en sus primeras etapas de utilidad. La adopción del Cloud Computing ha ampliado enormemente las posibilidades del big data, ofreciendo escalabilidad elástica y permitiendo a los desarrolladores simplificar clústeres ad hoc para analizar subconjuntos específicos de datos. Asimismo, las bases de datos de gráficos han ganado importancia al presentar grandes volúmenes de datos de manera rápida y completa, facilitando su análisis. En resumen, el big data ha recorrido un largo camino, pero su verdadero potencial apenas comienza a desvelarse.

Definición de Big Data

El término "big data" abarca datos que contienen una mayor variedad y que se presentan en volúmenes crecientes y a una velocidad superior. Esto también se conoce como "las tres V".

Dicho de otro modo, el big data está formado por conjuntos de datos de mayor tamaño y más complejos, especialmente procedentes de nuevas fuentes de datos. Estos conjuntos de datos son tan voluminosos que el software de procesamiento de datos convencional sencillamente no puede gestionarlos. Sin embargo, estos volúmenes masivos de datos pueden utilizarse para abordar problemas empresariales que antes no hubiera sido posible solucionar².

Acorde a lo mencionado por el McKinsey Global Institute Report los macrodatos o big data se convertirán en una base clave de la competencia, apuntalando nuevas oleadas de crecimiento de la productividad, innovación y excedente del consumidor, siempre y cuando existan las políticas y los facilitadores adecuados³.

En cuanto a “las tres V” el volumen se refiere a la inmensidad de los datos. La cantidad de datos importa. Con big data, tendrá que procesar grandes volúmenes de datos no estructurados de baja densidad. Puede tratarse de datos de valor desconocido, como feeds de datos de Twitter, flujos de clics de una página web o aplicación para móviles, o equipo con sensores. Para algunas organizaciones, esto puede suponer decenas de terabytes de datos. Para otras, incluso cientos de petabytes, la velocidad a la rapidez con la que se generan, es decir, es el ritmo al que se reciben los datos y (posiblemente) al que se aplica alguna acción. La mayor velocidad de los datos normalmente se transmite directamente a la memoria, en vez de escribirse en un disco. Algunos productos inteligentes habilitados para Internet funcionan en tiempo real o prácticamente en tiempo real y requieren una evaluación y actuación en tiempo real, y la variedad a la diversidad de fuentes y tipos de datos, los datos convencionales eran estructurados y podían organizarse claramente en una base de datos relacional. Con el auge del big data, los datos se presentan en nuevos tipos de datos no estructurados. Los tipos de datos no estructurados y semiestructurados, como el texto, audio o video, requieren un preprocesamiento adicional para poder obtener significado y habilitar los metadatos.

¿Cómo funciona el big data?

El funcionamiento del Big Data se basa en tres acciones clave que se realizan de manera secuencial para aprovechar las nuevas perspectivas que este enfoque proporciona, permitiendo así abrir paso a oportunidades y modelos de negocio innovadores.

Integración:

El proceso comienza con la integración de datos provenientes de diversas fuentes y aplicaciones. A diferencia de los métodos convencionales de integración de datos, como el proceso "extraer, transformar y cargar" (ETL), el Big Data requiere estrategias y tecnologías más avanzadas para manejar conjuntos de datos masivos, incluso de terabytes o petabytes. Durante esta fase, es crucial

² ¿Qué es el big data? (s. f.). <https://www.oracle.com/es/big-data/what-is-big-data/>

³ Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: the next frontier for innovation, competition, and productivity. En *McKinsey & Company*. <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation>

incorporar, procesar y formatear los datos de manera que estén disponibles y listos para ser utilizados por los analistas empresariales.

Gestión:

La siguiente acción clave implica la gestión de los datos, que incluye su almacenamiento. La solución de almacenamiento puede residir en la nube, en las instalaciones (on-premises) o utilizar ambos enfoques. La elección del lugar de almacenamiento se puede realizar según las necesidades específicas, y la flexibilidad de la nube se vuelve atractiva al permitir incorporar recursos de manera escalonada según la demanda. En este paso, se pueden aplicar requisitos de procesamiento específicos y utilizar motores de procesamiento según las características de los conjuntos de datos.

Análisis:

La rentabilidad de la inversión en Big Data se logra mediante el análisis y la utilización efectiva de los datos. Aquí es donde los analistas obtienen claridad al realizar análisis visuales de conjuntos de datos diversos. La exploración continua de datos lleva a nuevos descubrimientos, y compartir estos hallazgos con otros se convierte en una parte fundamental del proceso. Además, se promueve la construcción de modelos de datos utilizando técnicas de aprendizaje automático e inteligencia artificial, permitiendo poner los datos a trabajar de manera más efectiva.

c) OLAP (Online Analytical Processing)

“Es un software para realizar análisis multidimensionales a altas velocidades en grandes volúmenes de datos de un almacén de datos, data mart o algún otro almacén de datos unificado y centralizado” (IBM).

En el ámbito de los datos comerciales, es común encontrar conjuntos de datos que abarcan diversas dimensiones, representando varias categorías que permiten desglosar la información para su presentación, seguimiento o análisis. Por ejemplo, al examinar cifras de ventas, estas pueden descomponerse según dimensiones como ubicación (región, país, estado/provincia, tienda), tiempo (año, mes, semana, día), producto (ropa, categorías específicas, marca, tipo), entre otras.

Sin embargo, en un almacén de datos convencional, los conjuntos de datos se organizan en tablas, y cada una de ellas puede estructurar la información únicamente en dos dimensiones a la vez. Aquí es donde entra en juego OLAP (Procesamiento Analítico en Línea, por sus siglas en inglés), que extrae datos de múltiples conjuntos de datos relacionales y los reorganiza en un formato multidimensional. Este enfoque posibilita un procesamiento veloz y análisis detallados, ya que permite explorar y comprender la información desde diversas perspectivas al mismo tiempo. Es una solución que agiliza y enriquece la interpretación de datos, superando las limitaciones de los almacenes de datos tradicionales y brindando mayor flexibilidad en el análisis multidimensional.

Cubo OLAP

El cubo OLAP constituye la piedra angular de la mayoría de los sistemas OLAP, presentando una base de datos multidimensional estructurada en arreglos. Esta infraestructura posibilita el procesamiento

y análisis eficiente de múltiples dimensiones de datos, superando con creces la capacidad de una base de datos relacional convencional.

A diferencia de una tabla de base de datos relacional que se asemeja a una hoja de cálculo, organizando registros en un formato bidimensional de fila por columna, el cubo OLAP introduce una estructura más avanzada. Cada dato, o "hecho", en la base de datos se localiza en la intersección de dos dimensiones, asemejándose a la intersección de región y ventas totales.

Aunque las herramientas de informes basadas en bases de datos relacionales y SQL pueden consultar y analizar datos multidimensionales almacenados en tablas, su rendimiento se ve comprometido con volúmenes de datos crecientes. Además, la labor de reorganizar los resultados para centrarse en diferentes dimensiones resulta laboriosa.

El cubo OLAP aborda estas limitaciones mediante la expansión de la tabla única con capas adicionales, cada una agregando dimensiones suplementarias, generalmente en el siguiente nivel de la "jerarquía de conceptos" de la dimensión. Por ejemplo, la capa inicial del cubo podría organizar las ventas por región, mientras que capas adicionales podrían abordar niveles más detallados como país, estado/provincia, ciudad e incluso tienda específica.

En teoría, un cubo tiene la capacidad de contener un número infinito de capas, y cuando representa más de tres dimensiones, a veces se denomina hipercubo. Además, es factible incluir cubos más pequeños dentro de las capas; por ejemplo, cada capa de la tienda podría contener cubos que clasifican las ventas por vendedor y producto.

En la práctica, los analistas de datos se enfocarán en la creación de cubos OLAP que contengan únicamente las capas necesarias para un análisis específico, garantizando un rendimiento optimizado. Este enfoque estratégico equilibra la flexibilidad del cubo con la eficiencia y relevancia del análisis de datos.

En el contexto del análisis de datos, existen varias operaciones que nos permiten explorar la información de manera más detallada o resumida. Una de estas operaciones es la profundización (Drill-down), que convierte datos menos detallados en datos más detallados. Por ejemplo, si observamos las ventas trimestrales de una organización, podemos profundizar para ver las ventas mensuales.

En contraposición, la consolidación (Roll-up) es la operación que agrega datos en un cubo OLAP subiendo en la jerarquía de conceptos o reduciendo el número de dimensiones. Por ejemplo, podríamos consolidar los datos de ventas a nivel de país en lugar de ciudad.

Otra operación es el desglose (Slice and dice), que crea un subcubo seleccionando una única dimensión del cubo principal. Por ejemplo, podríamos desglosar y resaltar todos los datos del primer trimestre fiscal de una organización.

La rotación (Pivot) es una función que gira la vista del cubo para visualizar una nueva representación de los datos mediante vistas multidimensionales dinámicas. A diferencia de las tablas dinámicas en programas como Excel, las rotaciones de OLAP son más fáciles de utilizar y tienen un tiempo de respuesta más rápido.

A continuación, en la figura 1 se puede ver de manera resumida el proceso OLAP.

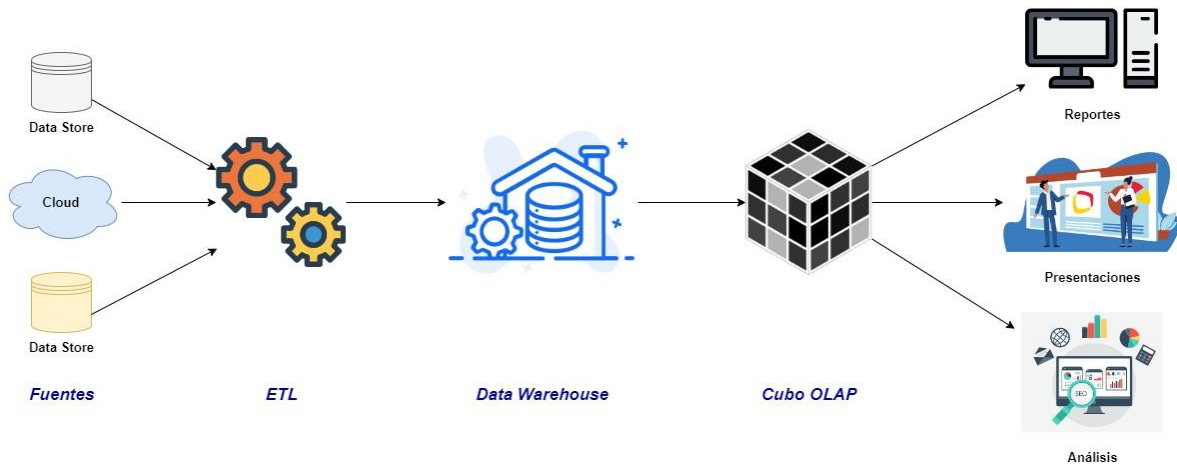


Figura 1 Proceso OLAP

d) Data Warehouse

Un datawarehouse es un almacén de datos, normalmente de gran volumen, cuya estructura facilita el acceso y el análisis de información. Su contenido puede provenir de distintas fuentes de datos estructurados y no estructurados, los cuales son actualizados periódicamente.

La finalidad de la creación de un datawarehouse es contener información precisa, que sirve para llevar a cabo procesos analíticos y la medición de resultados que sean de interés para la entidad que lo implementa.

Un datawarehouse se crea desde una necesidad de obtención de información, desde el principio debe estar orientado a uno o varios procesos, y a los resultados que se desean obtener, todo su desarrollo debe estar regido por el objetivo de su implementación.

El primer paso es el proceso de análisis de la información con la que se cuenta, e identificar su nivel de detalle, con el fin de determinar si es posible obtener los resultados deseados después de su procesamiento. Una vez se realiza este paso se crea un diseño de datawarehouse, usando el modelo y las técnicas que mejor se adecúen a la solución deseada, en la figura 2 se muestra la arquitectura de un Data Warehouse.

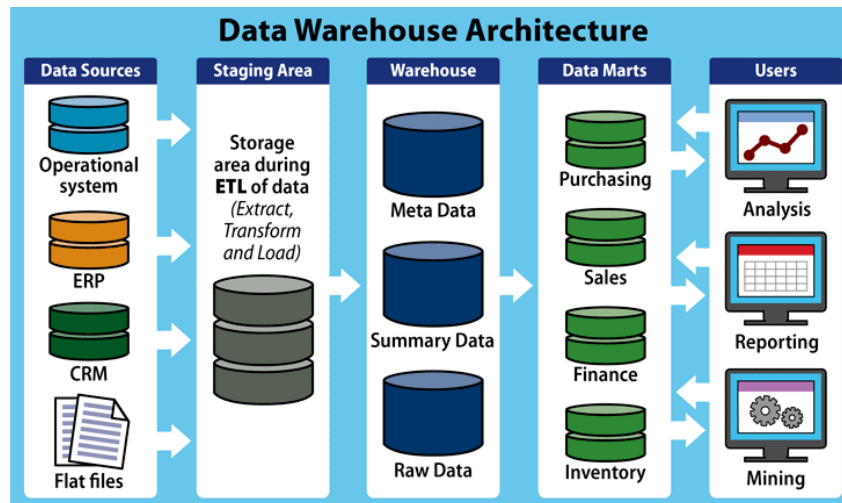


Figura 2 Arquitectura de un Data Warehouse

Componentes

Data Sources: una vez identificados los procesos de negocio que se desean evaluar y los resultados analíticos que se quieren obtener, hay que identificar todas las fuentes de datos que contienen información relevante para realizar el ejercicio, estas pueden ser sistemas transaccionales, bases de datos relacionales y no relacionales, archivos de datos en cualquier formato, etc.

Staging Area: cuando se obtienen los datos requeridos, es necesario empezar un proceso de transformación de dicha información, por medio de procesos ETL (Extracción, Transformación y Carga), estos comprenden actividades como la extracción, limpieza de datos, conversión de formatos y la adecuación al modelo del datawarehouse a implementar, entre otras.

Warehouse: este componente hace referencia al espacio donde se aloja la información en sus diferentes fases, se secciona la información según el proceso donde se encuentra, las 3 áreas más comúnmente usadas son:

- **Raw data:** contiene la información en estado puro como se obtiene de los sistemas fuentes.
- **Summary data:** en esta área se contiene solo la información relevante y se omite lo que no será útil en el modelo del datawarehouse.
- **Meta data:** aquí se almacena la información relevante ajustada al modelo creado, y lista para ser usada en procesos analíticos.

Data Marts: este componente es opcional a utilizar, en caso de querer segmentar la información de acuerdo a los diferentes procesos o departamentos que maneje la empresa, pero provee una mejor organización de los datos.

Users: finalmente los usuarios que tengan acceso al datawarehouse pueden realizar procesos analíticos con los datos, elaborar reportes por medio de aplicaciones de inteligencia de negocio e incluso esos datos pueden seguir siendo analizados con el fin de obtener patrones de

comportamiento, trayectorias de información, resultados probabilísticos y estadísticos etc, dando lugar a la minería y a la ciencia de datos.

Características de un datawarehouse

- Contiene datos de distintas fuentes, y esa es una de sus mayores ventajas, el poder integrar todo tipo de información para la obtención de mejores resultados.
- Un datawarehouse está orientado a temas o procesos específicos, no almacena información que no sea relevante para dicho proceso.
- Se debe actualizar periódicamente
- Contiene un histórico de información, un datawarehouse permite no solamente actualizar datos, sino mantener un registro de los cambios que se han dado en distintos períodos de tiempo.
- Permite procesar grandes cantidades de información.

e) Modelo de Inmon

La metodología de Inmon para el diseño de Data Warehouses, desarrollada por William Inmon, es conocida por su enfoque estructurado y orientado a la integración de datos. Esta metodología se basa en la premisa fundamental de construir un Data Warehouse centralizado y único que sirva como fuente única de la verdad para toda la organización. A continuación, se explica la metodología en un formato de párrafos.

La metodología de Inmon comienza con la creación de un almacén de datos central, denominado el Data Warehouse Corporativo. Este almacén actúa como un depósito único de datos, consolidando información de diversas fuentes dentro de la organización. Inmon aboga por un enfoque top-down, donde la construcción del Data Warehouse comienza con la identificación de los procesos comerciales clave y la comprensión de los requisitos empresariales.

El proceso de diseño implica la creación de un modelo de datos relacional que refleje las entidades comerciales y sus relaciones. Este modelo, conocido como el modelo dimensional, se construye utilizando técnicas normalizadas para garantizar la coherencia y la integridad de los datos. Los modelos dimensionales se centran en la representación de las dimensiones clave para el negocio y las medidas asociadas.

Inmon destaca la importancia de la integridad y la calidad de los datos en todo el proceso. Propone la utilización de metadatos detallados para documentar y gestionar la información sobre la procedencia, significado y calidad de los datos almacenados en el Data Warehouse.

Otro componente esencial de la metodología de Inmon es la arquitectura de procesos ETL (Extract, Transform, Load). Este proceso se encarga de extraer datos de diversas fuentes, transformarlos para

que se ajusten al modelo dimensional y cargarlos en el Data Warehouse. La ETL juega un papel crucial en la garantía de la consistencia y la calidad de los datos almacenados.

La metodología de Inmon también abarca la creación de Data Marts, que son subconjuntos especializados del Data Warehouse Corporativo destinados a satisfacer necesidades específicas de departamentos o grupos de usuarios. Estos Data Marts se derivan del almacén central, asegurando consistencia en la información a lo largo de toda la organización.

La metodología de Inmon destaca la importancia de un enfoque centralizado, la comprensión de los procesos comerciales, la calidad de los datos y la utilización de modelos dimensionales para construir un Data Warehouse robusto y alineado con los objetivos estratégicos de la organización.

f) Modelo de Kimball

La metodología de Kimball para Data Warehouse, desarrollada por Ralph Kimball, es una aproximación ampliamente reconocida y utilizada en el diseño e implementación de almacenes de datos. Esta metodología se basa en principios clave que buscan garantizar la efectividad y la utilidad del Data Warehouse para la toma de decisiones en una organización.

En primer lugar, la metodología de Kimball aboga por un enfoque pragmático y basado en proyectos. En lugar de tratar de construir un Data Warehouse completo de una sola vez, propone dividir el proyecto en fases más pequeñas y manejables llamadas "dimensiones". Cada dimensión representa un aspecto clave del negocio y se construye incrementalmente, permitiendo una implementación ágil y adaptativa.

Un componente central de la metodología Kimball es la importancia otorgada a las dimensiones y hechos. Las dimensiones representan las características o categorías de interés (como clientes, productos o fechas), mientras que los hechos son las métricas numéricas que se miden y analizan (como ventas o ingresos). La modelización dimensional, a través de esquemas como estrella y copo de nieve, facilita la estructuración de datos de manera que sea intuitiva y fácil de entender para los usuarios finales.

En el proceso de desarrollo, Kimball destaca la necesidad de la participación activa de los usuarios de negocio. Se fomenta la colaboración estrecha entre los equipos de tecnología y los usuarios para garantizar que el Data Warehouse cumpla con las necesidades reales del negocio. Esto se logra a través de la realización de entrevistas y la identificación clara de requisitos comerciales.

Además, Kimball subraya la importancia de los datos de calidad y de la gestión de metadatos. La integridad y la consistencia de los datos son esenciales para la toma de decisiones informadas, y los metadatos facilitan la comprensión y la interpretación de los datos almacenados en el almacén.

La metodología de Kimball destaca la necesidad de un enfoque iterativo y centrado en el negocio para la construcción de un Data Warehouse. Al dividir el proyecto en dimensiones manejables, centrarse en la colaboración con los usuarios finales, y mantener un enfoque dimensional, la metodología de Kimball ha demostrado ser efectiva en la creación de almacenes de datos que realmente respaldan las necesidades comerciales.

4. Capítulo I: Especificación del proyecto

a) Situación actual

i) Antecedentes

La importancia de la información en el ámbito empresarial es difícil de subestimar; de hecho, constituye el cimiento sobre el cual se construyen decisiones, estrategias y operaciones efectivas. En un entorno empresarial cada vez más complejo y dinámico, la información se erige como un recurso estratégico clave que impulsa la toma de decisiones informadas. Al comprender y analizar datos relevantes, las empresas pueden anticipar tendencias, evaluar el rendimiento pasado y actual, y planificar de manera más efectiva para el futuro.

La información no solo es esencial para el proceso de toma de decisiones, sino que también juega un papel fundamental en la optimización de operaciones internas. La gestión eficiente de datos permite a las organizaciones mejorar la eficiencia de los procesos, reducir costos y mitigar riesgos. Desde la cadena de suministro hasta la gestión de recursos humanos, contar con información precisa y oportuna facilita la toma de medidas correctivas y la implementación de estrategias para mejorar la eficacia operativa.

En el contexto actual de mercados globalizados, la información se convierte en un facilitador crucial para la competitividad empresarial. Aquellas empresas que pueden acceder, analizar y aplicar rápidamente la información relevante están mejor posicionadas para adaptarse a cambios en el entorno empresarial, identificar oportunidades emergentes y enfrentar desafíos de manera proactiva. La información efectiva se convierte así en un impulsor de la innovación, permitiendo a las empresas desarrollar nuevos productos y servicios, mejorar la experiencia del cliente y mantenerse a la vanguardia en sus respectivas industrias.

La gestión exitosa de la información también fortalece las relaciones con los clientes y socios comerciales. Al comprender las necesidades y expectativas de los clientes, las empresas pueden personalizar sus ofertas, mejorar la satisfacción del cliente y construir lealtad a largo plazo. Asimismo, compartir información de manera eficiente con socios comerciales contribuye a la construcción de relaciones sólidas y colaborativas.

Para el desarrollo del presente proyecto se analizarán los datos brindados por la empresa Inversiones Digitales, empresa dedicada a la elaboración de sillas y escritorios de oficina.

La gestión de todos los procesos de la empresa lo realizan a través del software Odoo ERP, plataforma conocida anteriormente como Open ERP o TinyERP, que ofrece consultoría e implementación de ERP, especializada en gestionar los procesos principales de una empresa como lo son por ejemplo compras, ventas, inventario, gastos, gestión de usuarios, permisos y sucursales, este software está escrito en lenguaje Python. El cliente se comunica con éste a través de interfaces XML-RPC y JSON, y se comunica principalmente con la base de datos PostgreSQL.

La empresa cuenta con múltiples tipos de usuario, desde administradores hasta cajeros de punto de venta, que son personalizados según sea necesario. En forma general se dividen en Administradores, Gerentes y empleados de diferentes áreas, cada persona tiene su usuario con correo y contraseña.

Para la obtención de la información financiera, la empresa debe de realizar flujos financieros partiendo desde la definición de su catálogo de cuentas que estarán ligadas a cada movimiento de dinero de la empresa.

Para comenzar se definen las sucursales de la empresa, posteriormente se enlistan los posibles proveedores de materia prima y proveedores de productos, bancos asociados a la empresa que estarán financiando.

Para realizar una compra se realiza una orden de compra a los diferentes proveedores, posteriormente se valida facturando y pagando (al instante o a futuro) el producto recibido. Una vez teniendo la materia prima se puede comenzar el proceso de fabricación de productos, que se realiza utilizando insumos comprados a los proveedores, y así ir abasteciendo el inventario para las diferentes sucursales.

Una vez abastecido el inventario se puede realizar la venta de los productos en el punto de venta, donde se vende a clientes minoritarios o de mayoreo, a los cuales se les puede aplicar descuentos si así lo establecen. Además, se permite la devolución de producto tanto al proveedor como a la tienda siempre bajo restricciones.

Además del flujo normal, se incluyen pagos de planilla, gastos con los que contará la empresa, pagos y cobros de dinero, así como también de impuestos (IVA 13%). En sí posteriormente a todo el flujo de entrada y salida de dinero es donde se entra a lo importante que es el análisis de la información financiera de la empresa y poder apoyar con el diseño del Data Warehouse, cabe destacar que los módulos de Odoó están mayormente relacionados entre sí para poder complementarse. Como dato adicional se proporciona una breve descripción de los principales módulos que provee Odoó para la gestión de la empresa son los siguientes:

Ventas: Este módulo nos ayuda con la optimización del flujo de las ventas en diferentes períodos mediante la automatización, ofreciéndonos el catálogo de productos además de poder realizar las siguientes acciones en dicho módulo:

- Gestión de Clientes:
 - ✓ Registro de Clientes: Permite la creación y gestión de perfiles de clientes, incluyendo información detallada como datos de contacto, historial de compras y preferencias.
 - ✓ Seguimiento de Clientes: Facilita el seguimiento de interacciones con los clientes, como llamadas, reuniones o correos electrónicos.
- Cotizaciones y Pedidos:
 - ✓ Creación de Cotizaciones: Permite la generación de cotizaciones detalladas para los clientes, incluyendo productos, precios y términos.
 - ✓ Registro de Pedidos: Facilita la conversión de cotizaciones a pedidos confirmados con seguimiento del estado de cada pedido.
- Gestión de Productos y Catálogos:
 - ✓ Base de Datos de Productos: Permite mantener un catálogo completo de productos con detalles como descripciones, precios y existencias en tiempo real.
 - ✓ Variantes de Productos: Admite la gestión de variantes de productos, como tallas o colores, simplificando la administración de inventarios.
- Seguimiento de Inventario:

- ✓ Control de Stock: Integra el módulo de ventas con el módulo de inventario, proporcionando un seguimiento preciso de los niveles de stock y evitando la sobreventa o la falta de productos.
- Facturación y Pagos:
 - ✓ Generación de Facturas: Automatiza el proceso de facturación basado en pedidos confirmados.
 - ✓ Seguimiento de Pagos: Facilita el seguimiento de los pagos recibidos y pendientes, proporcionando una visión clara de la salud financiera de las transacciones.
- Automatización y Flujos de Trabajo:
 - ✓ Flujos de Aprobación: Permite la configuración de flujos de trabajo para la aprobación de cotizaciones y pedidos, garantizando un proceso eficiente y consistente.
 - ✓ Recordatorios y Alertas: Facilita la configuración de recordatorios automáticos para seguimientos de ventas y fechas importantes.
- Análisis y Reportes:
 - ✓ Informes de Ventas: Ofrece informes predefinidos y personalizables para analizar el rendimiento de las ventas, las tendencias y la rentabilidad.
 - ✓ Paneles de Control: Proporciona paneles visuales que permiten a los usuarios tener una visión rápida y clara del estado de las ventas y el desempeño del equipo.
- Integración con Otros Módulos:
 - ✓ Integración con Contabilidad: Garantiza una contabilidad precisa al vincular automáticamente las transacciones de ventas con el módulo de contabilidad.
 - ✓ CRM Integrado: Facilita la gestión de relaciones con el cliente al integrar el módulo de ventas con el CRM de Odoo.

Punto de venta: Con este módulo Inversiones Digitales puede administrar sus 5 sucursales con mayor facilidad, cabe mencionar que los movimientos de los productos se registran automáticamente en su stock, obteniendo las estadísticas en tiempo real, consolidando los datos para cada una de las sucursales, se describen a continuación las acciones que es posible realizar en dicho módulo:

- Gestión de Productos:
 - ✓ Permite cargar y gestionar un catálogo completo de productos con detalles como descripciones, imágenes y precios.
 - ✓ Admite la creación rápida de productos durante la transacción.
- Procesamiento de Ventas:
 - ✓ Facilita la creación y finalización rápida de pedidos y facturas directamente desde el punto de venta.
 - ✓ Admite diversas formas de pago, como efectivo, tarjetas de crédito y otros métodos.
- Gestión de Clientes:
 - ✓ Permite la creación y gestión de perfiles de clientes desde el punto de venta, incluyendo detalles como datos de contacto y historial de compras.
- Descuentos y Promociones:
 - ✓ Admite la aplicación de descuentos y promociones directamente en el punto de venta, mejorando la flexibilidad en la gestión de precios.
- Terminal de Punto de Venta Móvil:

- ✓ Permite la movilidad al utilizar dispositivos móviles como tablets para procesar transacciones en cualquier ubicación de la tienda.
- Gestión de Inventarios en Tiempo Real:
 - ✓ Actualiza automáticamente los niveles de inventario a medida que se realizan las ventas, evitando problemas de sobreventa o falta de productos.
- Conexión en Tiempo Real con otros Módulos:
 - ✓ Se integra de manera efectiva con otros módulos de Odoo, como el de Inventarios y Contabilidad, asegurando coherencia en toda la plataforma.
- Personalización de Recibos y Facturas:
 - ✓ Permite personalizar los recibos y facturas que se generan desde el punto de venta, incorporando detalles específicos del negocio.
- Gestión de Cajeros y Sesiones:
 - ✓ Facilita la administración de cajeros y sesiones, lo que es crucial en entornos donde varios usuarios pueden compartir una misma terminal de punto de venta.
- Registro de Movimientos de Caja:
 - ✓ Proporciona un seguimiento detallado de las transacciones de caja, lo que facilita la reconciliación y la auditoría.
- Informes de Punto de Venta:
 - ✓ Ofrece informes detallados sobre las transacciones realizadas en el punto de venta, proporcionando datos valiosos para análisis y toma de decisiones.

Facturación: con este módulo se gestionan contratos, facturas y cotizaciones. Se podrá personalizar las facturas para los clientes a los que irá dirigida detallando el producto, cantidades, precios unitarios, el impuesto a pagar e importes adeudados, como valor agregado dentro de este módulo se encuentra parte de la gestión contable, se enlista a continuación las acciones posibles de este módulo:

- Creación y Gestión de Facturas:
 - ✓ Permite la generación fácil y rápida de facturas tanto para clientes como para proveedores.
 - ✓ Admite la personalización de plantillas de factura según las necesidades del negocio.
- Control de Pagos:
 - ✓ Facilita el seguimiento detallado de los pagos recibidos.
 - ✓ Permite registrar pagos parciales o completos y proporciona una visión clara de las cuentas por cobrar.
- Facturación Automática:
 - ✓ Automatiza la generación de facturas recurrentes, como suscripciones mensuales o contratos de servicio.
 - ✓ Mejora la eficiencia al eliminar la necesidad de generar manualmente facturas repetitivas.
- Gestión de Impuestos:
 - ✓ Permite la configuración de reglas fiscales para adaptarse a la complejidad de los sistemas impositivos locales e internacionales.
 - ✓ Automatiza el cálculo de impuestos en las facturas y proporciona informes precisos.
- Facturación por Proyectos:
 - ✓ Integración con el módulo de Proyectos de Odoo para la facturación de horas de trabajo y otros costos asociados a proyectos específicos.

- ✓ Facilita la alineación precisa de costos con los ingresos generados.
- Registros Contables Automáticos:
 - ✓ Automatiza la generación de registros contables a medida que se crean y validan las facturas.
 - ✓ Ofrece una contabilidad precisa y actualizada en tiempo real.
- Conciliación Bancaria:
 - ✓ Facilita la conciliación bancaria al rastrear las transacciones financieras y conciliarlas con los registros del banco.
 - ✓ Mejora la precisión en la gestión de cuentas y evita discrepancias.
- Notas de Crédito y Débito:
 - ✓ Permite la creación de notas de crédito y débito para corregir errores en las facturas emitidas.
 - ✓ Proporciona flexibilidad en la corrección de transacciones.
- Informes Financieros:
 - ✓ Ofrece una variedad de informes financieros, como balances y estados de pérdidas y ganancias, para proporcionar una visión completa de la salud financiera de la empresa.
 - ✓ Facilita el análisis y la toma de decisiones basada en datos.
- Integración con Otros Módulos:
 - ✓ Se integra sin problemas con otros módulos de Odoo, como Ventas e Inventario, para garantizar coherencia en los datos en toda la plataforma.
- Seguridad y Auditoría:
 - ✓ Proporciona funciones de seguridad avanzadas para garantizar que solo usuarios autorizados tengan acceso a funciones críticas.
 - ✓ Facilita la auditoría al rastrear cambios en las facturas y registros contables.

Módulo de inventario para Inversiones Digitales es de importancia ya que este es un sistema de gestión de almacén, donde se llevará el control de materia prima para sillas y escritorios a elaborar, como también almacenamiento de productos terminados, podemos ver a continuación un listado de acciones que se pueden realizar en este módulo:

- Creación y Gestión de Productos:
 - ✓ Facilita la creación y gestión de productos, permitiendo detalles como nombre, categoría, precio y unidades de medida.
 - ✓ Permite definir variantes de productos para artículos con diferentes atributos, como tamaño o color.
- Control de Stock en Tiempo Real:
 - ✓ Proporciona una visión en tiempo real de los niveles de stock, evitando problemas de sobreventa o falta de productos.
 - ✓ Admite la actualización automática de inventarios con cada transacción.
- Movimientos de Inventario:
 - ✓ Permite registrar movimientos de inventario, como recepciones de productos, traslados entre ubicaciones y ajustes de inventario.
 - ✓ Facilita el seguimiento detallado de todas las transacciones relacionadas con el inventario.
- Gestión de Ubicaciones y Almacenes:

- ✓ Admite la creación de múltiples almacenes y ubicaciones, facilitando la organización y seguimiento de productos en diversas áreas.
- ✓ Permite la asignación de ubicaciones específicas a productos según requisitos operativos.
- Valuación de Inventario:
 - ✓ Ofrece métodos de valuación de inventario, como el costo promedio o el costo FIFO (primero en entrar, primero en salir).
 - ✓ Facilita la determinación precisa del valor total del inventario.
- Transferencias de Stock:
 - ✓ Permite la transferencia eficiente de productos entre almacenes y ubicaciones.
 - ✓ Facilita el seguimiento de la ruta de los productos en todo momento.
- Integración con Ventas y Compras:
 - ✓ Se integra sin problemas con los módulos de Ventas y Compras de Odoo.
 - ✓ Automatiza la actualización de inventarios en función de transacciones de ventas y compras.
- Reservas y Disponibilidad:
 - ✓ Facilita la gestión de reservas de productos para pedidos pendientes o clientes específicos.
 - ✓ Proporciona información sobre la disponibilidad de productos en tiempo real.
- Seguimiento de Lotes y Números de Serie:
 - ✓ Permite el seguimiento preciso de lotes de productos y números de serie.
 - ✓ Es esencial en sectores como la fabricación o la distribución de productos con requisitos específicos de trazabilidad.
- Informes de Inventario:
 - ✓ Ofrece informes detallados sobre la situación actual del inventario, movimientos recientes y valoración.
 - ✓ Facilita la toma de decisiones informada y la planificación de inventarios.
- Auditoría y Seguridad:
 - ✓ Proporciona funciones avanzadas de seguridad para controlar el acceso a funciones críticas.
 - ✓ Facilita la auditoría al rastrear cambios en los niveles de stock y movimientos de inventario.
- Integración con Contabilidad:
 - ✓ Se integra con el módulo de Contabilidad para garantizar la coherencia en la gestión de transacciones financieras relacionadas con inventario.

El módulo de fabricación ayuda a la sucursal matriz a programar, planificar y procesar las órdenes de fabricación desde la etapa de diseño hasta la fase de finalización del producto, a continuación, vemos un listado de acciones que se pueden llevar a cabo en este módulo:

- Gestión de Órdenes de Fabricación:
 - ✓ Facilita la creación y gestión de órdenes de fabricación para productos específicos.
 - ✓ Permite la planificación y programación de las actividades de producción.
- Bill of Materials (BOM) o Lista de Materiales:
 - ✓ Permite definir la estructura detallada de productos, especificando los materiales, componentes y operaciones de ensamblaje necesarios.
 - ✓ Facilita la gestión de variantes de productos y versiones de BOM.
- Planificación de la Producción:
 - ✓ Ofrece herramientas para planificar la producción en función de la demanda del mercado, los niveles de inventario y otros factores.

- ✓ Facilita la asignación de recursos y la programación eficiente de la producción.
- Seguimiento de la Producción en Tiempo Real:
 - ✓ Proporciona visibilidad en tiempo real de las operaciones de fabricación.
 - ✓ Permite el seguimiento de la progresión de las órdenes de fabricación y la identificación de posibles cuellos de botella.
- Gestión de Recursos y Máquinas:
 - ✓ Permite la asignación de máquinas y recursos a operaciones específicas.
 - ✓ Facilita la gestión de la capacidad y la eficiencia de los recursos de producción.
- Control de Calidad:
 - ✓ Integra funciones de control de calidad en el proceso de fabricación.
 - ✓ Permite la definición de puntos de control y la realización de pruebas para garantizar la calidad del producto final.
- Movimientos de Inventarios Automáticos:
 - ✓ Automatiza los movimientos de inventario asociados con la producción.
 - ✓ Actualiza automáticamente los niveles de inventario a medida que se fabrican productos.
- Gestión de Costos de Producción:
 - ✓ Permite el seguimiento detallado de los costos asociados con la producción, incluyendo materiales, mano de obra y gastos generales.
 - ✓ Facilita la evaluación de la rentabilidad de las operaciones de fabricación.
- Asociación con Ventas y Compras:
 - ✓ Se integra con los módulos de Ventas e Inventarios para garantizar la coherencia en la gestión de transacciones relacionadas con la producción.
 - ✓ Facilita la actualización automática de inventarios y la sincronización con pedidos de venta.
- Seguimiento de Lotes y Números de Serie:
 - ✓ Permite el seguimiento de lotes de productos y números de serie en el proceso de fabricación.
 - ✓ Esencial para industrias con requisitos específicos de trazabilidad.
- Informes de Producción:
 - ✓ Ofrece informes detallados sobre la eficiencia de la producción, costos y otras métricas clave.
 - ✓ Facilita la toma de decisiones informada y la mejora continua.
- Auditoría y Seguridad:
 - ✓ Proporciona funciones avanzadas de seguridad para controlar el acceso a funciones críticas.
 - ✓ Facilita la auditoría al rastrear cambios en las órdenes de fabricación y registros asociados.
- Personalización de Procesos:
 - ✓ Permite la configuración de procesos de fabricación específicos para adaptarse a las necesidades únicas de la empresa.
 - ✓ Facilita la optimización de operaciones de acuerdo con los métodos de trabajo preferidos.

El módulo de compras ayudará a la organización para apoyar el inventario, realizando órdenes de compra y métodos de reabastecimiento para cada uno de los productos de materia prima que se adquieran, este módulo nos permite realizar las siguientes acciones:

- Creación y Gestión de Órdenes de Compra:
 - ✓ Facilita la creación y gestión de órdenes de compra para la adquisición de bienes y servicios.

- ✓ Permite la personalización de órdenes de compra con detalles como términos de pago y entrega.
- Seguimiento de Proveedores:
 - ✓ Permite el registro y gestión de información detallada sobre proveedores, incluyendo datos de contacto, historial de compras y términos contractuales.
 - ✓ Facilita la evaluación y comparación de proveedores.
- Integración con Inventarios:
 - ✓ Se integra de manera efectiva con el módulo de Inventarios de Odoo.
 - ✓ Automatiza la actualización de inventarios en función de transacciones de compras.
- Solicitudes de Cotización (RFQ):
 - ✓ Facilita la creación de solicitudes de cotización para obtener ofertas de proveedores.
 - ✓ Mejora la transparencia y competitividad en el proceso de adquisición.
- Gestión de Contratos de Compra:
 - ✓ Permite la creación y seguimiento de contratos de compra con proveedores.
 - ✓ Facilita la gestión de acuerdos a largo plazo y condiciones contractuales específicas.
- Control de Costos y Presupuestos:
 - ✓ Permite el control de costos al establecer presupuestos para órdenes de compra.
 - ✓ Facilita la comparación de costos reales con presupuestos planificados.
- Recepción de Productos:
 - ✓ Automatiza el proceso de recepción de productos o servicios conforme a órdenes de compra.
 - ✓ Facilita la reconciliación de entregas con pedidos.
- Facturación Automática:
 - ✓ Permite la automatización de la facturación en función de órdenes de compra recibidas.
 - ✓ Facilita la gestión eficiente de cuentas por pagar.
- Evaluación de Proveedores:
 - ✓ Facilita la evaluación de proveedores basada en criterios como calidad, tiempo de entrega y precios.
 - ✓ Permite la toma de decisiones informada al seleccionar proveedores estratégicos.
- Informes de Compras:
 - ✓ Ofrece informes detallados sobre las transacciones de compras, costos y rendimiento de proveedores.
 - ✓ Facilita la toma de decisiones informada y la planificación estratégica.
- Auditoría y Seguridad:
 - ✓ Proporciona funciones avanzadas de seguridad para controlar el acceso a funciones críticas.
 - ✓ Facilita la auditoría al rastrear cambios en las órdenes de compra y registros asociados.
- Personalización de Flujo de Trabajo:
 - ✓ Permite la personalización del flujo de trabajo para adaptarse a los procesos específicos de adquisición de la empresa.
 - ✓ Facilita la optimización de procesos internos.
- Integración con Contabilidad:
 - ✓ Se integra con el módulo de Contabilidad para garantizar la coherencia en la gestión de transacciones financieras relacionadas con compras.

Recursos Humanos: En este apartado el módulo de empleados será de gran ayuda ya que nos ayuda con el registro de los empleados donde se podrá llevar el control de ciertas actividades como horas de trabajo, permisos, asistencias, y se puede llevar como valor importante la nómina de cada uno de los empleados, se detalla en el siguiente listado lo que se puede realizar en este módulo:

- Gestión de Empleados:
 - ✓ Permite la creación y gestión de perfiles de empleados con detalles como datos personales, historial laboral y documentos asociados.
 - ✓ Facilita el seguimiento del progreso y cambios en el estado del empleado.
- Contratación y Onboarding:
 - ✓ Facilita el proceso de contratación mediante la generación de contratos, la asignación de recursos y la configuración de permisos.
 - ✓ Simplifica el proceso de integración y formación de nuevos empleados.
- Evaluación del Desempeño:
 - ✓ Permite la configuración y seguimiento de evaluaciones de desempeño.
 - ✓ Facilita la identificación de áreas de mejora y el reconocimiento del rendimiento destacado.
- Gestión de Vacaciones y Ausencias:
 - ✓ Automatiza el seguimiento de las solicitudes de vacaciones y ausencias de los empleados.
 - ✓ Facilita la planificación y asignación eficiente de recursos durante períodos de ausencia.
- Administración de Nóminas:
 - ✓ Facilita la gestión completa del ciclo de nómina, desde la configuración de salarios hasta la generación de recibos de sueldo.
 - ✓ Permite la definición de reglas de nómina y la automatización de cálculos.
- Administración de Beneficios:
 - ✓ Permite la configuración y gestión de beneficios para empleados, como seguros de salud y planes de pensiones.
 - ✓ Facilita la comunicación de información relevante sobre beneficios.
- Registro de Tiempos y Asistencia:
 - ✓ Ofrece herramientas para el registro preciso de las horas trabajadas y la asistencia de los empleados.
 - ✓ Permite la generación de informes detallados sobre la gestión del tiempo.
- Gestión de Competencias y Habilidades:
 - ✓ Facilita la creación de perfiles de competencias y la asignación de habilidades a empleados.
 - ✓ Mejora la asignación eficiente de tareas y proyectos basada en habilidades específicas.
- Portal del Empleado:
 - ✓ Proporciona un portal dedicado para que los empleados accedan a información personal, documentos y solicitudes.
 - ✓ Mejora la comunicación interna y la autoadministración de los empleados.
- Integración con Otros Módulos:
 - ✓ Se integra de manera efectiva con otros módulos de Odoo, como Contabilidad y Proyectos.
 - ✓ Garantiza la coherencia de datos en toda la plataforma.
- Seguridad y Privacidad:
 - ✓ Proporciona funciones avanzadas de seguridad para garantizar la confidencialidad de la información del empleado.

- ✓ Cumple con regulaciones de privacidad y protección de datos.
- Informes y Analíticas:
 - ✓ Ofrece informes detallados sobre la gestión de recursos humanos, como costos de personal, eficiencia y rotación de empleados.
 - ✓ Facilita la toma de decisiones informada y la planificación estratégica.
- Gestión de Formación:
 - ✓ Permite la planificación y seguimiento de programas de formación y desarrollo para empleados.
 - ✓ Facilita la mejora continua de habilidades y competencias.

ii) Descripción del problema

La empresa Inversiones Digitales se enfrenta a la necesidad crítica de mejorar su capacidad para analizar y gestionar datos financieros de manera efectiva. Para abordar este desafío se propone la creación de un Data Warehouse financiero que permita medir las siguientes métricas clave:

1. **Beneficio Neto:** Este indicador es esencial para evaluar la rentabilidad global de las operaciones de Inversiones Digitales. El Data Warehouse facilitará el cálculo y seguimiento continuo de este indicador representados por los beneficios que quedan después de restar los costos que van asociados a la venta de productos, proporcionando una visión clara de la salud financiera de la empresa.
2. **Retorno sobre la Inversión (ROI):** Evaluar el rendimiento de las inversiones es fundamental. El Data Warehouse integrará datos de inversión con los resultados obtenidos, ofreciendo una visión completa del ROI. Esto permitirá tomar decisiones más informadas sobre futuras inversiones y estrategias.
3. **Tasa de Crecimiento en Utilidades por Período:** Comprender cómo evolucionan las utilidades a lo largo del tiempo es crucial. El Data Warehouse facilitará el seguimiento de la tasa de crecimiento en utilidades, proporcionando información valiosa para la planificación financiera y la identificación de tendencias a largo plazo.
4. **Liquidez:** Mantener una posición financiera líquida es esencial para la estabilidad. El Data Warehouse integrará datos relevantes sobre activos, pasivos y flujos de efectivo, proporcionando una visión integral de la liquidez y permitiendo una gestión más efectiva de los recursos financieros.
5. **Ratio de eficiencia en cobros:** El análisis de esta métrica por medio del Data Warehouse ayudará a la empresa a medir la eficiencia con la que se están cobrando las cuentas por cobrar, un ratio más alto indica que esto se lleva de manera eficiente.

iii) Planteamiento del problema

A continuación, se presenta en la figura 3 el planteamiento del problema por medio del enfoque de sistemas.

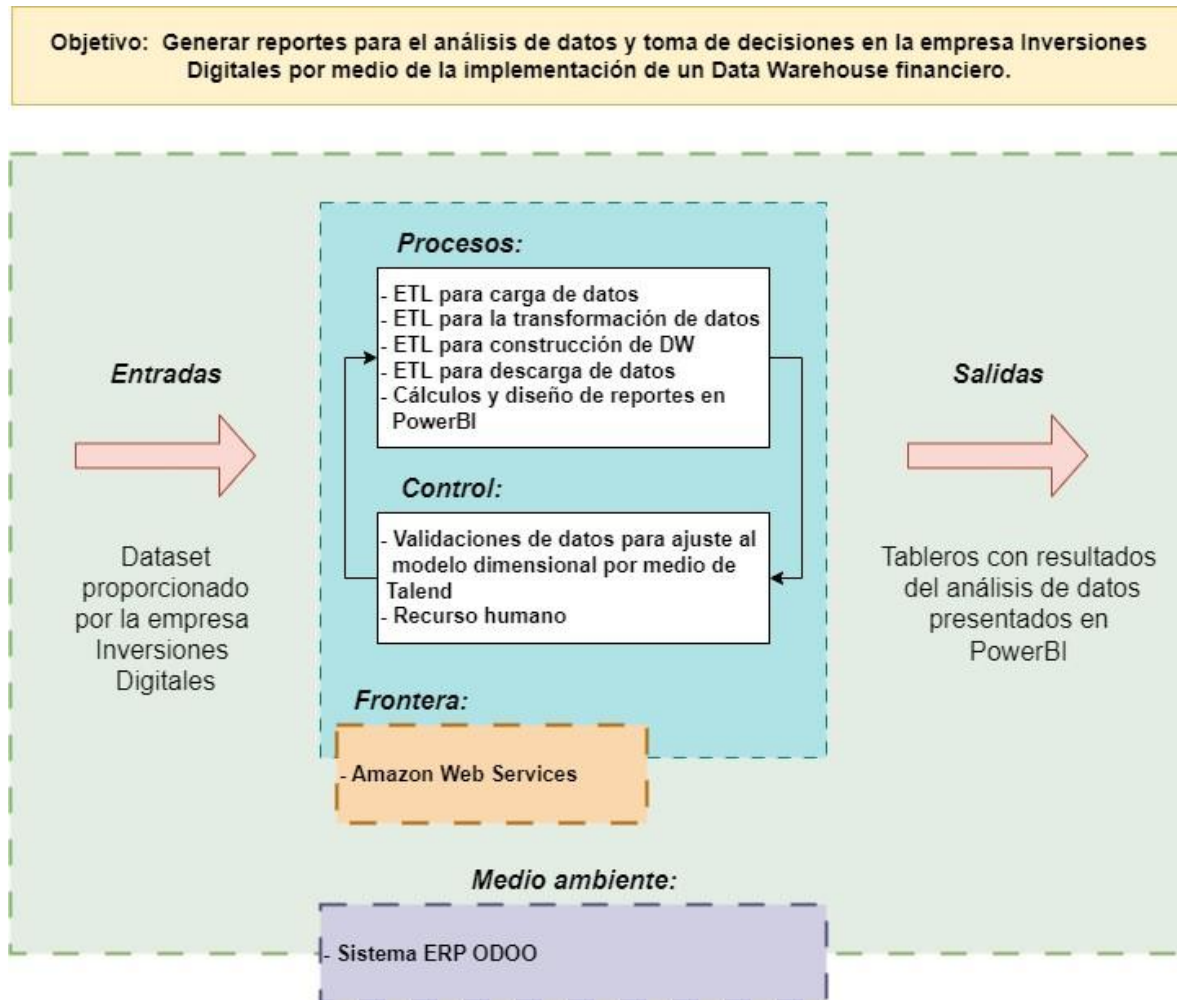


Figura 3: Enfoque de sistemas para el planteamiento del problema

Objetivo: Generar reportes para el análisis de datos y toma de decisiones en la empresa Inversiones Digitales por medio de la implementación de un Data Warehouse financiero.

Entradas:

- **Data set proporcionado por la empresa Inversiones Digitales:** Las entradas representan todo aquello que el sistema recibe desde su medio ambiente en este caso es la base datos transaccional proveniente del sistema ERP ODOO.

Procesos:

- **ETL para carga de datos:** Este se refiere al proceso para la extracción de datos de la base de datos para luego realizar la carga de dichos datos en el servicio S3 de Amazon Web Services en la zona designada para el almacenamiento de data cruda.
- **ETL para transformación de datos:** Este proceso se refiere a extraer de la zona de data cruda los datos necesarios y realizar con ellos la limpieza y transformaciones necesarias en el software Talend Open Studio para su posterior almacenamiento en el servicio de S3 en la zona de procesamiento.
- **ETL para construcción de DW:** Este proceso se refiere a tomar los datos que ya se encuentran con formatos idóneos de la zona de procesamiento de datos y llevar cada uno a las diferentes dimensiones o tabla de hechos según sea necesario para su posterior carga en la zona de acceso de S3 en Amazon Web Services.
- **ETL para descarga de datos:** Se refiere al proceso para poder descargar datos de la zona de datos crudos y de procesamiento de datos que se encuentran almacenados en las respectivas carpetas de Raw y Process en el servicio de S3 de Amazon Web Services.
- **Cálculos y diseño de reportes en PowerBI:** Este proceso se refiere a los cálculos realizados en el software PowerBI para el diseño de los reportes de las métricas a analizar.

Control:

- **Validaciones de datos para ajuste al modelo dimensional por medio de Talend Open Studio:** Es un mecanismo interno del sistema que se encarga de monitorear las salidas y para este caso se trata de del software Talend Open Studio donde se verifican que todas las salidas de datos se realicen de forma correcta y los datos tengan los formatos adecuados, además de tener los datos necesarios para que se acoplen de manera correcta al modelo dimensional propuesto.
- **Recurso humano:** En este caso, se refiere al personal que se encarga del ingreso de datos, además de que ellos son quienes controlan que los procesos se realicen de forma adecuada.

Frontera:

- **Amazon Web Services:** Se ha definido como la frontera los servicios de AWS ya que, es ahí donde se almacena la data cruda, los datos procesados y los datos de las dimensiones y tabla de hechos, y es por medio de AWS que se lleva a cabo el intercambio de información entre las entradas y salidas del sistema.

Medio ambiente:

- **Sistema ERP ODOO:** El medio ambiente es todo aquello que se encuentra fuera del sistema, pero proporciona las entradas al mismo, y es el sistema ERP ODOO el que genera la base de datos transaccional de la cual se parte para el diseño de la solución por medio del Data Warehouse financiero.

b) Alcances

El proyecto de implementación de una solución de Big Data a través de un Data Warehouse financiero para Inversiones Digitales tiene alcances definidos que a continuación se listan:

- Diseño de un modelo dimensional por medio del cual se pueda generar una solución para las métricas establecidas por la empresa para ayudar a la toma de decisiones.
- Creación de estructura de carpetas que se seguirá en el software Talend Open Studio para el almacenamiento de data cruda (Raw), data procesada (Process) y la data que estará accesible (Access) para la herramienta de visualización de datos (Power BI).
- Creación de estructura de carpetas para el almacenamiento de datos en el servicio de S3 y Redshift de Amazon Web Services.
- Creación de procesos ETL cuyos resultados serán almacenados en el servicio S3 de AWS, para lo anterior se implementarán el proceso ETL para carga de datos, ETL para transformación de datos, ETL para construcción de Data Warehouse, ETL para descarga de datos.
- Diseño de tableros en PowerBI que incluirán gráficas, listas o tablas según sea necesario para la presentación de las métricas establecidas por la empresa que de detallan a continuación:
 - ✓ Beneficio neto.
 - ✓ Retorno sobre la inversión (ROI).
 - ✓ Tasa de crecimiento en utilidades por período.
 - ✓ Liquidez.
 - ✓ Ratio de eficiencia en cobros.

c) Justificación

Inversiones Digitales se encuentra en un entorno empresarial altamente dinámico y competitivo, donde la toma de decisiones informada es esencial para el éxito y la sostenibilidad.

Es así que la implementación de una solución de Big Data a través de un Data Warehouse financiero para Inversiones Digitales se justifica en virtud de optimizar la gestión y análisis de datos financieros, abordando las métricas clave que son fundamentales para la toma de decisiones estratégicas y operativas. La creación de un entorno consolidado permitirá obtener, procesar y analizar eficientemente métricas cruciales como el de beneficio neto, el margen comercial, el retorno sobre la inversión (ROI), la tasa de crecimiento en utilidades por período y la liquidez.

La necesidad de una solución de Big Data se fundamenta en la creciente complejidad de los datos financieros de Inversiones Digitales, los cuales requieren un enfoque más avanzado y robusto para su manejo. Un Data Warehouse financiero proporcionará un repositorio centralizado donde se integrarán datos de diversas fuentes, permitiendo una visión unificada y coherente de la información financiera de la empresa.

La implementación de esta solución no solo se traducirá en una mayor eficiencia operativa, sino que también facilitará la generación de tableros de control mediante herramientas como Power BI. Estos tableros permitirán visualizar de manera clara y accesible las métricas financieras clave, brindando a los tomadores de decisiones una visión instantánea y detallada del rendimiento financiero de la empresa.

La importancia de contar con tableros de control en Power BI radica en su capacidad para transformar datos complejos en información accionable. Al presentar las métricas de interés de manera visual y comprensible, se facilita la identificación de tendencias, patrones y áreas de mejora. Esta solución no solo mejora la capacidad de respuesta ante cambios en el entorno financiero, sino que también fortalece la capacidad de la empresa para anticipar y mitigar riesgos, así como para capitalizar oportunidades estratégicas.

Es por todo esto que la creación de una solución de Big Data mediante un Data Warehouse financiero, combinada con la implementación de tableros de control en Power BI, no solo aborda las necesidades específicas de Inversiones Digitales en términos de métricas financieras, sino que también representa una inversión estratégica que potenciará la toma de decisiones basada en datos y optimizará el rendimiento financiero global de la empresa.

d) Diagrama del data set

El data set consta de 7 tablas provenientes de la base de datos transaccional proporcionada por la empresa Inversiones Digitales, la cual contiene datos de sus operaciones entre los meses de junio a diciembre del 2023.

Para efectos de una mejor visualización en la figura 4 se han incluido los primeros 10 campos de cada una de las tablas, el en diccionario de datos se puede visualizar todos los campos con los que cuenta cada una de las tablas.

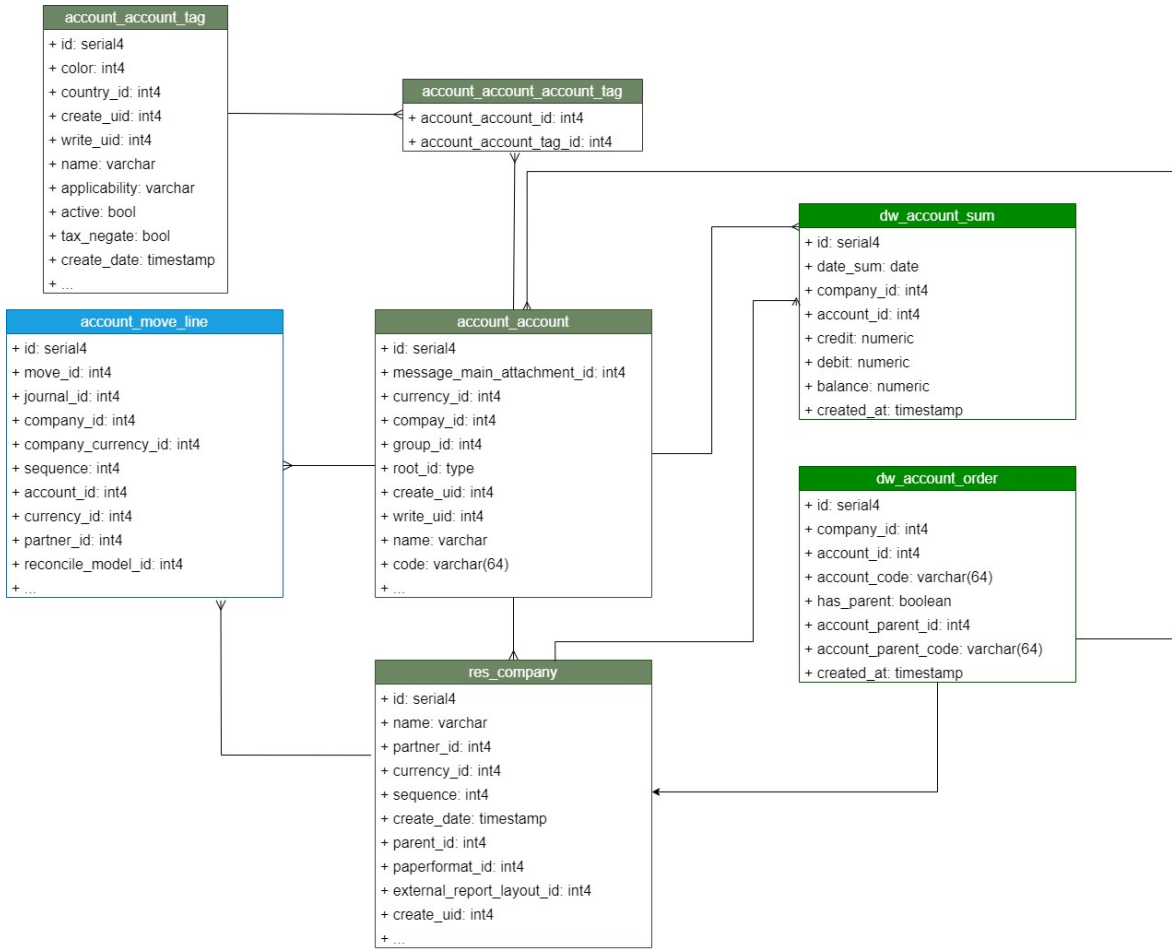


Figura 4 Diagrama del data set

e) Diccionario de datos del data set

Tabla: account_account

Descripción: La tabla 1 contiene el catálogo de cuentas utilizado.

No.	Nombre	Llave	Tipo de dato	Tamaño	Atributos	Nulo	Detalle	Descripción
1	id	PK	integer		Secuencia	Si	Auto incremental	Identificador de la tabla y llave primaria.
2	message_main_attachment_id	FK	integer			Si		Llave foránea de la tabla ir_attachment
3	currency_id	FK	integer			Si		Llave foránea de la tabla res_currency
4	company_id	FK	integer			No		Llave foránea de la tabla res_company
5	group_id	FK	integer			Si		Llave foránea de la tabla account_group
6	root_id	FK	integer			Si		Número de fuente de la cuenta
7	create_uid	FK	integer			Si		Llave foránea de la tabla res_users
8	write_uid	FK	integer			Si		Llave foránea de la tabla res_users
9	name		varchar			No		Nombre de la cuenta

10	code		varchar	64		No		Código de la cuenta
11	account_type		varchar			No		Tipo de cuenta
12	internal_group		varchar			Si		Grupo de cuenta
13	note		text			Si		Notas
14	deprecated		boolean			Si		Indica si la cuenta está obsoleta
15	include_initial_balance		boolean			Si		Indica si lleva el saldo de las cuentas
16	reconcile		boolean			Si		Indica si permite conciliación
17	is_off_balance		boolean			Si		Indica si está fuera de balance
18	non_trade		boolean			Si		Indica si la cuenta es no comercial
19	create_date		timestamp			Si		Fecha de creación
20	write_date		timestamp			Si		Fecha de modificación

Tabla 1 Descripción de datos del data set. Tabla: account_account

Tabla: account_account_tag

Descripción: La tabla 2 contiene categorías con su descripción y detalles de cada una de las cuentas del catálogo.

No.	Nombre	Llave	Tipo de dato	Tamaño	Atributos	Nulo	Detalle	Descripción
1	id	PK	integer		Secuencia	No	Auto incremental	Identificador de la tabla y llave primaria.
2	color		integer			Si		Color de la etiqueta

3	country_id	FK	integer			Si		Llave foránea de la tabla res_country
4	create_uid	FK	integer			Si		Llave foránea de la tabla res_users de creación
5	write_uid	FK	integer			No		Llave foránea de la tabla res_users de actualización
6	name		varchar			No		Nombre de etiqueta
7	applicability		varchar			Si		Aplicabilidad
8	active		boolean			Si		Indica si está activa
9	tax_negate		boolean			Si		Indica si se negarán saldo de impuestos
10	create_date		timestamp			Si		Fecha de creación
11	write_date		timestamp			Si		Fecha de actualización

Tabla 2 Descripción de datos del data set. Tabla: account_account_tag

Tabla: account_account_account_tag

Descripción: La tabla 3 es presenta la relación muchos a muchos, donde enlaza las cuentas del catálogo a las diferentes categorías o etiquetas a las que pueden pertenecer.

No.	Nombre	Llave	Tipo de dato	Tamaño	Atributos	Nulo	Detalle	Descripción
1	account_account_id	FK	integer			No		Llave foránea de la tabla account_account
2	account_account_tag_id	FK	integer			No		Llave foránea de la tabla account_account_tag

Tabla 3 Descripción de datos del data set. Tabla: account_account_account_tag

Tabla: account_move_line

Descripción: La tabla 4 contiene cada uno de los movimientos financieros en las diferentes cuentas de la empresa.

No.	Nombre	Llave	Tipo de dato	Tamaño	Atributos	Nulo	Detalle	Descripción
1	id	PK	integer		Secuencia	No	Auto incremental	Identificador de la tabla
2	move_id	FK	integer			No		Llave foránea de la tabla account_move
3	journal_id	FK	integer			Si		Llave foránea de la tabla account_journal
4	company_id	FK	integer			Si		Llave foránea de la tabla res_company
5	company_currency_id	FK	integer			Si		Llave foránea de la tabla res_currency
6	sequence		integer			Si		Secuencia
7	account_id	FK	integer			Si		Llave foránea de la tabla account_account
8	currency_id	FK	integer			No		Llave foránea de la tabla res_currency
9	partner_id	FK	integer			Si		Llave foránea de la tabla res_partner
10	reconcile_model_id	FK	integer			Si		Llave foránea de la tabla account_reconcile_model
11	payment_id	FK	integer			Si		Llave foránea de la tabla account_payment
12	statement_line_id	FK	integer			Si		Llave foránea de la tabla account_bank_statement_line
13	statement_id	FK	integer			Si		Llave foránea de la tabla account_bank_statement
14	group_tax_id	FK	integer			Si		Llave foránea de la tabla account_tax
15	tax_line_id	FK	integer			Si		Llave foránea de la tabla account_tax
16	tax_group_id	FK	integer			Si		Llave foránea de la tabla account_tax_group
17	tax_repartition_line_id	FK	integer			Si		Llave foránea de la tabla account_tax_repartition_line

18	full_reconcile_id	FK	integer			Si		Llave foránea de la tabla account_full_reconcile
19	account_root_id	FK	integer			Si		Raíz de la cuenta
20	product_id	FK	integer			Si		Llave foránea de la tabla product_product
21	product_uom_id	FK	integer			Si		Llave foránea de la tabla uom_uom
22	create_uid	FK	integer			Si		Llave foránea de la tabla res_users creación
23	write_uid	FK	integer			Si		Llave foránea de la tabla res_users modificación
24	move_name		varchar			Si		Nombre del movimiento
25	parent_state		varchar			Si		Estado del movimiento
26	ref	FK	varchar			Si		Referencia del movimiento
27	name	FK	varchar			Si		Etiqueta del movimiento
28	tax_audit		varchar			Si		Cadena de auditoría fiscal
29	matching_number		varchar			Si		Número coincidencia
30	display_type		varchar			No		Tipo de visualización
31	date		date			Si		Fecha del movimiento
32	date_maturity		date			Si		Fecha de vencimiento del movimiento
33	discount_date		date			Si		Fecha de descuento del movimiento
34	analytic_distribution		jsonb			Si		Distribución analítica
35	debit		numeric			Si		Monto de débito
36	credit		numeric			Si		Monto de crédito
37	balance		numeric			Si		Monto de balance
38	amount_currency		numeric			Si		Importe en moneda
39	tax_base_amount		numeric			Si		Monto de impuesto base
40	amount_residual		numeric			Si		Monto residual
41	amount_residual_currency		numeric			Si		Monto residual en moneda

42	quantity		numeric			Si	Cantidad de productos
43	price_unit		numeric			Si	Precio unitario
44	price_subtotal		numeric			Si	Monto del subtotal
45	price_total		numeric			Si	Total
46	discount		numeric			Si	Porcentaje de descuento
47	discount_amount_currency		numeric			Si	Cantidad de descuento en moneda
48	discount_balance		numeric			Si	Balance del descuento
49	tax_tag_invert		boolean			Si	Indica si se deben invertir las etiquetas de impuesto
50	reconciled		boolean			Si	Indica si el movimiento fue conciliado
51	blocked		boolean			Si	Indica si el movimiento fue bloqueado
52	create_date		timestamp			Si	Fecha de creación
53	write_date		timestamp			Si	Fecha de modificación
54	discount_percentage		float			Si	Porcentaje de descuento por defecto
55	is_downpayment		boolean			Si	Indica si se realiza un pago inicial
56	purchase_line_id	FK	integer			Si	Llave foránea de la tabla purchase_order_line
57	asset_category_id	FK	integer			Si	Llave foránea de la tabla account_asset_category
58	asset_start_date		date			Si	Fecha de inicio del activo
59	asset_end_date		date			Si	Fecha de fin del activo
60	asset_mrr		float			Si	Ingresos recurrentes mensuales
61	followup_line_id	FK	integer			Si	Llave foránea de la tabla followup_line
62	followup_date		date			Si	Fecha de último seguimiento

Tabla 4 Descripción de datos del data set. Tabla: account_move_line

Tabla: dw_account_order

Descripción: La tabla 5 muestra los datos sobre la jerarquía de las cuentas del catálogo de cuentas

No.	Nombre	Llave	Tipo de dato	Tamaño	Atributos	Nulo	Detalle	Descripción
1	id	PK	integer		Secuencia	Si	Auto incremental	Identificador de la tabla y llave primaria.
2	company_id	FK	integer			No		Llave foránea de la tabla res_company
3	account_id	FK	integer			No		Llave foránea de la tabla account_account
4	account_code		varchar	64		No		Código de la cuenta
5	account_name		varchar	100		No		Nombre de la cuenta
6	account_type		varchar	150		No		Tipo de la cuenta
7	account_group		varchar	15		No		Grupo a la que pertenece la cuenta
8	has_parent		boolean			No		Si la cuenta posee una cuenta padre el valor es true de lo contrario es false
9	parent_account_id	FK	integer			No		Llave foránea de la tabla account_account
10	parent_account_code		varchar	64		Si		Código de la cuenta padre
11	parent_account_name		varchar	100		Si		Nombre de la cuenta padre

12	created_at		datetime			No		Fecha y hora de creación del registro
----	------------	--	----------	--	--	----	--	---------------------------------------

Tabla 5 Descripción de datos del data set. Tabla: dw_account_order

Tabla: dw_account_sum

Descripción: La tabla 6 contiene los datos sumariados de cada cuenta del catálogo de cuentas, por día y sucursal

No.	Nombre	Llave	Tipo de dato	Tamaño	Atributos	Nulo	Detalle	Descripción
1	id	PK	integer		Secuencia	Si	Auto incremental	Identificador de la tabla y llave primaria.
2	date_sum		date			No		Fecha de los registros de las cuentas a sumar
3	company_id	FK	integer			No		Llave foránea de la tabla res_company
4	company_name		varchar	100		No		Nombre de la sucursal
5	account_id	FK	integer			No		Llave foránea de la tabla account_account
6	account_code		varchar	64		No		Código de la cuenta
7	account_name		varchar	100		No		Nombre de la cuenta
8	credit		numeric			No		
9	debit		numeric			No		

10	balance		numeric			No		
11	initial_balance		numeric			No		Valor inicial de la cuenta para un día en específico
12	final_balance		numerci			No		Valor final de la cuenta para un día específico
13	calculated_balance		boolean			No		Si el balance de la cuenta fue calculado tiene valor true, de lo contrario tiene valor false
14	created_at		datetime			No		Fecha y hora de creación del registro

Tabla 6 Descripción de datos del data set. Tabla: dw_account_sum

Tabla: res_company

Descripción: La tabla 7 contiene datos sobre las compañías o puntos de venta que se manejan en la aplicación.

No.	Nombre	Llave	Tipo de dato	Tamaño	Atributos	Nulo	Detalle	Descripción
1	id	PK	integer		Secuencia	No	Auto incremental	Identificador de la tabla y llave primaria.
2	name		varchar		Único	No		Nombre de la compañía
3	partner_id	FK	integer			No		Llave foránea de la tabla res_partner

4	currency_id	FK	integer			No		Llave foránea de la tabla res_currency
5	sequence	FK	integer			Si		Secuencia
6	create_date		timestamp			Si		Fecha de creación de la compañía
7	parent_id	FK	integer			Si		Llave foránea recursiva de la tabla res_company
8	paperformat_id	FK	integer			Si		Llave foránea de la tabla report_paperformat
9	external_report_layout_id	FK	integer			Si		Llave foránea de la tabla ir_ui_view
10	create_uid	FK	integer			Si		Llave foránea de la tabla res_users
11	write_uid	FK	integer			Si		Llave foránea de la tabla res_users
12	email		varchar			Si		Correo de la compañía
13	phone		varchar			Si		Teléfono de la compañía
14	mobile		varchar			Si		Teléfono móvil de la compañía
15	base_onboarding_company_state		varchar			Si		Estado de incorporación de la empresa
16	font		varchar			Si		Fuente
17	primary_color		varchar			Si		Color primario
18	secondary_color		varchar			Si		Color secundario
19	layout_background		varchar			No		Fondo del layout

20	report_footer		jsonb			Si		Informe del pie de página
21	report_header		text			Si		Informe del encabezado
22	company_details		text			Si		Detalle de la compañía
23	active		boolean			Si		Indica si la compañía está activa
24	write_date		timestamp			Si		Última fecha de actualización
25	logo_web		bytea			Si		Logo de la compañía
26	resource_calendar_id	FK	integer			Si		Llave foránea de la tabla resource_calendar
27	partner_gid		integer			Si		ID de la base de datos de la empresa
28	iap_enrich_auto_done		boolean			Si		Configuración de la compañía
29	snailmail_color		boolean			Si		Color del correo postal
30	snailmail_cover		boolean			Si		Portada del correo postal
31	snailmail_duplex		boolean			Si		Dúplex del correo postal
32	payment_provider_onboarding_state		varchar			Si		Estado de incorporación del proveedor de pago
33	payment_onboarding_payment_method		varchar			Si		Estado de incorporación del método de pago
34	message_main_attachment_id	FK	integer			Si		Llave foránea de la tabla ir_attachment
35	fiscalyear_last_day		integer			No		Último día del año fiscal

36	transfer_account_id	FK	integer			Si		Llave foránea de la tabla account_account
37	chart_template_id	FK	integer			Si		Llave foránea de la tabla account_chart_template
38	default_cash_difference_income_account_id	FK	integer			Si		Llave foránea de la tabla account_account
39	default_cash_difference_expense_account_id	FK	integer			Si		Llave foránea de la tabla account_account
40	account_journal_suspense_account_id	FK	integer			Si		Llave foránea de la tabla account_account
41	account_journal_payment_debit_account_id	FK	integer			Si		Llave foránea de la tabla account_account
42	account_journal_payment_credit_account_id	FK	integer			Si		Llave foránea de la tabla account_account
43	account_journal_early_pay_discount_gain_account_id	FK	integer			Si		Llave foránea de la tabla account_account
44	account_journal_early_pay_discount_loss_account_id	FK	integer			Si		Llave foránea de la tabla account_account
45	account_sale_tax_id	FK	integer			Si		Llave foránea de la tabla account_tax
46	account_purchase_tax_id	FK	integer			Si		Llave foránea de la tabla account_tax

47	currency_exchange_journal_id	FK	integer			Si		Llave foránea de la tabla account_journal
48	income_currency_exchange_account_id	FK	integer			Si		Llave foránea de la tabla account_account
49	expense_currency_exchange_account_id	FK	integer			Si		Llave foránea de la tabla account_account
50	property_stock_account_input_categ_id	FK	integer			Si		Llave foránea de la tabla account_account
51	property_stock_account_output_categ_id	FK	integer			Si		Llave foránea de la tabla account_account
52	property_stock_valuation_account_id	FK	integer			Si		Llave foránea de la tabla account_account
53	incoterm_id	FK	integer			Si		Llave foránea de la tabla account_incoterms
54	account_opening_move_id	FK	integer			Si		Llave foránea de la tabla account_move
55	account_default_pos_receivable_account_id	FK	integer			Si		Llave foránea de la tabla account_account
56	expense_accrual_account_id	FK	integer			Si		Llave foránea de la tabla account_account
57	revenue_accrual_account_id	FK	integer			Si		Llave foránea de la tabla account_account

58	automatic_entry_default_journal_id	FK	integer			Si		Llave foránea de la tabla account_journal
59	account_fiscal_country_id	FK	integer			Si		Llave foránea de la tabla res_country
60	tax_cash_basis_journal_id	FK	integer			Si		Llave foránea de la tabla account_journal
61	account_cash_basis_base_account_id	FK	integer			Si		Llave foránea de la tabla account_account
62	fiscalyear_last_month		varchar			Si		Último mes del año fiscal
63	bank_account_code_prefix		varchar			Si		Prefijo de las cuentas bancarias
64	cash_account_code_prefix		varchar			Si		Prefijo de las cuentas de caja
65	early_pay_discount_computation		varchar			Si		Reducción de impuestos de descuento en efectivo
66	transfer_account_code_prefix		varchar			Si		Prefijo de las cuentas de transferencia
67	tax_calculation_rounding_method		varchar			Si		Método de redondeo de cálculo de impuestos
68	account_setup_bank_data_state		varchar			Si		Estado de datos bancarios
69	account_setup_fy_data_state		varchar			Si		Estado del año fiscal
70	account_setup_coa_state		varchar			Si		Estado de gráfico de cuentas

71	account_setup_taxes_state		varchar			Si		Estado de impuestos
72	account_onboarding_invoice_layout_state		varchar			Si		Estado de facturas en el layout
73	account_onboarding_sale_tax_state		varchar			Si		Estado de los impuestos de venta
74	account_invoice_onboarding_state		varchar			Si		Estado de datos bancarios
75	account_dashboard_onboarding_state		varchar			Si		Estado de datos bancarios
76	terms_type		varchar			Si		Formato de términos y condiciones
77	account_setup_bill_state		varchar			Si		Estado de facturas
78	quick_edit_mode		varchar			Si		Codificación rápida
79	period_lock_date		date			Si		Fecha de bloqueo de entradas del diario
80	fiscalyear_lock_date		date			Si		Fecha de bloqueo de todos los usuarios
81	tax_lock_date		date			Si		Fecha de bloqueo de declaración de impuestos
82	account_opening_date		date			No		Entrada de apertura
83	invoice_terms		jsonb			Si		Términos y condiciones predeterminados

84	invoice_terms_html		jsonb			Si		Términos y condiciones predeterminados en el sitio web
85	expects_chart_of_accounts		boolean			Si		Indica si habrá un gráfico de cuentas
86	anglo_saxon_accounting		boolean			Si		Indica si usa contabilidad anglosajona
87	qr_code		boolean			Si		Indica si habrá un gráfico de cuentas
88	invoice_is_email		boolean			Si		Indica si muestra código QR en las facturas
89	invoice_is_print		boolean			Si		Indica si envía el correo electrónico por defecto de las facturas
90	account_use_credit_limit		boolean			Si		Indica si existe límite de crédito de ventas
91	account_onboarding_create_invoice_state_flag		boolean			Si		Indica si se crea cuenta para estados de factura
92	tax_exigibility		boolean			Si		Indica si los impuestos son obligatorios
93	account_storno		boolean			Si		Indica si habrá contabilidad de cancelaciones
94	invoice_is_snailmail		boolean			Si		Indica si se mandan las facturas por correo

95	quotation_validity_days		integer			Si		Validez de cotización predeterminada (días)
96	sale_quotation_onboarding_state		varchar			Si		Estado del panel de ventas
97	sale_onboarding_order_confirmation_state		varchar			Si		Estado de confirmación de ventas
98	sale_onboarding_sample_quotation_state		varchar			Si		Estado de cotización de ventas
99	sale_onboarding_payment_method		varchar			Si		Estado del método de pago de ventas
100	portal_confirmation_sign		boolean			Si		Indica si hay firma electrónica
101	portal_confirmation_pay		boolean			Si		Indica si se efectuarán pagos en línea
102	sale_order_template_id	FK	integer			Si		Llave foránea de la tabla sale_order_template
103	social_twitter		varchar			Si		Cuenta de Twitter
104	social_facebook		varchar			Si		Cuenta de Facebook
105	social_github		varchar			Si		Cuenta de Github
106	social_linkedin		varchar			Si		Cuenta de LinkedIn
107	social_youtube		varchar			Si		Cuenta de Youtube
108	social_instagram		varchar			Si		Cuenta de Instagram

109	hr_presence_control_email_amount		integer			Si		Cantidad de correos a enviar
110	hr_presence_control_ip_list		varchar			Si		Direcciones IP válidas
111	nomenclature_id	FK	integer			Si		Llave foránea de la tabla barcode_nomenclature
112	internal_transit_location_id	FK	integer			Si		Llave foránea de la tabla Internal stock_location
113	stock_mail_confirmation_template_id	FK	integer			Si		Llave foránea de la tabla mail_template
114	annual_inventory_day		integer			Si		Día del inventario anual
115	annual_inventory_month		varchar			Si		Mes del inventario anual
116	stock_move_email_validation		boolean			Si		Correo electrónico de confirmación en la recolección de productos
117	stock_sms_confirmation_template_id	FK	integer			Si		Llave foránea de la tabla sms_template
118	stock_move_sms_validation		boolean			Si		Plantilla de SMS
119	has_received_warning_stock_sms		boolean			Si		Habrará SMS de confirmación
120	point_of_sale_update_stock_quantities		varchar			Si		Recibirá un SMS de advertencia de existencias en inventario

121	point_of_sale_use_ticket_qr_code		boolean			Si		Actualizar cantidades en stock
122	security_lead		float			No		Use el código QR en el ticket
123	po_lock		varchar			Si		Días seguros de ventas
124	po_double_validation		varchar			Si		Modificación de orden de compra
125	po_double_validation_amount		numeric			Si		Niveles de aprobación
126	po_lead		float			No		Cantidad de validación doble
127	days_to_purchase		float			Si		Días para comprar
128	manufacturing_lead		float			No		Plazo de entrega de fabricación

Tabla 7 Descripción de datos del data set. Tabla: res_company

f) Data profiling

Tabla: account_account

Análisis numérico

	id	message_main_attachment_id	currency_id	company_id	group_id	root_id	create_uid	write_uid
Nº de filas	215	215	215	215	215	215	215	215
Nº de Nulos	0	215	0	0	215	0	0	0
Valor más alto	215	0	2	5	0	57,057	1	1
Valor más bajo	1	0	2	1	0	49,048	1	1
Suma	23,220	0	430	645	0	11,010,865	215	215
Media	108	0	2	3	0	51,213.33	1	1
Media Geométrica	80.431	0	2	2.605	0	51,154.54	1	1
Desviación Estándar	62.209	0	0	1.418	0	2,493.26	0	0
Varianza	3,870	0	0	2.009	0	6,216,338.17	0	0
Second moment	828,180	0	0	430	0	1,330,296,367.21	0	0
Sum of squares	#####	0	860	2,365	0	565,233,310,545	215	215

Tabla 8 Data profiling. Análisis numérico. Tabla: account_account

Análisis de String

	name	code	account_type	internal_group	note
Nº de filas	215	215	215	215	215
Nº de Nulos	0	0	0	0	215
Nº de espacios en blanco	0	0	0	0	0
Conteo sólo mayúsculas	0	0	0	0	0
Conteo sólo minúsculas	10	0	215	215	0
Conteo total de caracteres	4050	1290	2500	1345	0
Max Caracteres	35	6	21	9	0
Min Caracteres	5	6	6	5	0
Promedio de caracteres	18.837	6	11.628	6.256	0
Max de espacios en blanco	4	0	0	0	0
Min de espacios en blanco	0	0	0	0	0
Promedio de espacios en blanco	1.651	0	0	0	0
Caracteres en mayúscula	245	0	0	0	0
Caracteres en mayúscula (excepto 1ra letra)	40	0	0	0	0
Caracteres en minúscula	3415	0	2350	1345	0
Caracteres de dígitos	0	1290	0	0	0
Caracteres diacríticos	25	0	0	0	0
Caracteres sin letras	390	1290	150	0	0
Contador de palabras	570	215	215	215	0
Máximo de palabras	5	1	1	1	0
Mínimo de palabras	1	1	1	1	0

Tabla 9 Data profiling. Análisis String. Tabla: account_account

Análisis de Fecha/Tiempo

	create_date	write_date
Nº de filas	215	215
Nº de nulos	0	0
Fecha más alta	2023-06-01	2023-12-05
Fecha más baja	2023-07-07	2023-07-07
Tiempo más alto	22:01:25.268	22:01:25.268
Tiempo más bajo	00:45:42.866	00:45:42.866
Media	2023-07-07 22:34	2023-07-07 22:34

Tabla 10 Data profiling. Análisis fecha/tiempo. Tabla: account_acccount

Análisis de Booleanos

	deprecated	include_initial_balance	reconcile	is_off_balance	non_trade
Nº de filas	215	215	215	215	215
Nº de nulos	0	0	0	0	0
Nº de TRUE	0	135	50	0	0
Nº de FALSE	215	80	165	215	215

Tabla 11 Data profiling. Análisis booleanos. Tabla: account_account

Tabla: account_account_account_tag

Análisis numérico

	account_account_id	account_account_tag_id
Nº de filas	13	13
Nº de Nulos	0	0
Valor más alto	34	3
Valor más bajo	21	1
Suma	361	28
Media	27.769	2.154
Media Geométrica	27.464	1.948
Desviación Estándar	4.226	0.899
Varianza	17.859	0.808
Second moment	214.308	9.692
Sum of squares	10,239	70

Tabla 12 Data profiling. Análisis numérico. Tabla: account_account_account_tag

Tabla: account_account_tag

Análisis de string

	name	applicability
Nº de filas	3	3
Nº de Nulos	0	0
Nº de espacios en blanco	0	0
Conteo sólo mayúsculas	0	0
Conteo sólo minúsculas	0	3
Conteo total de caracteres	76	24

Max Caracteres	36	8
Min Caracteres	20	8
Promedio de caracteres	25.333	8
Max de espacios en blanco	3	0
Min de espacios en blanco	1	0
Promedio de espacios en blanco	1.667	0
Caracteres en mayúscula	7	0
Caracteres en mayúscula (excepto 1ra letra)	4	0
Caracteres en minúscula	63	24
Caracteres de dígitos	0	0
Caracteres diacríticos	0	0
Caracteres sin letras	6	0
Contador de palabras	8	3
Máximo de palabras	4	1
Mínimo de palabras	2	1

Tabla 13 Data profiling. Análisis String. Tabla: account_account_account_tag

Análisis de Fecha/Tiempo

	create_date	write_date
Nº de filas	3	3
Nº de nulos	0	0
Fecha más alta	1/6/2023	1/6/2023
Fecha más baja	5/12/2023	5/12/2023
Tiempo más alto	45:32.2	45:32.2
Tiempo más bajo	45:32.2	45:32.2
Media	15/7/2023 00:45	15/7/2023 00:45

Tabla 14 Data profiling. Análisis fecha/tiempo. Tabla: account_account_account_tag

Análisis de booleanos

	active	tax_negate
Nº de filas	3	3
Nº de nulos	0	3
Nº de TRUE	3	0
Nº de FALSE	0	0

Tabla 15 Data profiling. Análisis booleanos. Tabla: account_account_account_tag

Tabla: account_move_line

Análisis Numérico

campo	Nº de filas	Nº de Nulos	Valor más alto	Valor más bajo	Suma	Media	Media Geométrica	Desviación Estándar	Varianza
id	290	0	290	1	42,195	145.5	108.075	83.86	7,032.50
move_id	290	0	20	1	2,880	9.931	7.899	5.372	28.853
journal_id	290	0	9	2	1,405	4.845	4.043	2.662	7.087
company_id	290	0	5	1	870	3	2.605	1.417	2.007
company_currency_id	290	0	2	2	580	2	2	0	0
sequence	290	0	12,000	10	626,860	2,161.59	139.108	4,349.20	18,915,564.36
account_id	290	0	41	8	7,405	25.534	22.997	10.595	112.257
currency_id	290	0	2	2	580	2	2	0	0
partner_id	290	115	28	14	3,760	21.486	20.71	5.516	30.424
reconcile_model_id	290	290	0	0	0	0	0	0	0

payment_id	290	230	6	1	210	3.5	2.994	1.722	2.966
statement_line_id	290	230	6	1	210	3.5	2.994	1.722	2.966
statement_id	290	290	0	0	0	0	0	0	0
group_tax_id	290	290	0	0	0	0	0	0	0
tax_line_id	290	250	2	1	65	1.625	1.542	0.49	0.24
tax_group_id	290	250	3	3	120	3	3	0	0
tax_repartition_line_id	290	250	4	2	130	3.25	3.084	0.981	0.962
full_reconcile_id	290	210	8	1	360	4.5	3.764	2.306	5.316
account_root_id	290	0	54,052	49,048	14,789,180	50,997.17	50,953.39	2,132.22	4,546,372.64
product_id	290	210	38	10	2,230	27.875	26.524	7.58	57.453
product_uom_id	290	210	1	1	80	1	1	0	0
create_uid	290	0	2	2	580	2	2	0	0
write_uid	290	0	2	2	580	2	2	0	0
debit	290	0	12,204	0	376,608.05	1,298.65	0	2,323.65	5,399,358.29
credit	290	0	12,204	0	376,608.05	1,298.65	0	2,788.93	7,778,120.27
balance	290	0	12,204	-12,204	0	0	0	4,069.66	16,562,125.34
amount_currency	290	0	12,204	-12,204	0	0	0	4,069.66	16,562,125.34
tax_base_amount	290	250	10,800	189	144,520	3,613	1,928.68	3,379.16	11,418,709.74
amount_residual	290	0	6,363.03	-12,204	137,548.75	-474.306	0	2,261.42	5,114,000.32
amount_residual_currency	290	0	6,363.03	-12,204	137,548.75	-474.306	0	2,261.42	5,114,000.32
quantity	290	0	100	0	4,005	13.81	0	30.218	913.13
price_unit	290	0	2,000	0	16,477.50	56.819	0	265.883	70,693.56
price_subtotal	290	0	4,800	0	122,875	423.707	0	941.687	886,773.84
price_total	290	0	5,424	0	137,548.75	474.306	0	1,057.09	1,117,446.17
discount	290	0	0	0	0	0	0	0	0
discount_amount_currency	290	285	0	0	0	0	0	0	0

discount_balance	290	285	0	0	0	0	0	0	0
discount_percentage	290	265	0	0	0	0	0	0	0
asset_category_id	290	290	0	0	0	0	0	0	0
asset_mrr	290	0	0	0	0	0	0	0	0
followup_line_id	290	290	0	0	0	0	0	0	0
expense_id	290	285	1	1	5	1	1	0	0
purchase_line_id	290	215	15	1	600	8	6.423	4.35	18.919

Tabla 16 Data profiling. Análisis numérico. Tabla: account_move_line

Análisis de String

	move_name	parent_state	ref	name	tax_audit	matching_number	display_type
Nº de filas	290	290	290	290	290	290	290
Nº de Nulos	0	0	60	0	0	210	0
Nº de espacios en blanco	0	0	125	30	290	0	0
Conteo sólo mayúsculas	290	0	95	45	0	80	0
Conteo sólo minúsculas	0	290	0	0	0	0	290
Conteo total de caracteres	4875	1740	1455	8845	0	160	2080
Max Caracteres	18	6	18	82	0	2	12
Min Caracteres	15	6	0	0	0	2	3
Promedio de caracteres	16.81	6	6.326	30.5	0	2	7.172
Max de espacios en blanco	0	0	1	11	0	0	0
Min de espacios en blanco	0	0	0	0	0	0	0
Promedio de espacios en blanco	0	0	0.043	3.534	0	0	0
Caracteres en mayúscula	1225	0	395	895	0	80	0
Caracteres en mayúscula (excepto 1ra letra)	935	0	290	545	0	0	0
Caracteres en minúscula	0	1740	130	4790	0	0	2050

Caracteres de dígitos	2900	0	725	1410	0	80	0
Caracteres diacríticos	0	0	0	10	0	0	0
Caracteres sin letras	3650	0	930	3160	0	80	30
Contador de palabras	290	290	115	1285	0	80	290
Máximo de palabras	1	1	2	12	0	1	1
Mínimo de palabras	1	1	0	0	0	1	1

Tabla 17 Data profiling. Análisis String. Tabla: account_move_line

Análisis de Fecha/Tiempo

	date	date_maturity	discount_date	create_date	write_date	asset_start_date	asset_end_date	followup_date
Nº de filas	290	290	290	290	290	290	290	290
Nº de nulos	0	210	290	0	0	290	290	290
Fecha más alta	13/7/2023	13/7/2023	0	13/7/2023	13/7/2023	0	0	0
Fecha más baja	7/7/2023	7/7/2023	0	8/7/2023	8/7/2023	0	0	0
Tiempo más alto	00:00.0	00:00.0	0	12:35.5	12:35.5	0	0	0
Tiempo más bajo	00:00.0	00:00.0	0	59:22.2	59:53.6	0	0	0
Media	9/7/2023 23:10	9/7/2023 10:30	0	10/7/2023 15:04	10/7/2023 15:04	0	0	0

Tabla 18 Data profiling. Análisis fecha/tiempo. Tabla: account_move_line

Análisis de Booleanos

	tax_tag_invert	reconciled	blocked	is_downpayment
Nº de filas	290	290	290	290
Nº de nulos	0	0	0	290
Nº de TRUE	30	80	0	0
Nº de FALSE	260	210	290	0

Tabla 19 Data profiling. Análisis booleanos. Tabla: account_move_line

Tabla: dw_account_order

Análisis numérico

	id	company_id	account_id	parent_account_id
Nº de filas	216	216	216	216
Nº de Nulos	0	0	0	116
Valor más alto	216	5	216	198
Valor más bajo	1	1	1	2
Suma	23,436	646	23,436	10,100
Media	108.5	2.991	108.5	101
Media Geométrica	80.8	2.594	80.8	67.52
Desviación Estándar	62.498	1.421	62.498	62.043
Varianza	3,906	2.019	3,906	3,849.29
Second moment	839,790	433.981	839,790	381,080
Sum of squares	3,382,596	2,366	3,382,596	1,401,180

Tabla 20 Data profiling. Análisis numérico. Tabla: dw_account_order

Análisis de String

	account_code	account_name	account_type	account_group	parent_account_code	parent_account_name
Nº de filas	216	216	216	216	216	216
Nº de Nulos	0	0	0	0	116	116
Nº de espacios en blanco	0	0	0	0	0	0
Conteo sólo mayúsculas	0	0	0	0	0	0
Conteo sólo minúsculas	0	0	216	216	0	0
Conteo total de caracteres	1296	3333	2506	1351	600	1195
Max Caracteres	6	28	21	9	6	18
Min Caracteres	6	4	6	5	6	8
Promedio de caracteres	6	15.431	11.602	6.255	6	11.95

Max de espacios en blanco	0	3	0	0	0	1
Min de espacios en blanco	0	0	0	0	0	0
Promedio de espacios en blanco	0	1.194	0	0	0	0.65
Caracteres en mayúscula	0	423	0	0	0	165
Caracteres en mayúscula (excepto 1ra letra)	0	207	0	0	0	65
Caracteres en minúscula	0	2607	2356	1351	0	965
Caracteres de dígitos	1296	0	0	0	600	0
Caracteres diacríticos	0	0	0	0	0	0
Caracteres sin letras	1296	303	150	0	600	65
Contador de palabras	216	474	216	216	100	165
Máximo de palabras	1	4	1	1	1	2
Mínimo de palabras	1	1	1	1	1	1

Tabla 21 Data profiling. Análisis String. Tabla: dw_account_order

Análisis de Fecha/Tiempo

	created_at
Nº de filas	216
Nº de nulos	0
Fecha más alta	2023-11-01
Fecha más baja	2023-11-01
Tiempo más alto	12:10:56
Tiempo más bajo	12:10:50
	2023-11-01
Media	12:10:53

Tabla 22 Data profiling. Análisis fecha/tiempo. Tabla: dw_account_order

Análisis de booleanos

	has_parent
Nº de filas	216
Nº de nulos	0
Nº de TRUE	100
Nº de FALSE	116

Tabla 23 Data profiling. Análisis booleanos. Tabla: dw_account_order

Tabla: dw_account_sum

Análisis numérico

	Nº de filas	Nº de Nulos	Valor más alto	Valor más bajo	Suma	Media
id	197	0	1826	1630	340,416	1728
company_id	197	0	5	1	537	2.726
account_id	197	0	41	2	4,837	24.553
credit	197	0	12204	0	388,800	1973.606
debit	197	0	12204	0	388800.45	1973.606
balance	197	0	10,800	-12,204	0	0.00
initial_balance	197	0	24725	-7848.75	45269.62	36.902
final_balance	197	0	24890	-2035.2	87269.62	53.902

Tabla 24 Data profiling. Análisis numérico. Tabla: dw_account_sum

Análisis de String

	company_name	account_code	account_name
Nº de filas	197	197	197
Nº de Nulos	0	0	0
Nº de espacios en blanco	0	0	0
Conteo sólo mayúsculas	0	0	0
Conteo sólo minúsculas	0	0	0
Conteo total de caracteres	7647	1182	2617
Max Caracteres	49	6	25
Min Caracteres	33	6	4
Promedio de caracteres	38.817	6	13.284
Max de espacios en blanco	5	0	2
Min de espacios en blanco	3	0	0
Promedio de espacios en blanco	3.619	0	0.954
Caracteres en mayúscula	910	0	373
Caracteres en mayúscula (excepto 1ra letra)	713	0	176
Caracteres en minúscula	5569	0	2018
Caracteres de dígitos	0	1182	0
Caracteres diacríticos	0	0	0
Caracteres sin letras	1168	1182	226
Contador de palabras	910	197	385
Máximo de palabras	6	1	3
Mínimo de palabras	4	1	1

Tabla 25 Data profiling. Análisis String. Tabla: dw_account_sum

Análisis de Fecha/Tiempo

	date_sum	created_at
Nº de filas	197	197
Nº de nulos	0	0
Fecha más alta	1/6/2023	5/12/2023
Fecha más baja	7/7/2023	5/12/2023
Tiempo más alto	00:00.0	14:41.9
Tiempo más bajo	00:00.0	14:41.9
Media	29/7/2023 19:00	5/12/2023 21:14

Tabla 26 Data profiling. Análisis fecha/tiempo. Tabla: dw_account_sum

Análisis de booleanos

	calculated_balance
Nº de filas	197
Nº de nulos	0
Nº de TRUE	197
Nº de FALSE	0

Tabla 27 Data profiling. Análisis booleanos. Tabla: dw_account_sum

Tabla: res_company

Análisis numérico

	Nº de filas	Nº de Nulos	Valor más alto	Valor más bajo	Suma	Media	Media Geométrica	Desviación Estándar	Varianza
id	5	0	5	1	15	3	2.605	1.581	2.5
partner_id	5	0	10	1	35	7	5.502	3.536	12.5

currency_id	5	0	75	2	302	60.4	36.329	32.647	1,065.80
sequence	5	1	10	10	40	10	10	0	0
parent_id	5	5	0	0	0	0	0	0	0
paperformat_id	5	0	1	1	5	1	1	0	0
external_report_layout_id	5	0	202	202	1,010	202	202	0	0
create_uid	5	0	2	1	9	1.8	1.741	0.447	0.2
write_uid	5	0	2	1	6	1.2	1.149	0.447	0.2
partner_gid	5	5	0	0	0	0	0	0	0
resource_calendar_id	5	0	5	1	15	3	2.605	1.581	2.5
hr_presence_control_email_amount	5	5	0	0	0	0	0	0	0
message_main_attachment_id	5	5	0	0	0	0	0	0	0
fiscalyear_last_day	5	0	31	31	155	31	31	0	0
transfer_account_id	5	4	1	1	1	1	1	0	0
chart_template_id	5	4	1	1	1	1	1	0	0
default_cash_difference_income_account_id	5	4	23	23	23	23	23	0	0
default_cash_difference_expense_account_id	5	4	32	32	32	32	32	0	0
account_journal_suspend_account_id	5	4	38	38	38	38	38	0	0
account_journal_payment_debit_account_id	5	4	39	39	39	39	39	0	0
account_journal_payment_credit_account_id	5	4	40	40	40	40	40	0	0
account_journal_early_pay_discount_gain_account_id	5	4	37	37	37	37	37	0	0
account_journal_early_pay_discount_loss_account_id	5	4	36	36	36	36	36	0	0
account_sale_tax_id	5	4	1	1	1	1	1	0	0
account_purchase_tax_id	5	4	2	2	2	2	2	0	0
currency_exchange_journal_id	5	4	4	4	4	4	4	0	0
income_currency_exchange_account_id	5	4	22	22	22	22	22	0	0
expense_currency_exchange_account_id	5	4	31	31	31	31	31	0	0

property_stock_account_input_categ_id	5	4	4	4	4	4	4	0	0
property_stock_account_output_categ_id	5	4	5	5	5	5	5	0	0
property_stock_valuation_account_id	5	4	3	3	3	3	3	0	0
incoterm_id	5	5	0	0	0	0	0	0	0
account_opening_move_id	5	5	0	0	0	0	0	0	0
account_default_pos_receivable_account_id	5	4	35	35	35	35	35	0	0
expense_accrual_account_id	5	5	0	0	0	0	0	0	0
revenue_accrual_account_id	5	5	0	0	0	0	0	0	0
automatic_entry_default_journal_id	5	5	0	0	0	0	0	0	0
account_fiscal_country_id	5	4	209	209	209	209	209	0	0
tax_cash_basis_journal_id	5	4	5	5	5	5	5	0	0
account_cash_basis_base_account_id	5	5	0	0	0	0	0	0	0
quotation_validity_days	5	0	30	30	150	30	30	0	0
sale_order_template_id	5	5	0	0	0	0	0	0	0
nomenclature_id	5	0	1	1	5	1	1	0	0
internal_transit_location_id	5	0	17	13	75	15	14.933	1.581	2.5
stock_mail_confirmation_template_id	5	0	31	31	155	31	31	0	0
annual_inventory_day	5	0	31	31	155	31	31	0	0
stock_sms_confirmation_template_id	5	0	2	2	10	2	2	0	0
security_lead	5	0	0	0	0	0	0	0	0
expense_journal_id	5	5	0	0	0	0	0	0	0
company_expense_journal_id	5	5	0	0	0	0	0	0	0
po_double_validation_amount	5	0	5,00	5,00	25,00	5,000	5,000	0	0
po_lead	5	0	0	0	0	0	0	0	0
days_to_purchase	5	5	0	0	0	0	0	0	0
manufacturing_lead	5	0	0	0	0	0	0	0	0

Tabla 28 Data profiling. Análisis numérico. Tabla: res_company

Análisis de String

	Nº de filas	Nº de Nulos	Nº de espacios en blanco	Conteo mayúsculas	Conteo minúsculas	Conteo total de caracteres	Max Caracteres	Min Caracteres
name	5	0	0	0	0	186	49	33
email	5	0	0	0	5	150	35	13
phone	5	0	0	0	0	60	12	12
mobile	5	5	0	0	0	0	0	0
base_onboarding_company_state	5	0	0	0	5	36	8	4
font	5	0	0	0	0	20	4	4
primary_color	5	0	0	0	0	35	7	7
secondary_color	5	0	0	0	0	35	7	7
layout_background	5	0	0	0	0	25	5	5
report_header	5	5	0	0	0	0	0	0
company_details	5	0	0	0	0	203	91	28
social_twitter	5	5	0	0	0	0	0	0
social_facebook	5	5	0	0	0	0	0	0
social_github	5	5	0	0	0	0	0	0
social_linkedin	5	5	0	0	0	0	0	0
social_youtube	5	5	0	0	0	0	0	0
social_instagram	5	5	0	0	0	0	0	0
hr_presence_control_ip_list	5	5	0	0	0	0	0	0
payment_provider_onboarding_state	5	0	0	0	5	40	8	8
payment_onboarding_payment_method	5	5	0	0	0	0	0	0
fiscalyear_last_month	5	0	0	0	0	10	2	2
bank_account_code_prefix	5	4	0	0	0	4	4	4

cash_account_code_prefix	5	4	0	0	0	4	4	4
early_pay_discount_computation	5	0	0	0	5	40	8	8
transfer_account_code_prefix	5	4	0	0	0	4	4	4
tax_calculation_rounding_method	5	0	0	0	5	70	14	14
account_setup_bank_data_state	5	0	0	0	5	40	8	8
account_setup_fy_data_state	5	0	0	0	5	40	8	8
account_setup_coa_state	5	0	0	0	5	36	8	4
account_setup_taxes_state	5	0	0	0	5	40	8	8
account_onboarding_invoice_layout_state	5	0	0	0	5	36	8	4
account_onboarding_sale_tax_state	5	0	0	0	5	40	8	8
account_invoice_onboarding_state	5	0	0	0	5	40	8	8
account_dashboard_onboarding_state	5	0	0	0	5	40	8	8
terms_type	5	0	0	0	5	25	5	5
account_setup_bill_state	5	0	0	0	5	40	8	8
quick_edit_mode	5	5	0	0	0	0	0	0
sale_quotation_onboarding_state	5	0	0	0	5	38	8	6
sale_onboarding_order_confirmation_state	5	0	0	0	5	36	8	4
sale_onboarding_sample_quotation_state	5	0	0	0	5	40	8	8
sale_onboarding_payment_method	5	4	0	0	1	17	17	17
annual_inventory_month	5	0	0	0	0	10	2	2
po_lock	5	0	0	0	5	20	4	4
po_double_validation	5	0	0	0	5	40	8	8
point_of_sale_update_stock_quantities	5	0	0	0	5	35	7	7

Tabla 29 Data profiling. Análisis String. Tabla: res_company

Análisis de Fecha/Tiempo

	create_date	write_date	period_lock_date	fiscalyear_lock_date	tax_lock_date	account_opening_date
Nº de filas	5	5	5	5	5	5
Nº de nulos	0	0	5	5	5	0
Fecha más alta	1/6/2023	1/6/2023	0	0	0	1/1/2023
Fecha más baja	5/12/2023	5/12/2023	0	0	0	1/1/2023
Tiempo más alto	18:20.7	35:27.5	0	0	0	00:00.0
Tiempo más bajo	15:18.9	47:20.4	0	0	0	00:00.0
Media	8/7/2023 00:17	8/7/2023 01:08	0	0	0	1/1/2023 00:00

Tabla 30 Data profiling. Análisis fecha/tiempo. Tabla: res_company

5. Capítulo II: Análisis y diseño de la propuesta de solución

a) Metodología de trabajo

A lo largo de los tiempos, las empresas han ido evolucionando su gestión de los procesos de todas sus áreas, y con mayor importancia sobre las áreas administrativas donde sucede la gestión de los recursos humanos, materiales, financieros y de tiempo. Por lo ha sido necesario la mejora de los procesos de gestión de estos, y es ahí donde ha ingresado como un gran apoyo los equipos y software informáticos.

En la actualidad muchas empresas importantes han escalado su gestión de los procesos y de información a su implementación con bases de datos transaccionales, para almacenar y procesar en el día a día las gestiones de sus diferentes áreas, y así mismo al tener tanta información importante pero distribuida surge la necesidad de tener un apoyo para la toma de decisiones, dentro de esta necesidad surge la opción de la creación de Data Warehouse con el fin de mejorar el procesamiento de la información y su posterior análisis para la toma de mejores decisiones basadas en los hechos realizados en su empresa.

Para esta solución se ha tomado a bien utilizar la metodología de Ralph Kimball, que, en pocas palabras, lo que realiza es un análisis del proceso de negocio elegido, y a partir de ello realizar una estructuración de los datos en un modelo de Data Warehouse para su posterior uso en herramientas de análisis.

En esta modelación de datos basado en Ralph Kimball, se enfocará la creación de tablas que serán creadas a partir del análisis de la base de datos transaccional y su posterior procesamiento de la información. En el diseño de las tablas definiremos dos tipos:

- Tabla de hecho: Es aquella tabla que contiene las métricas resultantes para el proceso de negocios.
- Tabla de dimensión: Son tablas desnormalizadas que son fáciles de entender, y que contienen los atributos de las métricas.

Según la metodología de Kimball, se establecen los cuatro pasos para el diseño de un modelo dimensional, que deben de realizarse uno a la vez para construir de manera adecuada y eficiente un Data Warehouse. Los cuatro pasos son los siguientes:

- Selección del proceso de negocio.
- Definición de la granularidad.
- Identificación de las dimensiones.
- Identificación de las métricas.

A partir de los cuatro pasos de la metodología, para la implementación de este proyecto se realizó el proceso siguiente:

- Se realizó un análisis y evaluación de la base de datos transaccional para identificar la información apropiada, para la selección del dataset que abarque el proceso de negocio de análisis financiero.
- Se desarrollaron los cuatro pasos para el diseño del modelo dimensional para el proceso de análisis financiero.
- Se elaboró una matriz de bus para mostrar las dimensiones identificadas que van asociadas a la tabla de hechos.
- Se desarrolló el modelo dimensional de acuerdo con el análisis realizado en los pasos previos.

b) Descripción de la propuesta de solución

Para la creación del modelo dimensional para el proceso de negocio de Análisis financiero de la empresa “Inversiones Digitales” se desarrollaron los cuatro pasos para el diseño del modelo dimensional, detallados a continuación:

1. Proceso de negocio: Análisis financiero.
2. Definición de granularidad: Una fila de datos representa un consolidado de transacciones financieras que representa un ingreso y/o egreso por fecha, sucursal o cuenta.
3. Identificación de dimensiones:
 - DimCuenta: cuentaKey (SK)⁴, cuentald (BK)⁵, codigoCuenta, nombreCuenta, tipoCuenta, grupoCuenta, descripcionCuenta, tieneCuentaPadre, codigoCuentaPadre, nombreCuentaPadre
 - DimSucursal: sucursalKey (SK), sucursalld (BK), nombre.
 - DimFecha: fechaKey (SK), fecha, día, diaNumeroSemana, diaNumeroMes, diaNumeroAnio, mes, mesNumero, numeroSeamanaMes, numeroSemanaAnio, trimestre, trimestreNumero, semestre, semestreNumero, anio, diaLaboral.
4. Identificación de métricas:
 - Ingreso por ventas después de impuesto para un tiempo.
 - El índice de retorno de inversión (ROI) para un período de tiempo.
 - La tasa de crecimiento en utilidades para un período de tiempo.
 - Total, de activos corrientes contra los pasivos corrientes para un período de tiempo.
 - La razón de eficiencia de la empresa de sus cuentas por cobrar contra sus ventas para un período de tiempo.

Matriz de bus

Con la matriz de bus se identifican las dimensiones asociadas al proceso de negocio dentro del Data Warehouse, indicando todas las dimensiones involucradas.

⁴ SK: Surrogate Key o Llave Subrogada. Es la llave de la dimensión.

⁵ BK: Bussiness Key o Llave de negocio. Es la llave asociada a una tabla de la base de datos transaccional.

Dimensiones comunes

Procesos de negocio (Fact table)	DimFecha	DimSucursal	DimCuenta
FactTransaccionFinanciera	X	X	X

Tabla 31 Matriz de bus entre la tabla de hechos y sus dimensiones

Modelo dimensional propuesto

Modelo dimensional para el proceso de negocio de Análisis Finaciero se muestra en la figura 5.

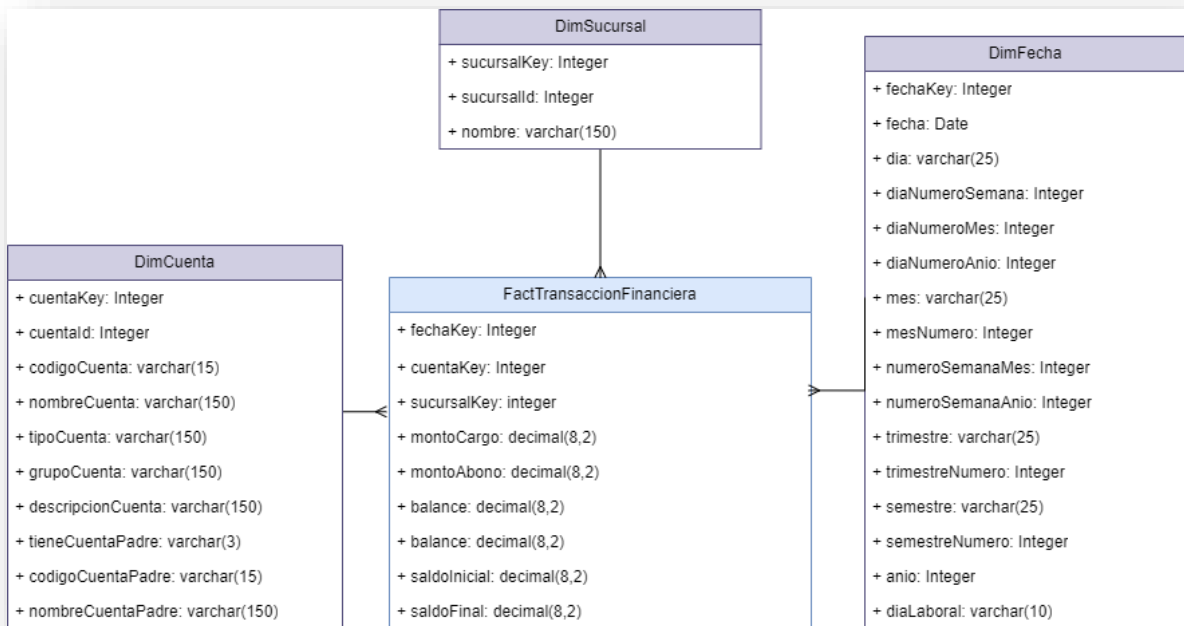


Figura 5 Modelo dimensional propuesto

c) Descripción de la tecnología a utilizar

Odoo

Odoo es un conjunto de aplicaciones de código abierto, software de ERP integrado que cubren todas las necesidades de tu empresa: Comercio electrónico, CRM, contabilidad, inventario, punto de venta, compras, ventas, gestión de proyectos, etc.

La propuesta única de valor de Odoo es ser muy fácil de usar y totalmente integrado. Cuenta con dos versiones, una versión "comunitaria" de código abierto bajo licencia LGPLv3 y una versión empresarial bajo licencia comercial que complementa la edición comunitaria con características y servicios comerciales.

La flexibilidad de Odoo es tal que se pueden añadir aplicaciones dependiendo del crecimiento de la empresa, añadiendo las aplicaciones de una en una a medida que las necesidades de la empresa cambien, estas aplicaciones pueden ser desarrolladas a la medida (lenguaje Python⁶) ya sea por desarrolladores propios o terceros. Gracias a la comunidad de código abierto, Odoo se mantiene totalmente a través de una gran base de desarrolladores, lo que permite responder a las cambiantes necesidades de los clientes y ofrecer aplicaciones nuevas e innovadoras, en algunos casos gratuitas y en las demás por lo general módulos de paga. En la figura 6 se muestra el logo del sistema ERP Odoo.



Figura 6 Logo de Odoo

PostgreSQL

PostgreSQL es una base de datos de código abierto que tiene una sólida reputación entre los sistemas gestores de bases de datos por su fiabilidad, flexibilidad y soporte de estándares técnicos abiertos. A diferencia de otros gestores soporta tipos de datos relacionales y no relacionales. Esto la convierte en una de las bases de datos relacionales más compatibles, estables y maduras disponibles actualmente.

Este gestor de bases de datos es muy utilizado a nivel mundial, para su uso elemental no se requiere aprender más que el lenguaje SQL estándar, por lo que es fácil e intuitivo de aprender a implementar bases de datos y sus datos.

⁶ Python es un lenguaje de programación ampliamente utilizado en las aplicaciones web, el desarrollo de software, la ciencia de datos y el machine learning (ML).

PostgreSQL para la seguridad utiliza una gestión de permisos para usuarios con autenticación por medio de contraseña, para el acceso a las bases de datos como a las tablas contenidas, por lo que al deseárselo se puede permitir únicamente a ciertos usuarios el acceso.

Como dato adicional para el acceso por parte de externos PostgreSQL utiliza para el permiso de acceso una lista blanca (lista que solo permite a ciertas entidades) para el acceso a las bases de datos.

En la figura 7 se muestra el logo de PostgreSQL.



Figura 7 Logo de PostgreSQL

DBeaver

DBeaver es un potente software gestor de bases de datos, de código abierto para múltiples sistemas operativos, que puede usar por la mayoría de los usuarios, tanto para desarrolladores, analistas, administradores de bases de datos y cualquier persona que tengan que trabajar con datos. Es compatible con todas las bases de datos SQL populares como MySQL, MariaDB, PostgreSQL, SQLite, Apache Family, Amazon Redshift, SAP, entre muchos más y en la figura 8 se puede ver el logo de DBeaver.

Con más de 13 años desde su nacimiento, DBeaver ha escalado popularidad por su facilidad de uso y su muy popular diagramador de relaciones ER para visualizar de forma más eficiente las relaciones entre las tablas adyacentes en una base de datos. A lo largo de los años ha ido incorporando compatibilidad con más bases de datos, hasta el punto de poder conectarse a una base de datos de Amazon Redshift.

DBeaver permite una gama de funcionalidades básicas de cualquier gestor de bases de datos y mucho más, de los cuales se pueden listar los siguientes:

- Crear todos los componentes de una base de datos y mostrar el script para generarlos en el CLI: esquemas, tablas, disparadores, funciones, usuarios, roles, etc.
- Definir los tipos de conexiones, si es básica, de entorno de desarrollo o producción.
- Crear/Modificar/Eliminar registros.
- Realizar consultas SQL y NoSQL
- Exportar y migrar datos en diferentes formatos.
- Generar copias de seguridad y restaurar las mismas.

- Visualizar y exportar diagramas del modelo entidad-relación.
- Mantener múltiples conexiones activas a diferentes bases de datos.



Figura 8 Logo de DBeaver

DataCleaner

DataCleaner es un conjunto de herramientas de uso gratuito que ayuda a analizar calidad de datos, permite perfilar, corregir y enriquecer sus datos. La gente lo utiliza para análisis ad hoc, limpiezas recurrentes y como navaja suiza en soluciones de comparación y gestión de datos maestros, como se ve en la figura 9 el logo de DaraCleaner.

El corazón de DataCleaner es un potente motor de creación de perfiles de datos para descubrir y analizar la calidad de sus datos. Encontrando los patrones, valores faltantes, conjuntos de caracteres y otras características de los valores de datos.

Tiene por defecto una serie de herramientas útiles, comenzando por las conexiones a las bases de datos para no necesitar exportar la información, sino más bien establecer contacto directo con la información de una base de datos transaccional, teniendo también uso para archivos exportados como CSV. Esta herramienta está construida para manejar desde cantidades pequeñas (como 15 registros) como a cantidades grandes (1 millón de registros), teniendo en cuenta que a mayor carga de datos se requiere mayor capacidad de recursos en el equipo informático si se desea que muestre resultados rápidos. Además, se pueden diseñar las propias reglas de limpieza de datos para ejecutarlas en diferentes escenarios, por ejemplo, para el reemplazo o búsqueda de datos y uso de expresiones regulares.

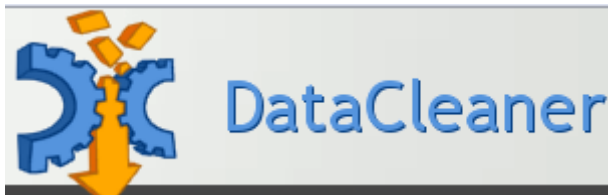


Figura 9 Logo de DataCleaner

Talend Open Studio

Talend Open Studio (TOS en abreviado) es una suite de software que aporta un conjunto muy completo de herramientas de integración de datos que se ofrece en una versión de código libre

Community Edition CE totalmente funcional. Dispone de una gran cantidad de componentes que permite hacer grandes cosas.

Precisamente debido a lo anterior, es una de las herramientas de integración ETL (extract, transform, load) más utilizadas dentro del mundo Big Data. Puede realizar múltiples conexiones a bases de datos, procesamiento de tareas condicionales y de tareas que contienen a otras tareas para la ejecución, conexión con Amazon S3 para subida de datos, entre muchas funcionalidades más, por lo cual deja a la imaginación del usuario la forma de ejecutar sus tareas para el procesamiento de datos, en la figura 10 se puede ver el logo de Talend Open Studio.



Figura 10 Logo de Talend Open Studio

Amazon S3

Amazon Simple Storage Service (Amazon S3) es un servicio de la gama de Amazon AWS, consiste en almacenamiento de objetos que ofrece muchas prestaciones, entre ellas escalabilidad, disponibilidad de datos, seguridad y rendimiento líderes en el sector. Clientes de todos los tamaños y sectores pueden almacenar y proteger cualquier cantidad de datos para prácticamente cualquier caso de uso, como los lagos de datos, las aplicaciones nativas en la nube y las aplicaciones móviles.

Gracias a las clases de almacenamiento rentables y a las características de administración fáciles de usar, es posible optimizar los costos, organizar los datos y configurar controles de acceso detallados para cumplir con requisitos empresariales, organizacionales y de conformidad específicos.

Los datos se almacenan como objetos dentro de recursos llamados “buckets”, y un solo objeto puede tener un tamaño de hasta 5 terabytes. Entre las características de S3, se incluyen capacidades para anexar etiquetas de metadatos a los objetos, mover y almacenar datos entre las clases de almacenamiento de S3, configurar y aplicar controles de acceso a datos, proteger los datos contra usuarios no autorizados, ejecutar análisis de big data y monitorear los datos en los niveles de objeto y bucket, y revisar el uso de almacenamiento y las tendencias de actividad en toda su organización. Se puede acceder a los objetos a través de los puntos de acceso de S3 o directamente a través del nombre de alojamiento del bucket, y en la figura 11 se muestra el logo de Amazon S3.



Figura 11 Logo de Amazon S3

Amazon RedShift

Amazon Redshift es un producto de almacenamiento de datos que forma parte de la plataforma de computación en la nube más grande Amazon Web Services, usado para manejar conjuntos de datos a gran escala y migraciones de bases de datos, en la figura 12 se puede ver el logo de Amazon Redshift.

Redshift se diferencia de la otra oferta de bases de datos alojadas de Amazon, Amazon RDS, en su capacidad para manejar cargas de trabajo analíticas en grandes conjuntos de datos almacenados mediante un principio DBMS orientado a columnas. Redshift permite hasta 16 petabytes de datos en un clúster⁷ en comparación con el tamaño máximo de 128 terabytes de Amazon RDS Aurora

Amazon Redshift utiliza SQL para analizar datos estructurados y semiestructurados en almacenamientos de datos, bases de datos operativas y lagos de datos, con hardware y machine learning diseñado por AWS para ofrecer rendimiento al mejor precio a cualquier escala. Sus precios iniciales van desde los \$0.25 USD por hora⁸.



Figura 12 Logo de Amazon RedShift

⁷ Un clúster es un grupo de colección de recursos informáticos llamados nodos. Que contienen especificaciones personalizadas por el usuario.

⁸ Precios de amazon RedShift: aws.amazon.com/es/redshift/pricing

d) Diagrama arquitectónico de la solución

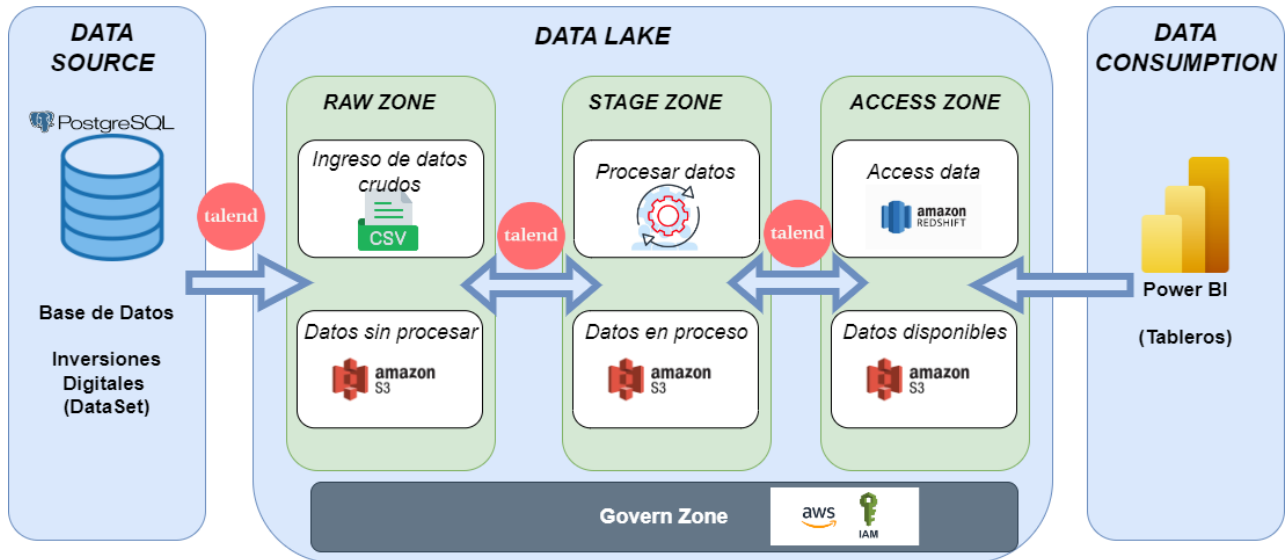


Figura 13 Diagrama arquitectónico de la solución – Arquitectura de Ralph Kimball

En la figura 13 se muestra el Diagrama arquitectónico de la solución basado en la arquitectura de Ralph Kimball.

Data sources: Se refiere a las principales fuentes de datos que alimentan al Data Lake. Estas fuentes de datos por lo general provienen de bases de datos transaccionales o CSV con datos puntuales que sirve como apoyo. Para el caso del proyecto se utilizó la fuente de dataset brindada por la base de datos de la empresa Inversiones Digitales y un CSV adicional referente al tratamiento de fechas.

Procesos ETL (Extracción, Transformación y Carga): El término proviene del inglés Extract, Transform and Load, y se refiere a los procesos con los cuales se realiza el movimiento de datos desde la fuente de datos, así como también los procesos para refactorizar, limpiar y cargar la información en otras fuentes de datos. Para el proyecto se hace uso de la herramienta Talend Open Studio para extraer la información desde la fuente origen en PostgreSQL, así como hacer los procesos correspondientes entre las diferentes zonas que componen el Data Lake.

Data Lake: Es un repositorio diseñado para almacenar, procesar y proteger grandes cantidades de datos, este repositorio debe ser centralizado. Para nuestro proyecto se utilizó Amazon S3 como repositorio para almacenar todos los archivos CSV que vaya generando el proceso de Data Warehouse.

Dentro del Data Lake nos encontramos 4 zonas:

- **Raw Zone:** Área donde se almacenan los datos crudos extraídos de las fuentes de datos origen. Estos datos se almacenan sin ningún tipo de transformación conservando así su formato original, tal cual están almacenados en las fuentes de datos. Para el proyecto se utilizó Amazon S3 dentro de la carpeta denominada "RAW", donde se guardan los datos crudos en formato CSV.

- **Stage Zone:** En esta área se almacenan los datos que están siendo procesados o han sido partícipes de alguna transformación final (cambiado de tipo de dato, formato del dato, renombrado de valores, entre otros), estos datos serán utilizados para la construcción de las respectivas tablas de hechos y dimensiones. Para el proyecto se utilizó Amazon S3 dentro de la carpeta denominada “Process”, donde se guardan los datos en proceso en formato CSV.
- **Access Zone:** Área donde se almacenan los datos que han sido procesados y se encuentran listos para su respectivo consumo por aplicaciones de Inteligencia de Negocios (BI). Estos datos ya están limpios, filtrados, transformados y organizados en la respectiva tabla de hechos (Fact Table) y dimensiones (Dimension Tables). Para el proyecto se utilizó Amazon S3 dentro de la carpeta denominada Access, donde se guardan los datos listos para ser consumidos en formato CSV.

Access Data: Dentro del área de Access Zone se hace uso de la herramienta Amazon Redshift la cual se encarga guardar la estructura de base de datos del modelo dimensional y sus datos. A través de esta herramienta se da acceso a los datos para que las aplicaciones de BI puedan consumirlos.

- **Govern Zone:** Área donde se administran las políticas, reglas y permisos de usuarios que podrán acceder a los datos que se utilizarán para la construcción del Data Warehouse, como a la administración de las diferentes herramientas utilizadas dentro del proceso. Es la capa de seguridad que se aplica en cada área que compone el Data Lake involucrado con Amazon S3 y Amazon RedShift. Para el proyecto se hace el uso de la herramienta Amazon IAM⁹.

Data consumption: Área donde podemos encontrar a las aplicaciones BI, quienes tienen la función de consumir los datos procesados por el Data Warehouse para dar respuesta a los requerimientos analíticos del proceso de negocio. Para el proyecto se hace uso de la herramienta Power BI en la cual se generan los respectivos tableros para cada métrica del proceso de negocio donde se analizan y diseñan los gráficos y reportes a presentar.

e) Descripción de cada componente de la solución

En este apartado se llevará a cabo la descripción y demostración de algunos de los componentes descritos anteriormente en el diagrama arquitectónico, además de algunos ejemplos de los procesos ETL que fueron realizados para el desarrollo del presente proyecto.

DATA SOURCE: Conexión a la base de datos origen.

⁹ Amazon IAM: Amazon Identity and Access Management, en español “Administración de identidades” de Amazon.

En esta sección como se ha mencionado en el documento, la fuente de origen de datos se encuentra alojada en el gestor de bases de datos PostgreSQL con el nombre de base de datos “inversiones_digitales_sv”.

Por lo que se procede a utilizar esos datos para establecer una conexión a través del software de Talend Open Studio.

En esta herramienta tiene diferentes opciones para guardar los componentes como los siguientes: Conexiones a bases de datos, ficheros de todo tipo, entre otros como se puede visualizar en la figura 14. Pero las que ocuparemos es la opción llamada “Metadatos” la cual se ocupará para almacenar la conexión con la base de datos de PostgreSQL, la conexión con S3, y con los ficheros CSV que se utilizarán a lo largo del proyecto, en la figura 15 podemos visualizar la configuración de la conexión de la base de datos.

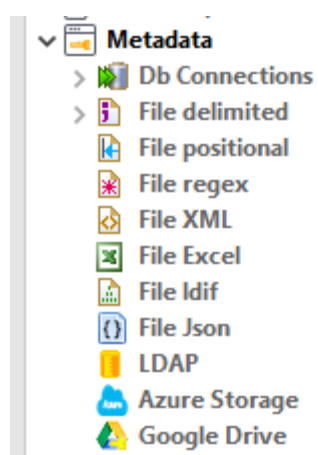


Figura 14 Conexiones de metadatos dentro de Talend Open Studio

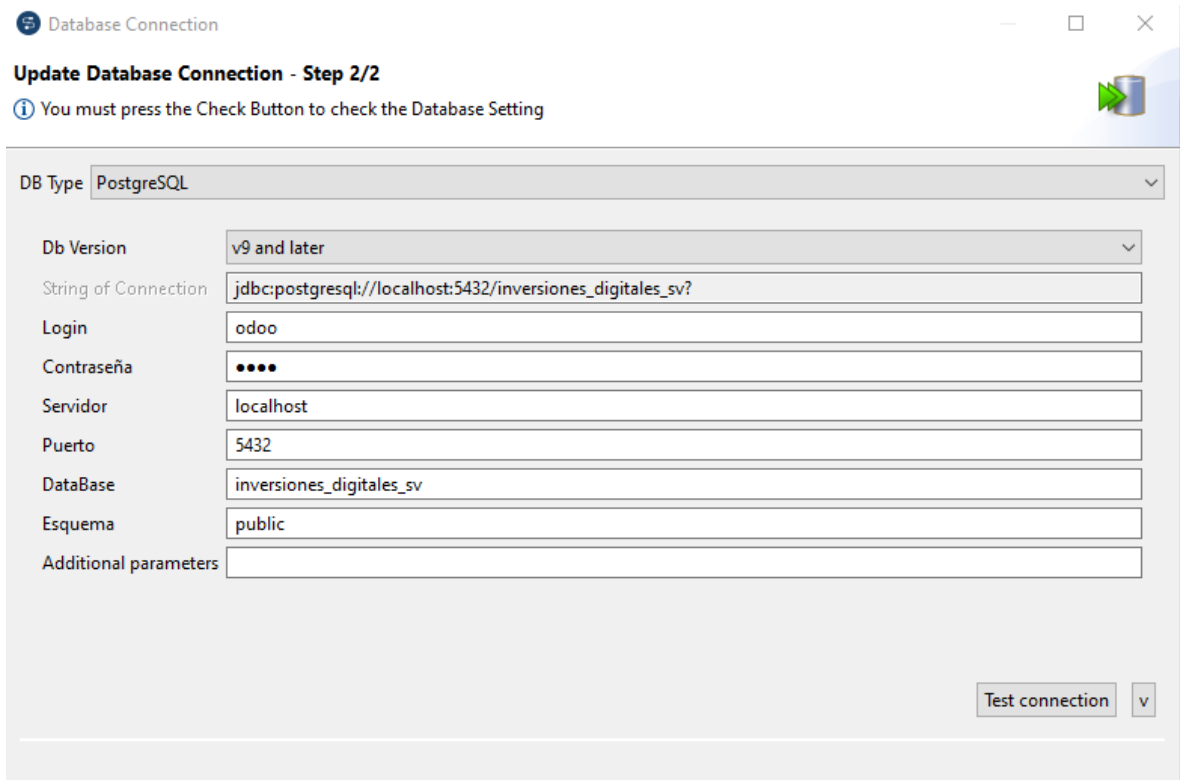


Figura 15 Conexión de Talend Open Studio hacia la base de datos de PostgreSQL

ETL realizados en el proyecto

Para la ejecución de los ETL que se describirán, la ejecución se programa para cada sábado a las 01 horas, con el fin de realizarse en el horario menos congestionado para la empresa.

ETL de extracción de información

Según la metodología escogida (Kimball), lo que se debe realizar primeramente es la extracción de la información desde la base de datos transaccional, realizando la extracción de los datos de forma íntegra y sin modificación alguna.

El proceso de extracción se realizó para cada una de las tablas que componen el Dataset como se puede ver en la figura 16.

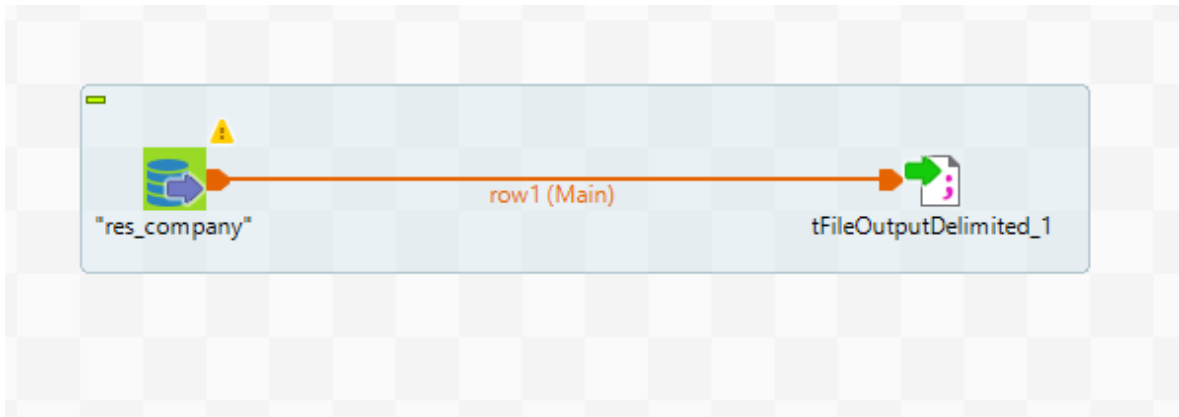


Figura 16 ETL de extracción de información. res_company

Este proceso está compuesto por los siguientes componentes:

- tDBInput, este refiere a la conexión de base de datos antes realizada, especificando el nombre de la tabla se desea extraer, además de especificar los campos con la sentencia SQL y obtener el esquema con la opción "Guess Schema". En la *Figura 17* podemos observar el tDBinput con la sentencia de extracción de los datos.

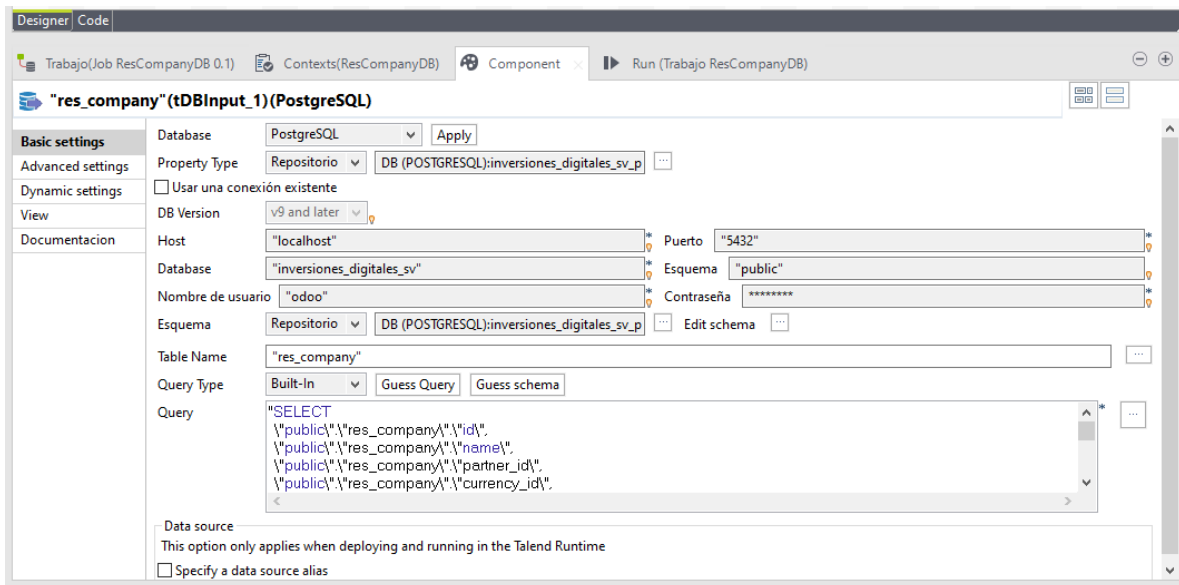


Figura 17 tDBinput-Extracción de los datos

- tFileOutputDelimited, este componente es el encargado de generar el archivo CSV al cual se le debe especificar una ruta donde va a almacenarse. En la *Figura 18* podemos observar los parámetros del componente con la especificación de salida.

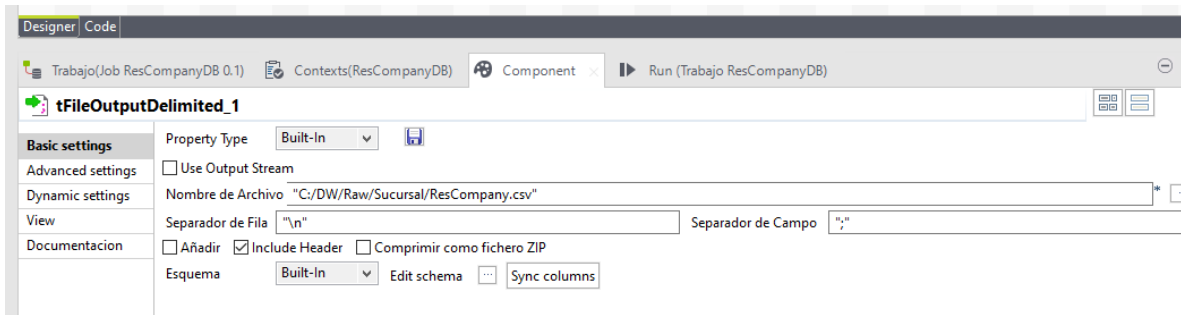


Figura 18 tFileOutputDelimited -Extracción de los datos

ETL de limpieza y transformación de datos

Posteriormente de la extracción de la información se procede realizar el siguiente paso que es la limpieza y transformación de la información, donde cada tabla es tratada por lo general de diferente forma dependiendo del contenido, por ejemplo, la limpieza de nullos, o selección de datos que serán de utilidad y descartando los que no, en otros casos de que no sea requerido realizar ningún proceso de limpieza y transformación debido a que los datos ya vienen limpios o se usarán como los proporciona el DataSet origen.

El proceso de limpieza y transformación que se muestra en la figura 19 es para el caso de la tabla `account_account_tag`, donde no fue requerido una transformación de datos, cabe aclarar que este es un ejemplo de uno de tantos procesos realizados.

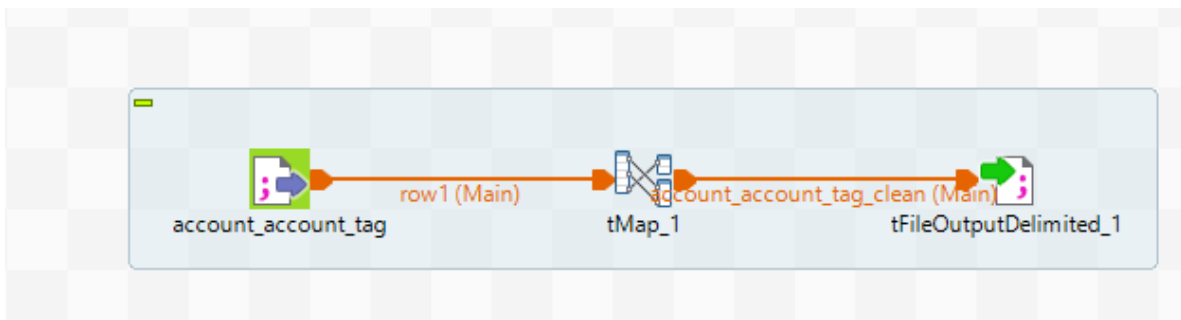


Figura 19 ETL limpieza y transformación. `account_account_tag`

Este proceso está compuesto por:

- tFileInputDelimited llamado “`account_account_tag`” el cual es un archivo de entrada CSV que fue previamente extraído en el paso anterior, en pocas palabras, sin ninguna modificación.
- tMap componente donde se pueden realizar variedad de procesos entre los cuales están: el filtrado de campos y registros, el uso de variables y expresiones intermedias, transformación de tipos de datos, entre otros. Se puede observar en la figura 20 para el caso de la tabla seleccionada.

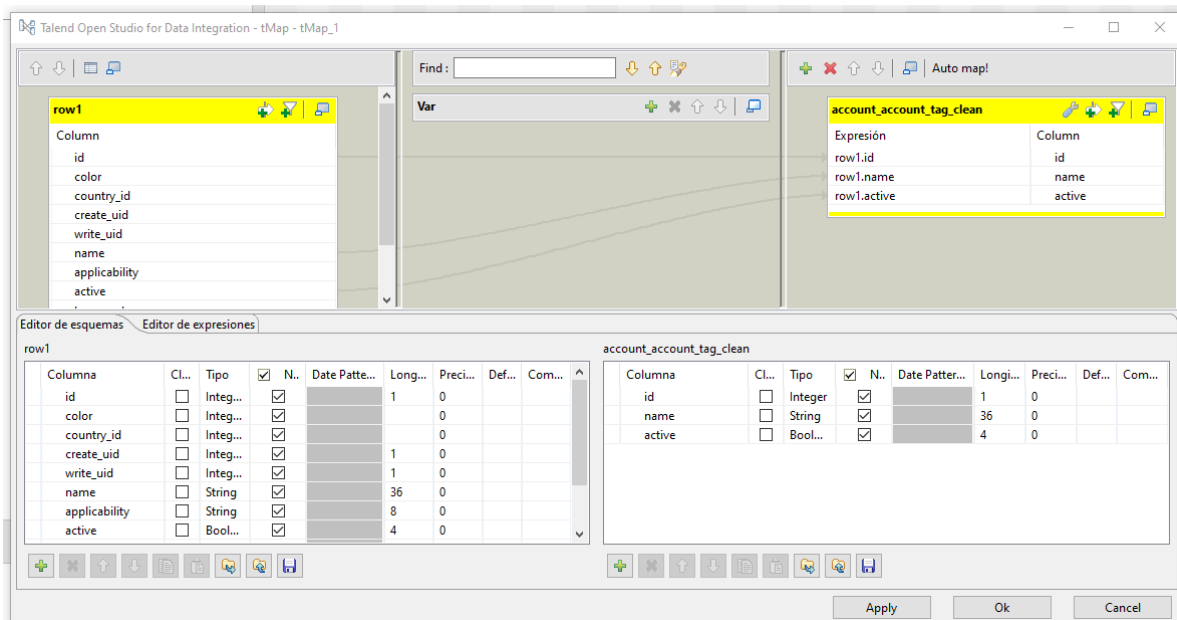


Figura 20 tMap, proceso de transformación account_account_tag

- tFileOutputDelimited, este componente como se ha visto previamente genera un archivo CSV con los datos limpios y transformados.

ETL de cálculos y preparación de datos

Al haber ejecutado los debidos procesos de limpieza y de transformación de datos se realizan los procesos de cálculos y preparación de datos los cuales se basan en generar información a través de expresiones lógicas o matemáticas, volver a filtrar campos o registros, realizar uniones (joins) entre tablas, ejecutar Jobs a detalle, entre otros procesos que se tuvieron que realizar a algunas tablas para obtener los datos listos para su consumo por aplicaciones de BI.

Los procesos ETL que agrupan esta parte, son los procesos más largos y de los que se requirió mayor análisis por parte del equipo, debido a que, por la complejidad de algunos datos y la normalización de la base de datos origen, se requirió realizar varios procesos para lograr llegar a la forma final de las tablas de hechos y dimensiones.

El proceso de cálculo y preparación de datos que se muestra en la *Figura 21 es para el caso del proceso de DimCuenta.*

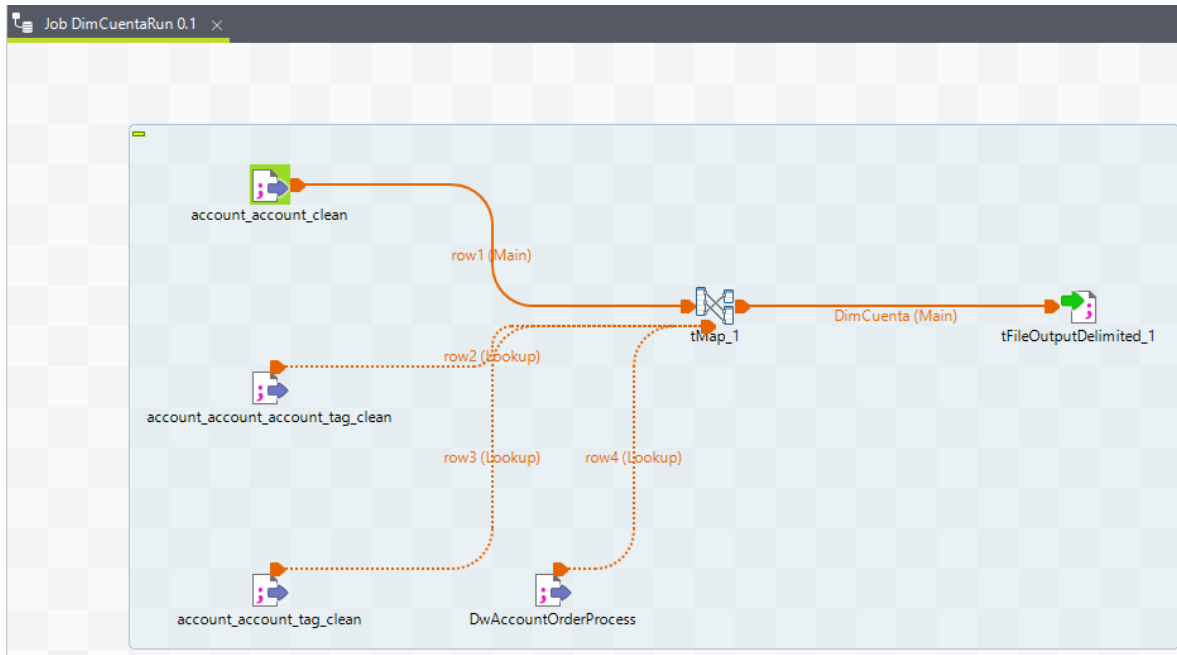


Figura 21 ETL Cálculo y preparación de datos. DimCuenta

Este proceso se encuentra formado por:

- cuatro fFileInputDelimited los cuales son cuatro archivos CSV limpios y transformados.
- tMap con el cual se realiza para este caso de la siguiente manera, como se visualiza en la figura 22.

The screenshot shows the configuration for the 'tMap' component. On the left, the 'row1' configuration is visible, showing columns: id, code, name, account_type, internal_group. The 'row2' configuration is also visible, showing columns: id, name. The 'Var' section contains the following expressions:

Expresión	Tipo	Variable
Numeric.sequence("1", "1")	int	cuentalKey
(row3.name == "" row3.name == null) ? "Sin Descri..."	String	descripcion
(row4.has_parent == true) ? "SI" : "No"	String	tieneCuentaPadre

At the bottom, the 'DimCuenta' column mapping table is shown:

Columna	CL	Tipo	N.	Date Pattern (Ctrl+Spac...)	Longitud	Precision	Default	Comentario
cuentalKey	<input checked="" type="checkbox"/>	int						
cuentalId	<input type="checkbox"/>	int			2	0		
codigoCuenta	<input type="checkbox"/>	int			6	0		
nombreCuenta	<input type="checkbox"/>	String			35	0		
tipoCuenta	<input type="checkbox"/>	String			21	0		
grupoCuenta	<input type="checkbox"/>	String			9	0		
descripcionCuenta	<input type="checkbox"/>	String			36	0		
tieneCuentaPadre	<input type="checkbox"/>	String						
codigoCuentaPadre	<input type="checkbox"/>	String						

Figura 22 tMap DimCuenta

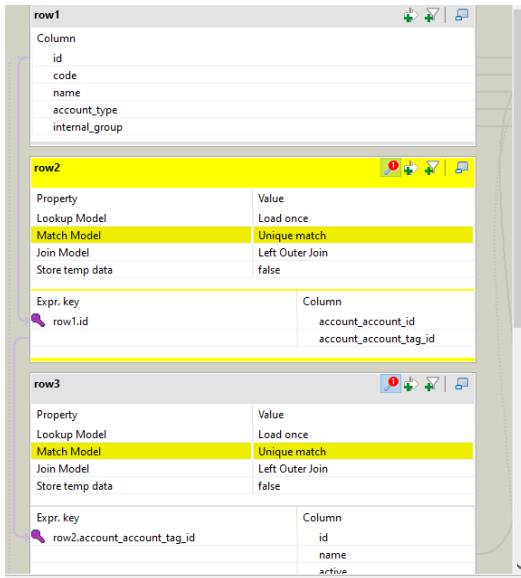


Figura 23 tMap ejemplo de Inner Join. DimCuenta

Proceso de Left Outer Join (como se muestra en la figura 23) para unir los archivos con la información requerida.

Así mismo como ejemplo se muestra en la figura 24 una expresión de ejemplo para evaluar si el registro tiene el campo has_parent como "true".

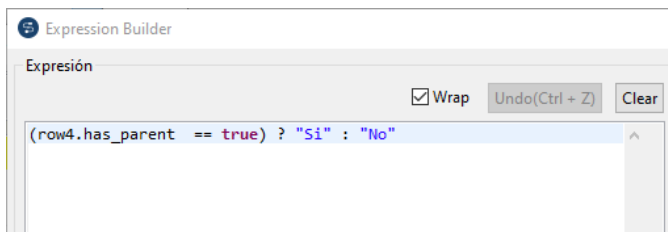


Figura 24 Tmap-Ejem. Expresión

- tFileOutputDelimited, este componente es el encargado de generar el archivo CSV (DimCuenta.csv) con los datos, para este caso preparado para el consumo.

Este mismo proceso se realizó en las demás tablas y fases para llegar a lo que sería el desarrollo del Data Warehouse y la presentación final de la tabla de hechos y las tablas de dimensiones correspondientes.

ETL: Carga y descarga de datos a Amazon S3

- ETL de descarga de archivos de S3

El proceso de descarga de datos se realiza en la zona de Stage y Access, debido a que estas zonas requieren la información de las zonas que les preceden, por ejemplo, la zona Access requiere de los archivos subidos a la Zona Stage.

Un ejemplo del proceso de descarga de archivos de S3 se muestra en la figura 25, para la zona de Access para el caso de la DimCuenta que requiere de los archivos dentro de la carpeta Process/ almacenada en S3.

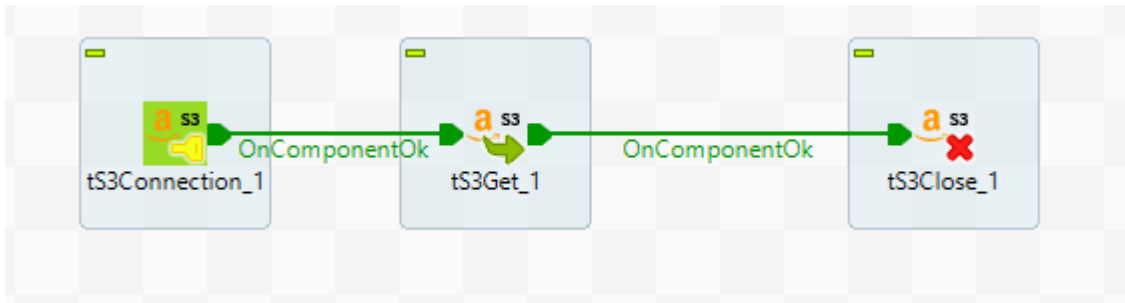


Figura 25 ETL-Descarga de archivo de S3

El proceso este compuesto por los siguientes componentes:

- tS3Connection, este elemento sirve para configurar la conexión con Amazon S3 colocando la llave de acceso, la llave secreta y la región del repositorio.
- tS3Get, en este componente debe de especificarse desde cuál bucket o repositorio se ira a descargar la información, el nombre del archivo CSV que se obtendrá y en que carpeta del sistema operativo se almacenará. Para el caso de descarga, hace referencia a la carpeta Process/Cuenta para obtener la información y guardarla en la carpeta Access/S3DownloadProcess como puede visualizarse en la figura 26.
- tS3Close, este componente tiene la función de cerrar la conexión abierta con Amazon S3.

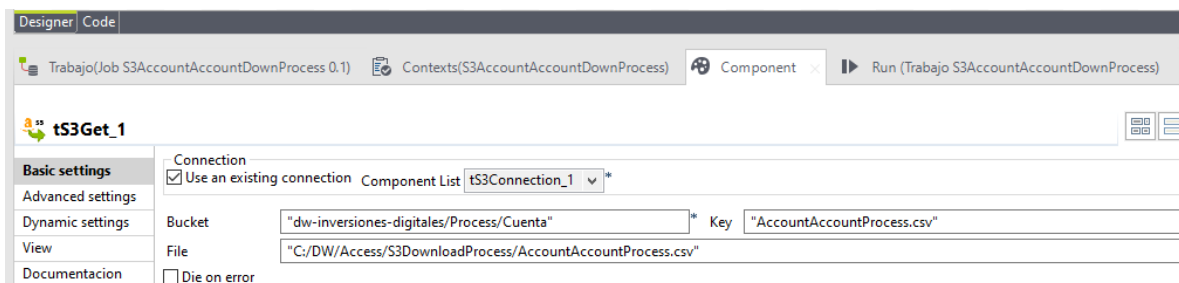


Figura 26 tS3Get

Cabe destacar que para cada archivo se realizó un Job individual, por lo que para descargar todos los relacionados con la DimCuenta se realizó otro job que ejecuta en serie la descarga de los archivos, esto con el fin de una mayor trazabilidad en caso de errores. Como se puede mostrar de ejemplo en la figura 27.

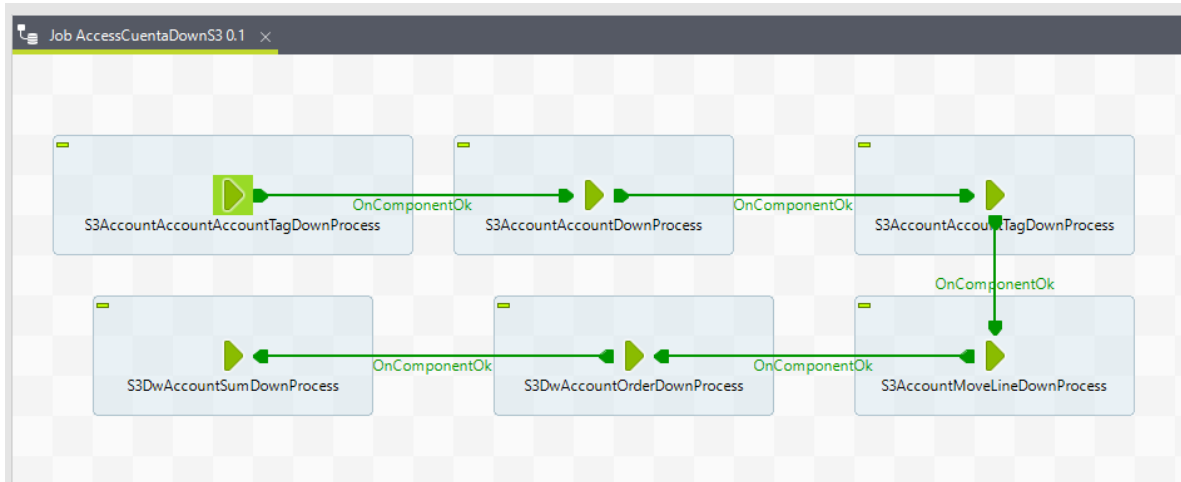


Figura 27 Job para la descarga de archivos de la Zona Process para la DimCuenta

- ETL de carga de archivos de S3

En cada zona que componen el Data Lake (RAW, Stage y Access) se ha realizado este proceso el cual consiste en cargar los datos, de la zona que esté siendo procesada, y almacenarlos al repositorio de datos en Amazon S3. Este proceso se realiza luego de:

- Finalizar los procesos de ETL: extracción de datos (procesos entre BD transaccional - RAW)
- Finalizar los procesos de ETL: limpieza y transformación (procesos entre RAW - Stage)
- Finalizar los procesos de ETL: cálculos y preparación de datos (procesos entre Stage - Access)

El proceso de carga de archivos a S3 que se muestra en la Figura 28, en este caso es para subir lo procesado en la zona de Stage para la tabla *res_company*.

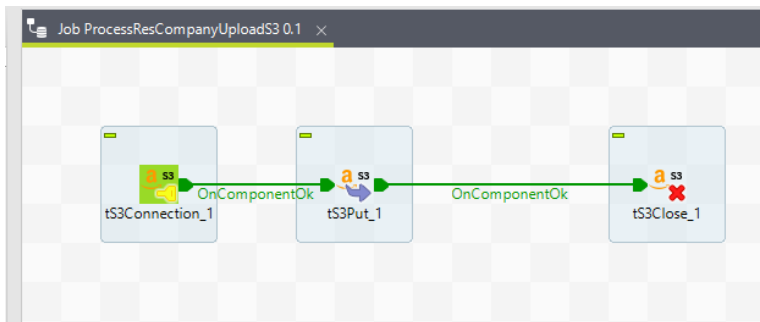


Figura 28 ETL de carga de archivo a S3

Este proceso está compuesto por los siguientes componentes:

- tS3Connection, este componente sirve para configurar la conexión con Amazon S3 colocando la llave de acceso, la llave secreta y la región del repositorio, en la figura 29.

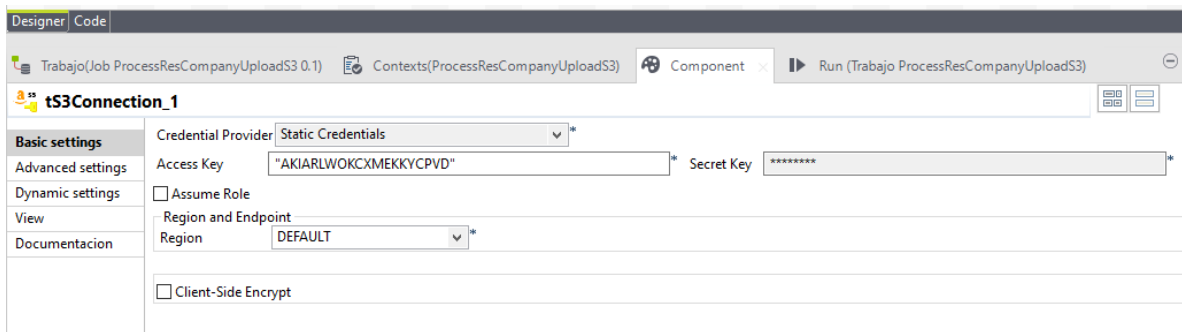


Figura 29 tS3Connection para subir a S3

- tS3Put, en este componente se debe configurar en que bucket se almacenará el dato. Para ello se especifica el nombre del bucket o repositorio, en este caso se incluye el nombre de la carpeta y se coloca el nombre que tendrá el archivo en S3, el cual debe ser único, además, se debe colocar desde que carpeta dentro del sistema operativo se va a subir. Para el caso de carga, hace referencia a la carpeta Process/Sucursal, donde se almacenan los datos generados en la zona Stage, se puede observar en la *figura 30*.

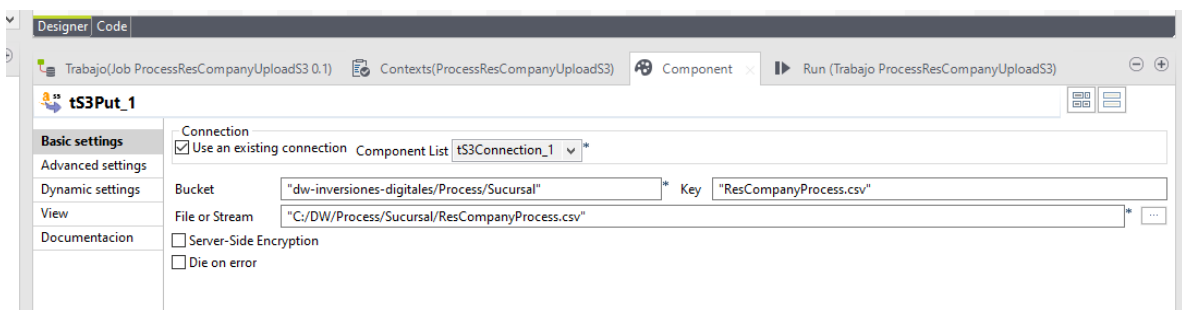


Figura 30 tS3Put para S3

Este proceso se repite de igual manera para la subida de los demás archivos CSV generados en Talend Open Studio, para las tres fases.

En caso para simplificar las pruebas se diseñaron Jobs que unificaran los procesos individuales como se puede visualizar en la figura 31.

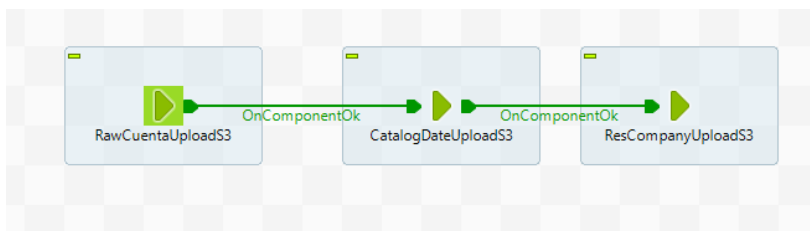


Figura 31 ejemplo de Job que sube todos los archivos a la zona RAW en S3

ETL: Carga de archivos a RedShift

Este proceso es un tanto diferente ya que consiste en la carga hacia Amazon RedShift de la información limpia de la tabla de hechos y las tablas dimensionales que se encuentran almacenadas en S3.

Para esto se utiliza el comando COPY el cual es directo de RedShift y su estructura es la siguiente:

```
COPY [nombre de la tabla] FROM ['URI de objeto en S3']  
CREDENTIALS ['llave de acceso del usuario'] ['llave secreta del usuario']  
DELIMITER ','  
IGNOREHEADER 1;
```

COPY: comando para la carga de la información de los archivos CSV alojados en S3 (en la Zona Access) hacia Redshift.

[nombre de la tabla]: nombre de la tabla en Redshift, para esto es necesario haber ejecutado el script que genera las tablas dentro de RedShift ubicado en Anexos (Anexo 4).

['URI de objeto en S3']: identificador del recurso que se quiere cargar en Redshift.

CREDENTIALS: esta contiene la llave de acceso y llave secreta de acceso del usuario asignado.

DELIMITER: delimitador la información del recurso que se quiere cargar. Se coloca el carácter que hace de delimitador.

IGNOREHEADER: para ignorar los encabezados.

El comando de la *Figura 32 es correspondiente* a la tabla dimFecha. Este mismo proceso se utilizó para llenar cada una de las demás tablas de dimensiones y la tabla de hechos.

```
1 --migrate data from S3 to redShift  
2 COPY dimFecha FROM 's3://dw-inversiones-digitales/Access/DimFecha/DimFecha.csv'  
3 CREDENTIALS 'aws_access_key_id=AKIARLNOKCXMEKKYCPVD;aws_secret_access_key=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'  
4 DELIMITER ','  
5 IGNOREHEADER 1;
```

Figura 32 Comando Copy en RedShift

Data Lake: Repositorio en AWS

El repositorio de los datos se encuentra alojado en el servicio de Amazon S3 y tiene de nombre “dw-inversiones-digitales”. Se creó una estructura de carpetas para una mejor organización de la información almacenada, en la figura 33 se mira de forma lógica y en la figura 34 como se visualiza en S3.

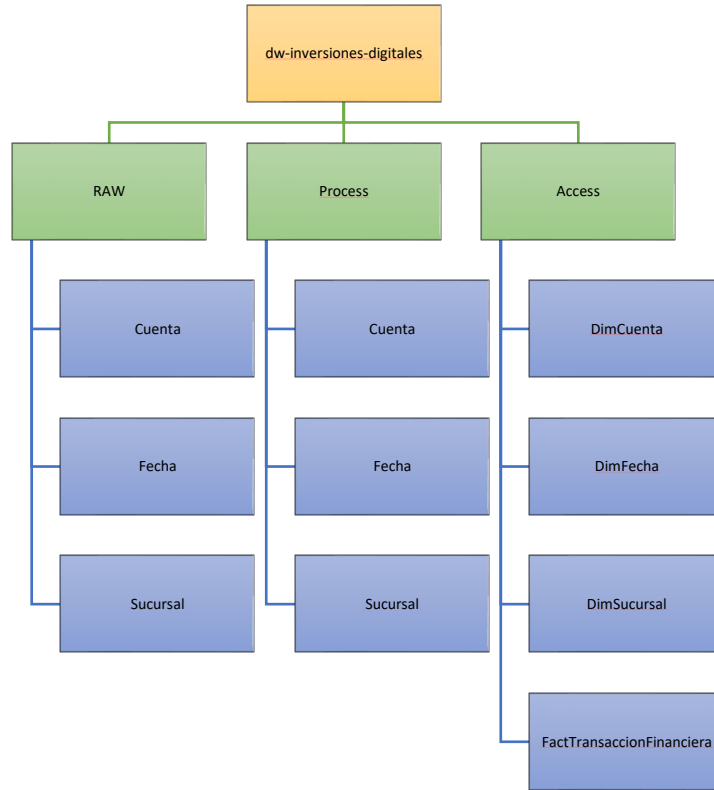


Figura 33 Estructura de carpetas dentro de S3

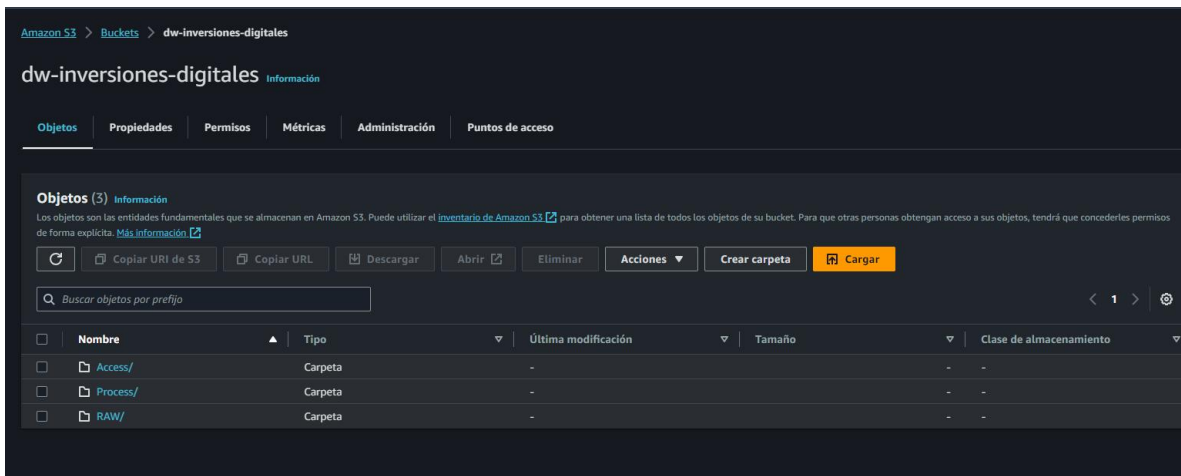


Figura 34 Repositorio en S3

En la RAW Zone (carpeta RAW) se almacena la información de forma cruda según fue extraída del origen, luego del proceso ETL de extracción de datos, se realiza el ETL de carga de datos a S3 como se puede visualizar en la figura 35.

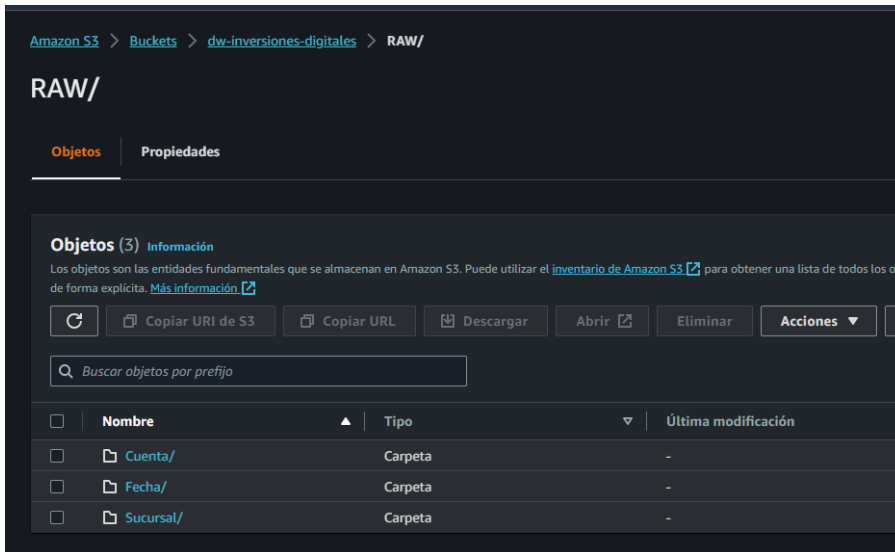


Figura 35 RAW Zone en S3

En la Stage Zone (carpeta Process) se almacena la información que ha sido procesada, luego del proceso ETL de limpieza y transformación de datos, se realiza el ETL de carga de datos a S3 como se puede visualizar en la figura 36.

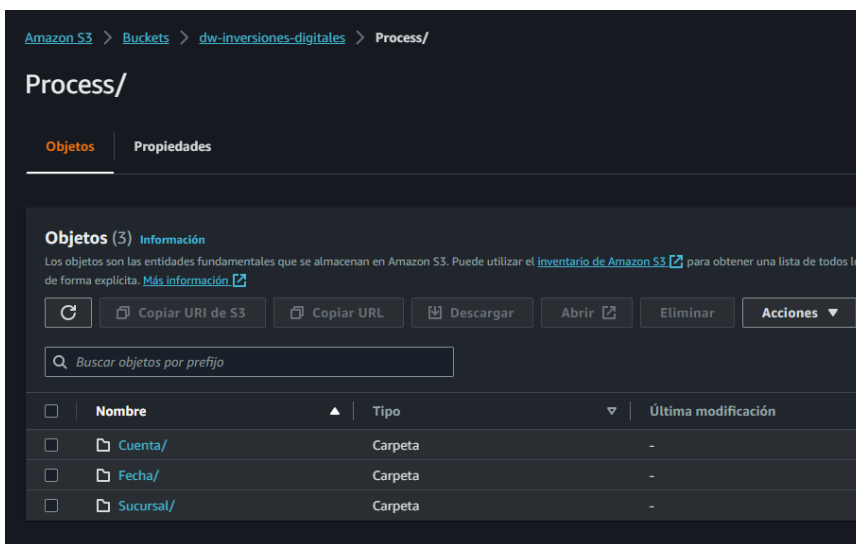


Figura 36 Stage Zone en S3

En la Access Zone (carpeta Access) se almacena la información que ha sido procesada para generar las dimensiones, luego del proceso ETL de cálculos y preparación de datos, se realiza el ETL de carga de datos a S3 como se puede visualizar en la figura 37.

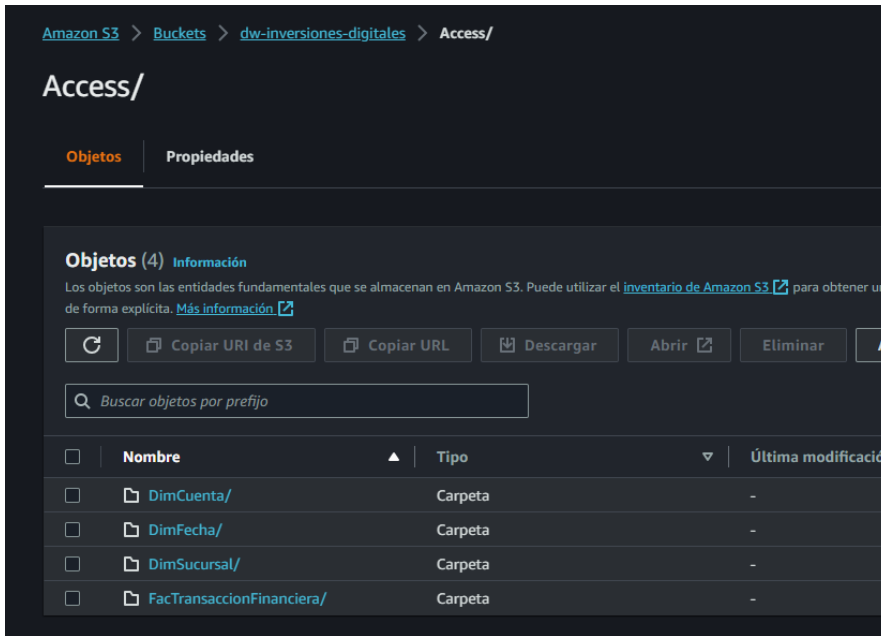


Figura 37 Access Zone en S3

Access Data: RedShift.

Amazon Redshift es la herramienta que se utilizó para guardar la estructura del modelo dimensional y dar acceso a la aplicación de BI, para ello se ha ejecutado el script que se encuentra Anexo 4 y posteriormente se ejecutaron los procesos ETL de carga de archivos a Redshift para llenar su estructura. En la figura 38 se puede observar el modelo dimensional en Redshift.

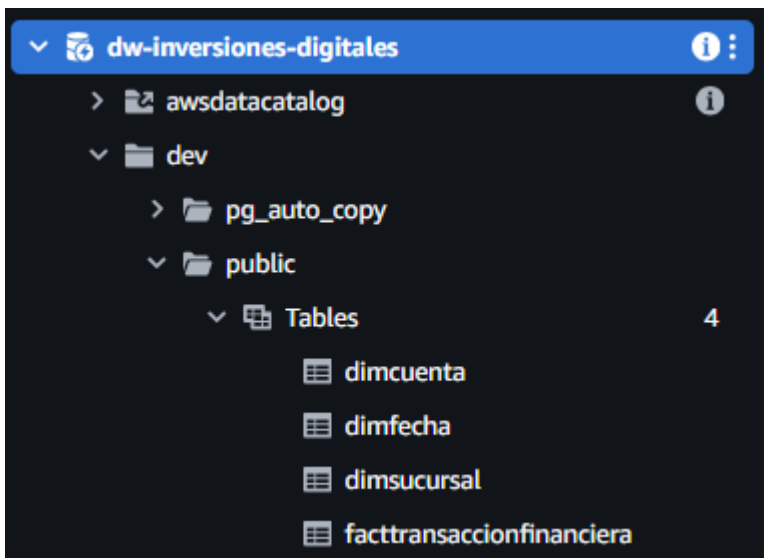


Figura 38 Modelo dimensional en Redshift

Govern Zone: Configuración de Amazon IAM.

Para llevar un mejor control de acceso a los servicio e información almacenada en el bucket, se crearon diferentes grupos usuarios a los cuales se les asignaron políticas diferentes (Asociadas a sus nombres), como se puede visualizar en la figura 39 y los usuarios que fueron creados en la figura 40.

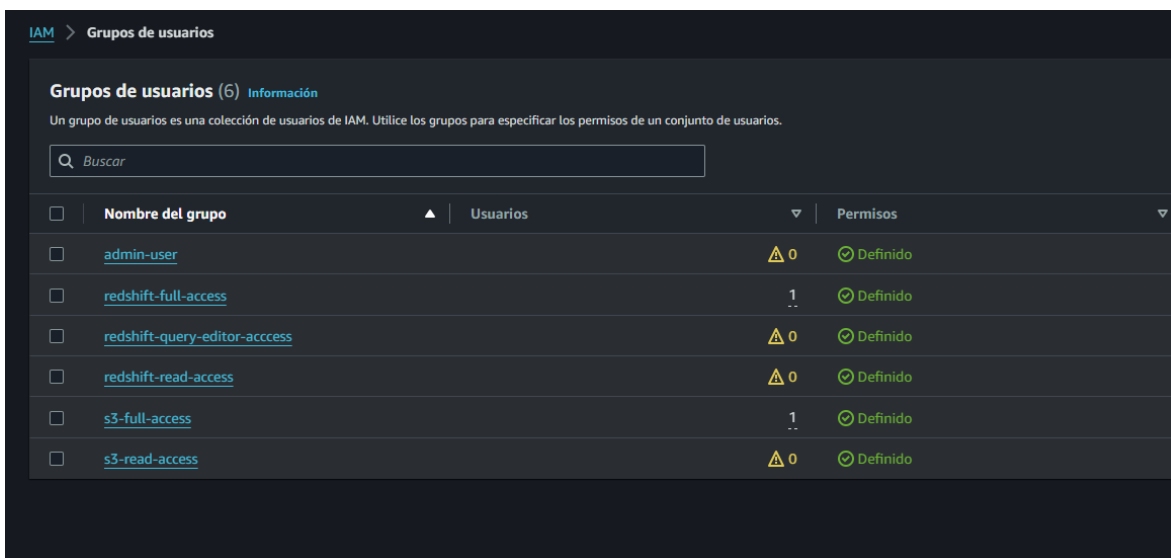


Figura 39 Grupos de usuarios en S3

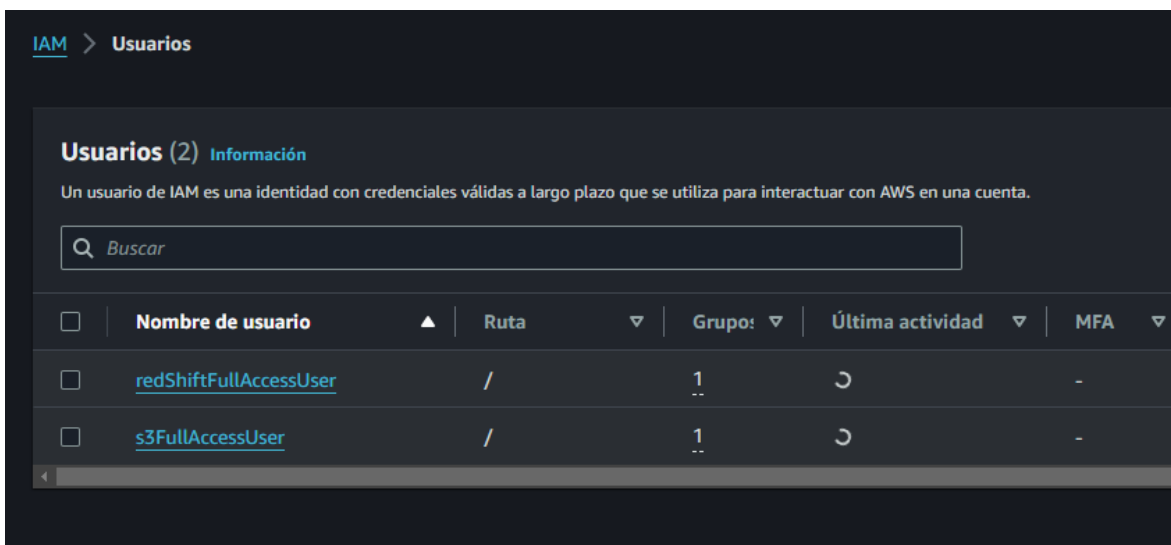


Figura 40 Usuarios en S3

- Nombre de usuario:** redShiftFullAccessUser
Descripción: Administrador general de RedShift
Política: AmazonRedshiftFullAccess
Descripción de la política: Proporciona acceso completo al servicio de RedShift.
- Nombre de usuario:** s3FullAccessUser
Descripción: Administrador de Amazon S3

Política: AmazonS3FullAccess

Descripción de la política: Proporciona acceso completo a todos los repositorios de S3.

Data consumption: Conexión a Power BI

Para la parte final, se procede a realizar el consumo de datos que están alojados en RedShift, desde Power BI, dando clic al ícono de “Obtener Datos” y buscar la opción de “Amazon RedShift” para poder ingresar credenciales para conectarse. En la figura 41 y 42 puede visualizarse los parámetros realizados para establecer la conexión con el servidor de Amazon RedShift.

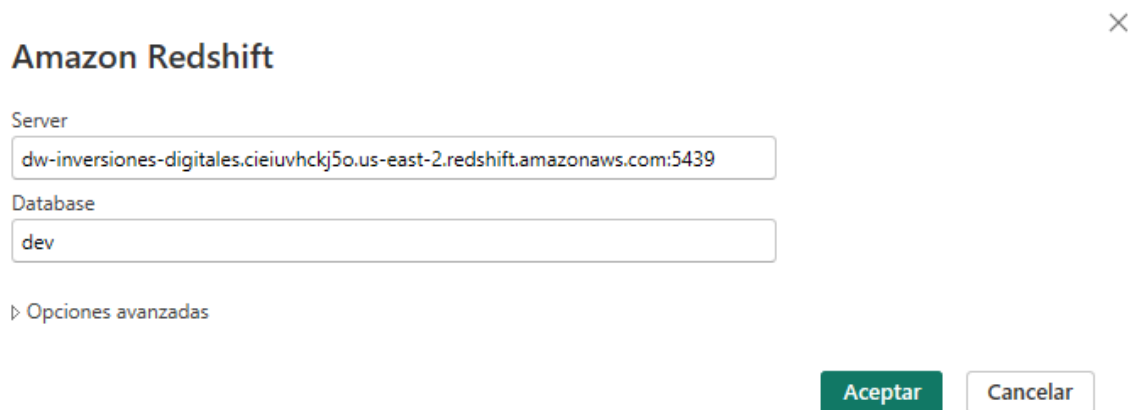


Figura 41 Power BI . Conexión a servidor RedShift



Figura 42 Power BI, ingreso de credenciales de RedShift para obtener información.

6. Capítulo III: Estrategia de implementación de la propuesta de solución

a) Estrategia de implementación

Una estrategia es un plan integral diseñado para alcanzar un conjunto específico de objetivos o metas. En un contexto empresarial, la estrategia guía la toma de decisiones y la asignación de recursos para lograr ventajas competitivas, mejorar el rendimiento y enfrentar los desafíos. Puede abordar diversos aspectos, como el crecimiento, la eficiencia operativa, la innovación, la expansión de mercado, entre otros.

Es así como por medio de un análisis exhaustivo del entorno interno de la empresa Inversiones Digitales por medio de sus antecedentes y teniendo una visión clara de la problemática que la misma enfrenta para poder tomar decisiones de forma más acertada por medio del análisis de métricas claves se han definido los de manera precisa los objetivos que se buscan lograr.

Teniendo en mente los resultados que requiere obtener se realizó la programación de las actividades necesarias como se puede ver en el Anexo 2 de este documento. A partir de esto, se muestra a continuación las siguientes actividades que se realizaron.

Creación de modelo dimensional

Utilizar la metodología de Kimball para diseñar un modelo dimensional que refleje de manera efectiva la estructura y las relaciones de los datos financieros de la empresa. Definir dimensiones críticas como cuentas, sucursal y tiempo, y establecer hechos clave como ingresos, costos y beneficios netos.

Instalación de herramientas de Extracción, Transformación y Carga (ETL)

Implementar Talend Open Studio para gestionar el proceso de ETL. Utilizar Talend para extraer datos de diversas fuentes internas y externas, transformarlos según las necesidades del modelo dimensional y cargarlos en el Data Warehouse además de garantizar la calidad y consistencia de los datos durante el proceso de ETL.

Por lo que el primer paso es verificar que se tiene **Java** instalado ya que Talend Open Studio hace uso de ella, en caso de no tenerlo instalado se deben seguir los siguientes pasos:

1. Descargar Java:

- Visitar el sitio web oficial de Oracle para Java¹⁰
- Hacer clic en "JDK Download" para acceder a la página de descarga.
- En la página de descarga, aceptar los términos y condiciones.
- Seleccionar la versión de Java que se desea instalar (por ejemplo, JDK 8 o JDK 11, siendo ésta última la versión recomendada para usar la versión 8.0 de Talen Open Studio) y hacer clic en el enlace de descarga para el sistema operativo (Windows x64 en la mayoría de los casos).

2. Descargar e Instalar:

- Una vez descargado el instalador, ejecutar haciendo doble clic.

¹⁰ <https://www.oracle.com/java/technologies/downloads/#jdk21-windows>.

- Seguir las instrucciones del instalador. Se puede dejar las opciones predeterminadas en la mayoría de los casos.

3. Configurar las Variables de Entorno:

- Abrir el "Panel de control" y buscar "Variables de entorno" en la barra de búsqueda.
- Hacer clic en "Editar las variables de entorno del sistema".
- En la sección "Variables del sistema", hacer clic en "Nuevo" y agrega una nueva variable con el nombre JAVA_HOME y la ruta donde se instaló Java (por ejemplo, C:\Program Files\Java\jdk-17).
- Buscar la variable Path en la lista de variables, hacer clic en "Editar" y agregar %JAVA_HOME%\bin al final.

4. Verificar la Instalación:

- Abrir una nueva ventana de comandos y escribir java -version y javac -version. Se debería visualizar información sobre la versión de Java y el compilador Java.

Luego de verificar que se tiene Java instalado se procede a realizar la instalación **de Talend Open Studio** como se indica a continuación:

Paso 1: Descargar Talend Open Studio

- Visitar el sitio web oficial de Talend: Talend Open Studio¹¹.
- Hacer clic en el botón de descarga. Dependiendo del sistema operativo, seleccionar la versión adecuada para Windows, Linux o macOS, como se indica en la figura 43.

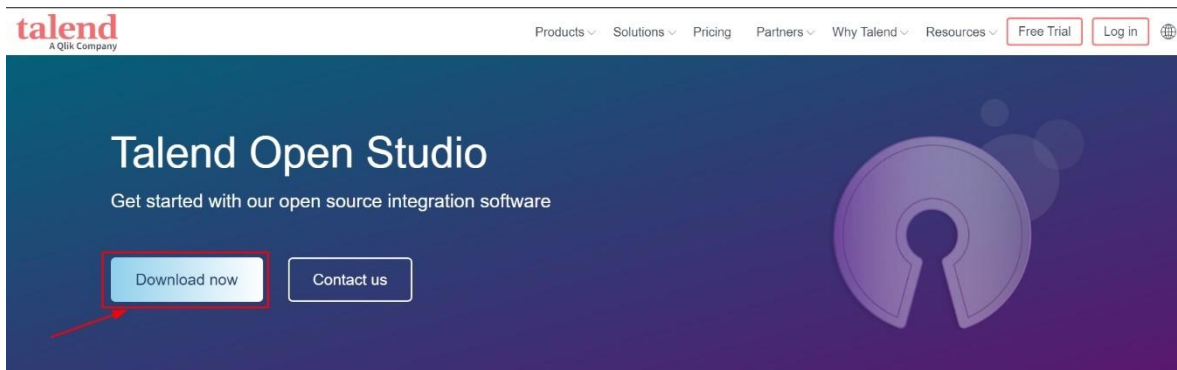


Figura 43 Descarga de Talend Open Studio

- Para poder realizar la descarga se deberá completar el formulario para crear una cuenta como se muestra en la figura 44.

¹¹ <https://www.talend.com/products/talend-open-studio/>

Open Studio for Data Integration

Talend Open Studio for Data Integration is free-to-download software to kickstart your first data integration and ETL projects.

Features:

- Free open source Apache license
- RDBMS connectors: Oracle, Teradata, Microsoft SQL server
- SaaS connectors: Marketo, Salesforce, NetSuite
- Packaged apps: SAP, Microsoft Dynamics, Sugar CRM

Get Started for Free

* First Name

* Last Name

* Business Email

* Company Name

* Phone Number

* Job Title

* Job Function

* Country

[Access Now](#)

Figura 44 Formulario de registro para descarga de Talend Open Studio

Paso 2: Instalar Talend Open Studio en Windows

- Una vez que se haya descargado el archivo de instalación, ejecutarlo haciendo doble clic.
- Se abrirá el instalador. Haz clic en "Next" para comenzar.
- Seleccionar Componentes: Elige los componentes que deseas instalar. Puedes dejar las opciones predeterminadas si no estás seguro.
- Hacer clic en "Next".
- Elegir Ubicación de Instalación: Seleccionar la ubicación donde deseas instalar Talend Open Studio.
- Haz clic en "Next".
- Configurar Integración de Java: Talend Open Studio requiere Java. Si ya se tiene Java instalado, la instalación debería detectarlo automáticamente. Si no se tiene Java, se puede descargar e instalar Java antes de continuar.
- Hacer clic en "Next".
- Completar la Instalación: Revisar la configuración y hacer clic en "Install" para comenzar la instalación.
- Esperar a que finalice la instalación y hacer clic en "Finish" para cerrar el instalador.

Paso 3: Ejecutar Talend Open Studio

- Buscar el acceso directo de Talend Open Studio en el menú de inicio o escritorio.
- Hacer clic para iniciar la aplicación.
- Configuración Inicial: La primera vez que se inicia Talend, se pedirán configuraciones de algunas preferencias. Completar la configuración según las preferencias y hacer clic en "Finish".
- Explorar el Entorno de Talend: Se puede comenzar creando un nuevo proyecto o abriendo un proyecto existente, en la figura 45 se muestra el espacio de trabajo básico de Talend Open Studio.

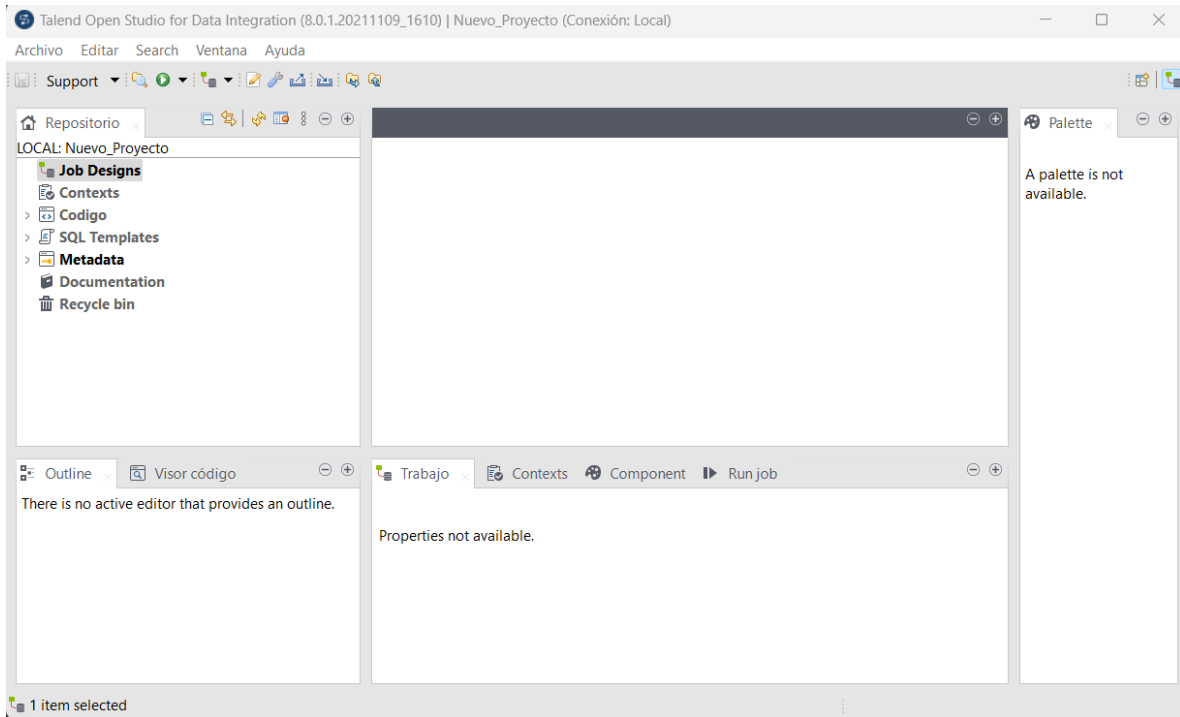


Figura 45 Espacio de trabajo de Taslend Open Studio

Creación de cuenta en AWS

- Abrir el navegador web y acceder al sitio web de AWS¹².
- Clic en "Crea una cuenta AWS" como se muestra en la figura 46.



Figura 46 Creación de cuenta en AWS

- Registrar un correo electrónico para que este pueda ser verificado y un nombre para la cuenta AWS como se muestra en la figura 47.

¹² <https://aws.amazon.com/>



Explore los productos de la capa gratuita con una cuenta de AWS nueva.

Para obtener más información, visite aws.amazon.com/free.



Registrarse en AWS

Dirección de correo electrónico del usuario raíz
Se utiliza para la recuperación de cuentas y algunas funciones administrativas

Nombre de la cuenta de AWS

Elija un nombre para la cuenta. Podrá cambiarlo en la configuración de la cuenta después de registrarse.

Verificar la dirección de correo electrónico

O

Iniciar sesión en una cuenta de AWS existente

Figura 47 Inicio de registro para creación de cuenta AWS

- Luego de que el correo ha sido confirmado deberá ingresar una contraseña segura como se indica en la figura 48 y luego dar “clic” en el botón “Continuar”.

Explore los productos de la capa gratuita con una cuenta de AWS nueva.

Para obtener más información, visite aws.amazon.com/free.



Registrarse en AWS

Cree la contraseña

✓ Hemos verificado su identidad. ✕
La dirección de correo electrónico se ha verificado correctamente.

La contraseña proporciona acceso de inicio de sesión a AWS, por lo que es importante que este proceso se realice de forma correcta.

Contraseña de usuario raíz

Confirmar la contraseña del usuario raíz

Continuar (paso 1 de 5)

Figura 48 Ingreso de contraseña para creación de cuenta AWS

- Añadir la Información de Contacto: Proporcionar información de contacto, incluyendo número de teléfono. Este paso es necesario para la verificación de la cuenta y la seguridad como se muestra en la figura 49.

Ofertas de la capa gratuita

Todas las cuentas de AWS pueden explorar 3 tipos diferentes de ofertas gratuitas, en función del producto utilizado.



Siempre gratis
Nunca vence



12 meses gratis
Inicio a partir de la fecha de registro inicial



Pruebas
Inicio a partir de la fecha de activación del servicio

Registrarse en AWS

Información de contacto

¿Cómo tiene previsto utilizar AWS?

- Empresarial: para su trabajo, escuela u organización
- Personal: para sus propios proyectos

¿A quién debemos contactar para consultar sobre esta cuenta?

Nombre completo

Número de teléfono

País o región

Dirección

Apartamento, suite, unidad, edificio, planta, etc.

Ciudad

Estado, provincia o región

Código postal

He leído y acepto los términos del [Contrato de usuario de AWS](#).

Continuar (paso 2 de 5)

Figura 49 Formulario de ingreso de datos de contacto para creación de cuenta AWS

- Configurar la Opción de Soporte Básico: Seleccionar la opción de soporte (Básico o Developer) que se desee. Puedes optar por el soporte básico de forma gratuita.
- Seleccionar un Plan de Soporte (Opcional): Si se desea, se puede seleccionar un plan de soporte de pago. Esto es opcional y se puede optar por no elegir un plan en este momento.
- Aceptar los Términos y Condiciones: Leer los términos y condiciones y marcar la casilla para indicar que se ha leído y aceptar las condiciones.
- Añadir Información de Pago: AWS solicitará información de pago para verificar la identidad. Esto puede incluir la información de tarjeta de crédito. No se realizarán cargos a menos que se elija usar servicios de pago.
- Verificación Telefónica: AWS enviará un código de verificación por mensaje de texto o llamada telefónica. Ingresar el código para verificar el número de teléfono.

- Una vez que se haya completado los pasos anteriores, la cuenta de AWS estará creada. Se puede iniciar sesión en la Consola de AWS utilizando la dirección de correo electrónico y la contraseña que se eligió durante el proceso de registro y se podrá ver una pantalla similar a la de la figura 50.

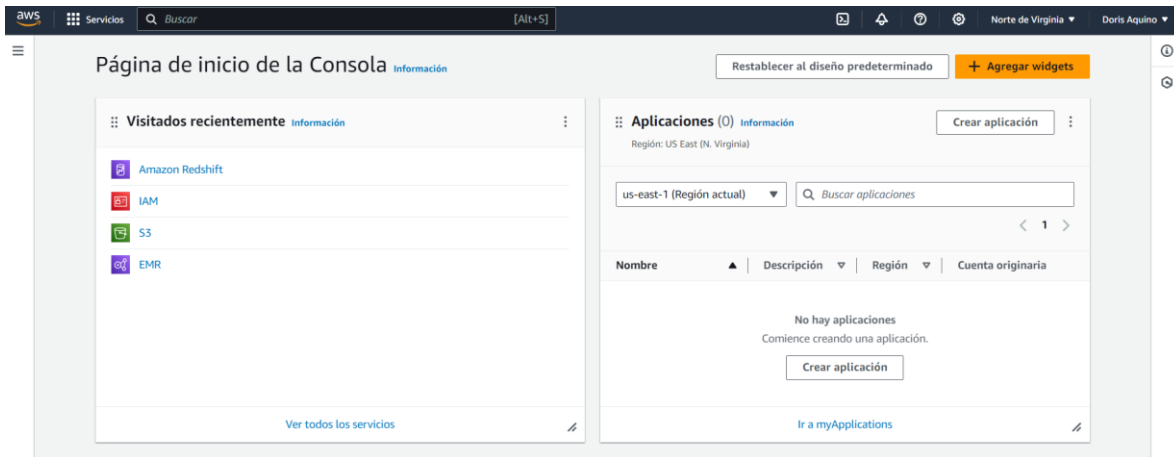


Figura 50 Tablero de AWS básico

Configuración de servicio S3 de Amazon Web Services

Para comenzar a utilizar como tal el S3 se necesita crear un bucket para almacenar los archivos que estarán en los diferentes procesos.

Para ello se debe dirigir al panel de S3 y posteriormente se hace clic al botón de “Crear bucket” como se observa en la figura 51.

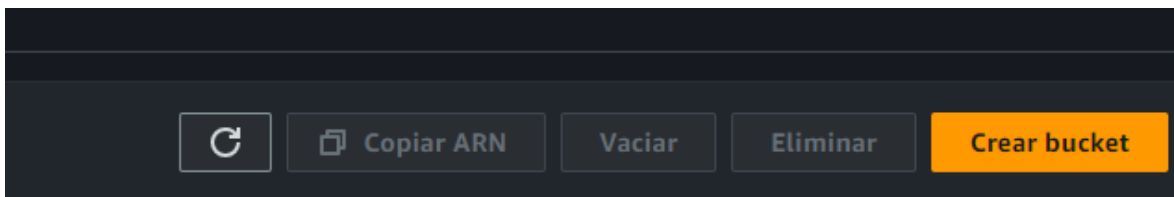


Figura 51 Botón para creación de bucket en S3

Posteriormente se debe ingresar el nombre del bucket deseado, elegir la región, se elige las ACL Deshabilitadas, y con el bloqueo al público activado, el control de versiones desactivado, y las demás opciones se deben dejar por defecto, como se puede ver en la figura 52 y 53.

Configuración general

Región de AWS

EE. UU. Este (Ohio) us-east-2

Nombre del bucket [Información](#)

dw-inversiones-digitales

El nombre del bucket debe ser único dentro del espacio de nombres global y seguir las reglas de nomenclatura del bucket. [Consulte las reglas para la asignación de nombres de buckets](#)

Copiar la configuración del bucket existente: *opcional*
Solo se copia la configuración del bucket en los siguientes ajustes.

Elegir el bucket

Formato: s3://bucket/prefijo

Propiedad de objetos [Información](#)

Controle la propiedad de los objetos escritos en este bucket desde otras cuentas de AWS y el uso de listas de control de acceso (ACL). La propiedad de los objetos determina quién puede especificar el acceso a los objetos.

ACL deshabilitadas (recomendado)

Todos los objetos de este bucket son propiedad de esta cuenta. El acceso a este bucket y sus objetos se especifica solo mediante políticas.

ACL habilitadas

Los objetos de este bucket pueden ser propiedad de otras cuentas de AWS. El acceso a este bucket y sus objetos se puede especificar mediante ACL.

Propiedad del objeto

Aplicada al propietario del bucket

Configuración de bloqueo de acceso público para este bucket

Se concede acceso público a los buckets y objetos a través de listas de control de acceso (ACL), políticas de bucket, políticas de puntos de acceso o todas las anteriores. A fin de garantizar que se bloquee el acceso público a todos sus buckets y objetos, active Bloquear todo el acceso público. Esta configuración se aplica exclusivamente a este bucket y a sus puntos de acceso. AWS recomienda activar Bloquear todo el acceso público, pero, antes de aplicar cualquiera de estos ajustes, asegúrese de que las aplicaciones funcionarán correctamente sin acceso público. Si necesita cierto nivel de acceso público a los buckets u objetos, puede personalizar la configuración individual a continuación para adaptarla a sus casos de uso de almacenamiento específicos. [Más información](#)

Bloquear todo el acceso público

Activar esta configuración equivale a activar las cuatro opciones que aparecen a continuación. Cada uno de los siguientes ajustes son independientes entre sí.

- Bloquear el acceso público a buckets y objetos concedido a través de *nuevas* listas de control de acceso (ACL)

Figura 52 Creación de bucket en s3, sección 1

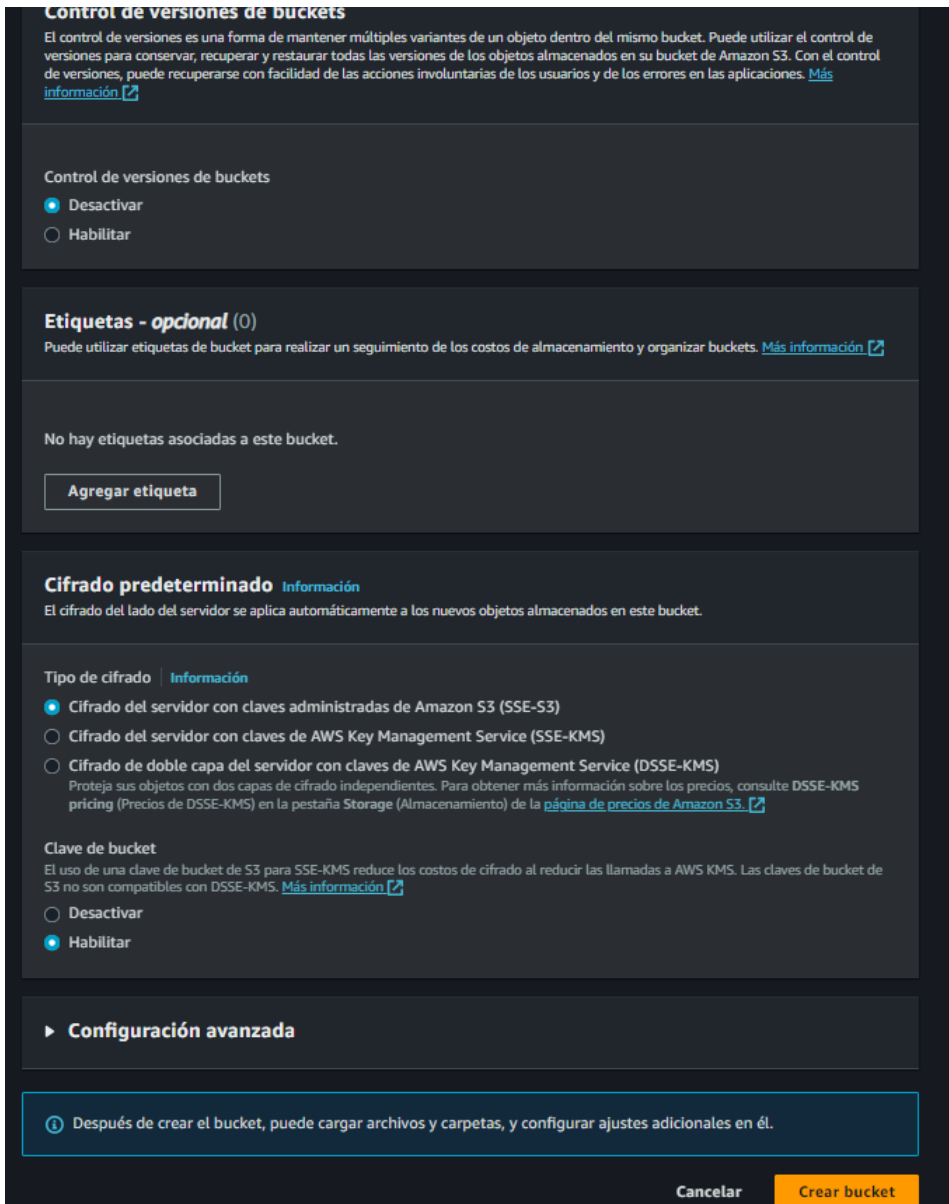


Figura 53 Creación de bucket en s3, seccion 2

Una vez creado el bucket aparecerá en el listado de buckets de S3, tal como se ve en la figura 54.

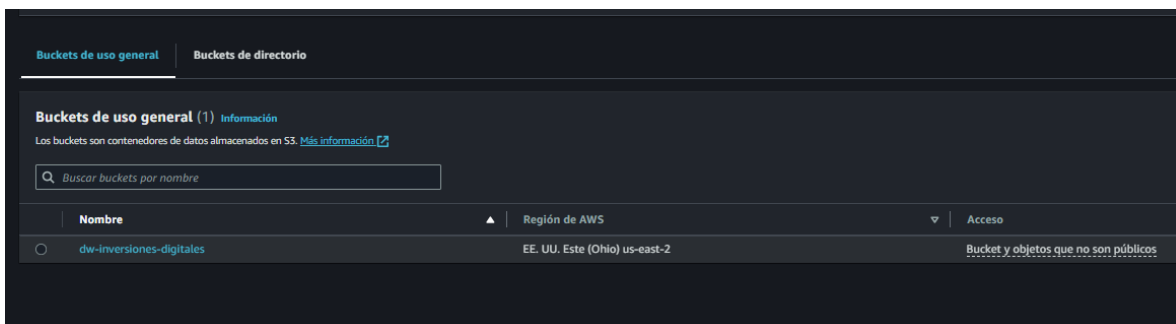


Figura 54 lista de buckets en S3

Luego de ello se debe ingresar el bucket creado, y seleccionar el botón de crear carpeta. Donde solo se ingresa el nombre de las carpetas y el tipo de clave de cifrado, en este caso: Claves administradas por Amazon S3 (SSE-S3). Se crean las carpetas hasta llegar a la estructura de carpetas que se muestra en la Figura 32.

Configuración y creación de Data Warehouse en el servicio de Redshift de Amazon Web Services

Amazon Redshift se utilizará como herramienta para el acceso a los datos, se encargará de almacenar la estructura del modelo dimensional y cargar la información almacenada en S3, para dejarla a disposición de su consumo. Las configuraciones que se realizaron fueron las siguientes:

- Iniciar sesión en la consola de AWS, buscar el componente de Amazon RedShift como se visualiza en la figura 55.

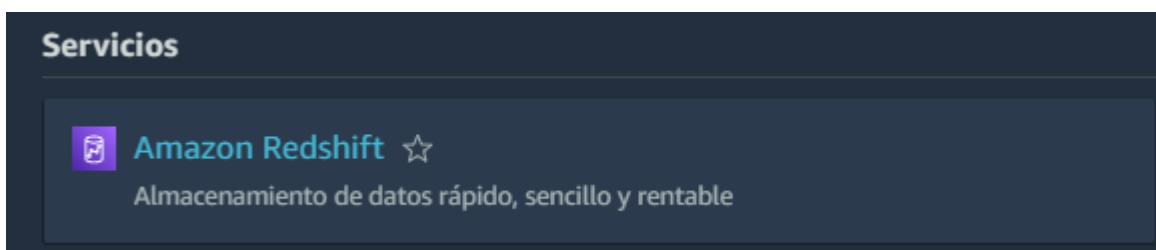


Figura 55 Componente de Amazon RedShift en AWS

- Al ingresar se crea un clúster con el botón de “Crear clúster”
- Se le asigna un nombre al clúster y se selecciona para que se quiere utilizar y de acuerdo con ello se elige la configuración que tendrá.
- Se selecciona el tipo y la cantidad de nodos que tendrá el clúster. Esto dependerá de la capacidad que necesite la empresa, en nuestro caso es un 1 nodo, puede visualizarse en la figura 56.
- Luego se asigna el nombre del usuario de administrador y su contraseña.

Configuración del clúster

Identificador del clúster
Esta es la clave exclusiva que identifica un clúster.

redshift-cluster-1

El identificador debe tener entre 1 y 63 caracteres. Los caracteres válidos son a-z (solo minúsculas) y - (guion).

Elegir el tamaño del clúster

Yo elegiré

Ayúdeme a elegir

Tipo de nodo [Información](#)
Seleccione un tipo de nodo que cumpla los requisitos de CPU, RAM, capacidad de almacenamiento y tipo de unidad.

dc2.large

Número de nodos
Escriba el número de nodos que necesita.

1

Intervalo (1-32)

Resumen de configuración [Información](#)
dc2.large | 1 nodo

\$182,50/mes

Precio de cómputo bajo demanda estimado

Ahorre más del 60 % de sus costos mediante la compra de nodos reservados. [Más información](#)

160 GB

Almacenamiento comprimido total

La capacidad de almacenamiento total del clúster si implementa el número de nodos elegido.

Figura 56 Creación de Cluster de RedShift

- Después de haber creado el clúster se debe configurar la accesibilidad pública, para ello se debe ingresar al clúster y en la desglosar el botón “Acciones” y seleccionar la opción de “Modificar la configuración de accesibilidad pública”. Esta acción se puede observar en la figura 57.

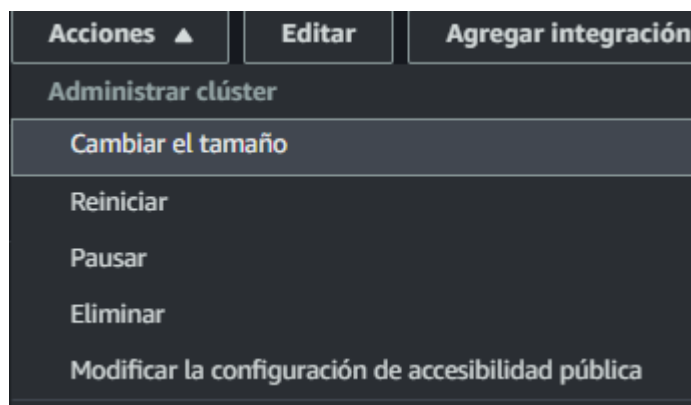


Figura 57 Redshift - Modificar la configuración de acceso público

- Marcar con un cheque que se permita la accesibilidad pública. Esta acción se puede observar en la figura 58.

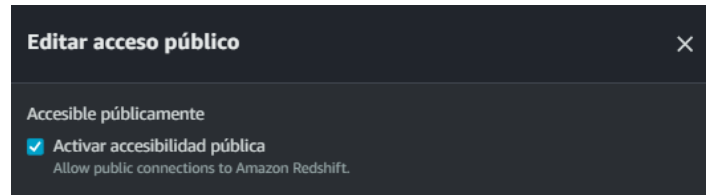


Figura 58 Redshift - Habilitar acceso público

▪ Usuarios IAM

Amazon IAM, tiene la funcionalidad de permitir y administrar usuarios y permisos de usuario o grupos de usuarios en AWS. El usuario de IAM se crea para alguien que necesite acceder a la consola de AWS, o cuando se posea una aplicación que necesite hacer llamadas API a los servicios de Amazon Web Services. Esto es para agregar una capa adicional de seguridad a la cuenta de AWS.

Para ello procederemos con los siguientes pasos:

- Iniciar sesión en la consola de AWS y buscar en el selector de búsqueda "IAM", donde saldrá la opción que se puede visualizar en la figura 59.

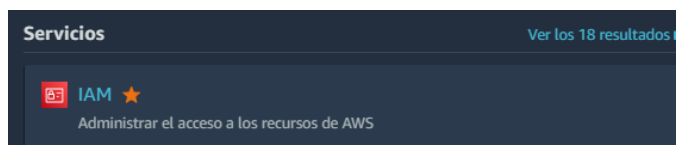


Figura 59 Seleccionar servicio Amazon IAM

- Seleccionar la opción "política" y dar clic en el botón "Crear política".
- En servicios se busca S3 y se selecciona el nivel de acceso que se desea asignar.
- En la sección de recursos en el apartado de Bucket agregamos el ARN del bucket al que le queremos dar acceso.
- Clic en revisar política y se le asigna un nombre a la política recién creada.
- Seleccionar "usuario"s y clic en el botón de "añadir usuario".
- Configurar el detalle del nuevo usuario, el nombre y la contraseña.
- Agregar los permisos al nuevo usuario que estamos por crear, se selecciona la política anteriormente creada y se continua con el siguiente paso.
- Luego dar clic en "Crear usuario".
- En este punto AWS genera una URL la cual se debe guardar En este punto AWS genera una URL la cual será usada por el usuario IAM recién creado para acceder a la consola de AWS.

Instalación de herramienta para visualización de datos

Utilizar PowerBI como la herramienta de visualización y generación de informes. Configurar conexiones directas desde PowerBI a Redshift para acceder y analizar datos en tiempo real y así poder crear paneles interactivos y visualizaciones personalizadas que reflejen los indicadores clave definidos en los objetivos, para esto los pasos para instalar PowerBI son los siguientes:

1. Descarga de Power BI Desktop:

- Visita el sitio web de Power BI ¹³.
- Haz clic en "Productos" en el menú superior y selecciona "Power BI Desktop" como se muestra en la figura 60.



Figura 60 Menú para acceder a la descarga de PowerBI

- En la página de Power BI Desktop, haz clic en "Descargar gratis" como se indica en la figura 61.



Figura 61 Botón para descarga de PowerBI

- Inicia la Descarga: Después de seleccionar la versión, haz clic en "Descargar" para iniciar la descarga del instalador.
- ## 2. Ejecuta el Instalador:
- Una vez que se complete la descarga, ejecuta el archivo de instalación que has descargado (por ejemplo, PBIDesktopSetup_x64.exe).
 - Configura la Instalación: El instalador te guiará a través del proceso de instalación. Haz clic en "Next" para avanzar en cada paso.

¹³ <https://powerbi.microsoft.com/>

- Acepta los Términos de Licencia: Lee los términos de licencia y, si estás de acuerdo, marca la casilla para aceptarlos. Luego, haz clic en "Next".
- Elige la Ubicación de Instalación: Selecciona la ubicación donde deseas instalar Power BI Desktop. Por lo general, puedes dejar la ubicación predeterminada y hacer clic en "Next".
- Selecciona Componentes (Opcional): Puedes elegir instalar componentes adicionales si lo deseas. En la mayoría de los casos, puedes dejar las opciones predeterminadas y hacer clic en "Install".
- Finaliza la Instalación: Espera a que la instalación se complete. Una vez finalizada, haz clic en "Finish" para cerrar el instalador.

3. Inicia Power BI Desktop: Después de la instalación, busca y ejecuta Power BI Desktop desde el menú de inicio o el escritorio de tu computadora y se verá una pantalla similar a la figura 62.

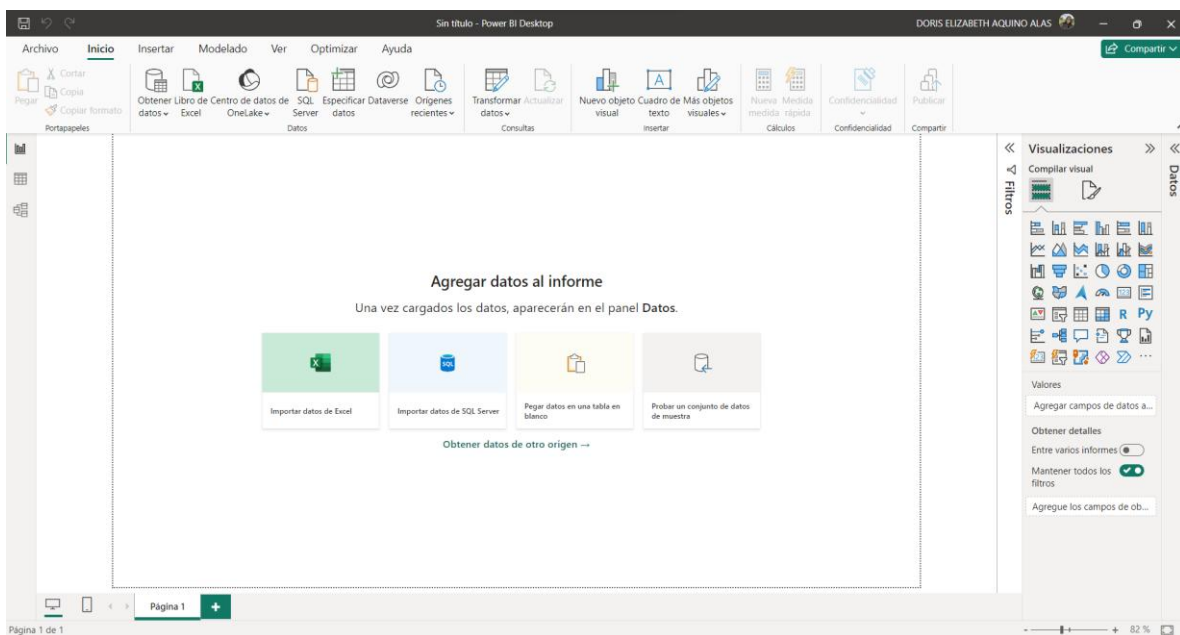


Figura 62 Pantalla de inicio de PowerBI

b) Presupuesto de implementación

El presupuesto de implementación ha sido diseñado con los elementos esenciales que son necesarios para la elaboración de la solución propuesta, con la intención que sea implementado y llevado a un ambiente de producción, estimando a través de una investigación se presenta el siguiente presupuesto:

Recurso Humano

Para el presupuesto de recurso humano se realizó una estimación de mercado tanto para el ingeniero de datos ¹⁴ como para el analista de datos ¹⁵ para establecer los salarios.

Cargo	Funciones para realizar	Salario mensual	Cantidad requerida	Costo
Ingeniero de datos	<ul style="list-style-type: none"> • Análisis de la base de datos transaccional. • Diseño del modelo dimensional. • Diseño e implementación de los ETL en el software Talend Open Studio. • Manejo de los servicios Amazon S3, Amazon RedShift y Amazon IAM. 	\$2,500	1	\$2,500
Analista de datos	<ul style="list-style-type: none"> • Comunicación con la API de Amazon para consumir datos desde Amazon RedShift a Power BI. • Creación de reportes en Power BI. • Análisis de resultados del proceso de negocios. 	\$1,700	1	\$1,700
Total, costo mensual				\$ 4,200

Tabla 32 Presupuesto de Implementación. Recurso humano

Equipo informático, Hardware

El presente presupuesto se realizó estimando los valores promedio actuales del mercado en el país. Cabe aclarar que las características son las mínimas para un funcionamiento óptimo.

Componente	Descripción	Precio	Cantidad	Total
Disco duro	SSD nVME 1 TB	\$172,32	2	\$344,64
Memoria RAM	16 GB RAM DDR4	\$95,48	2	\$190,96
Procesador	i5 12Th	\$295,00	2	\$590,00
Fuente de poder	80 Plus, 600 W, certificada	\$79,00	2	\$158,00
UPS	APC 1000VA	\$144,00	2	\$288,00
Otros componentes	Teclado, mouse, case, motherboard...	\$350,00	2	\$700,00
Total				\$2.271,60

Tabla 33 Presupuesto de implementación. Equipo informático, Hardware

¹⁴ Información para Ingeniero de datos: <https://worldsalaries.com/average-data-warehousing-specialist-salary-in-el-salvador/>

¹⁵

Información salarial para analista de datos: Información para analista de datos: <https://worldsalaries.com/average-data-analyst-salary-in-el-salvador/>

Licencias

En la tabla 34 se presenta el presupuesto de licencias para Power BI.

Licencia	Cantidad	Costo Unitario	Subtotal
Licencia power BI Pro	15	\$10	\$150
Total, mensual			\$150

Tabla 34 Presupuesto implementación, Licencias

Servicios AWS

Según la calculadora de servicios de Amazon¹⁶ se presenta en la tabla 35 el presupuesto estimado por los servicios de Amazon.

Servicio	Descripción	Costo	Subtotal
S3 Standard	<ul style="list-style-type: none">Almacenamiento y datos devueltos/escaneados para 10 TB.10,000 solicitudes PUT,COPY,POST y LIST desde y hacia S3.10,000 peticiones GET, SELECT y otras peticiones desde S3 Standard.	\$235.57	\$235.57
Amazon RedShift	<ul style="list-style-type: none">2 nodos dc2.large, con disponibilidad del 100% del tiempo.	\$325	\$325
Total mensual		\$560.57	

Tabla 35 Presupuesto implementación, servicios AWS

Los servicios de AWS pueden variar según vayan cambiando las necesidades de la empresa, y las políticas por parte de Amazon, por lo cual es un estimado para la actualidad y es sujeta a cambios.

Elemento	Costo
Recurso humano	\$4,200
Equipo informático, hardware	\$2,271.60
Licencias	\$150
Servicios Amazon AWS	\$560.57
Total	\$7182.11

Tabla 36 Presupuesto de implementación

Nota: los costos estimados son mensuales a excepción del equipo informático que es un costo único.

¹⁶ AWS Pricing Calculator <https://calculator.aws/>

c) Análisis de resultados

Reporte 1: Reporte financiero de beneficio neto

Descripción: La métrica de beneficio neto es una medida financiera fundamental que le indica a la empresa la cantidad de ingresos que ha generado después de deducir todos los costos y gastos asociados con sus operaciones, impuestos y otros elementos financieros. Es así que en informe se listan las sucursales, con sus ingresos por ventas, impuestos y posteriormente el total de beneficio neto para un rango de fecha específica. Se presenta además de la gráfica de beneficios por sucursal, la sucursal con mayores ventas, y el filtro para el tiempo, siendo estos valores positivos, representando que la empresa ha generado ganancias.

Inversiones Digitales - Reporte Financiero Beneficio Neto

El beneficio neto representa los ingresos por venta después del pago de impuesto

Sucursal	Ingreso por venta	Impuestos	Beneficio neto
Inversiones Digitales (Santa Tecla)	\$4.685,00	\$468,50	\$4.216,50
Inversiones Digitales (San Vicente)	\$4.630,00	\$463,00	\$4.167,00
Inversiones Digitales(Casa Matriz - San Salvador)	\$4.500,00	\$450,00	\$4.050,00
Inversiones Digitales (Santa Ana)	\$4.486,00	\$448,60	\$4.037,40
Inversiones Digitales (San Miguel)	\$4.175,00	\$417,50	\$3.757,50
Total	\$22.476,00	\$2.247,60	\$20.228,40

Sucursal con mayores ventas después de impuesto

Inversiones Digitales (Santa Tecla)
\$4.216,50

Beneficio neto por Sucursal

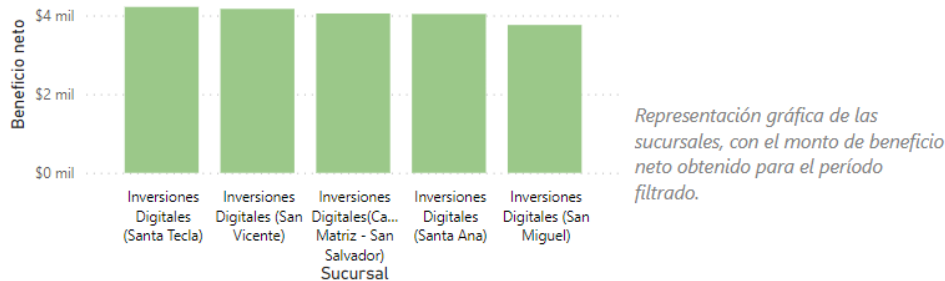


Figura 63 Reporte financiero sobre el beneficio neto

Reporte 2: Reporte financiero de índice de retorno de inversión (ROI)

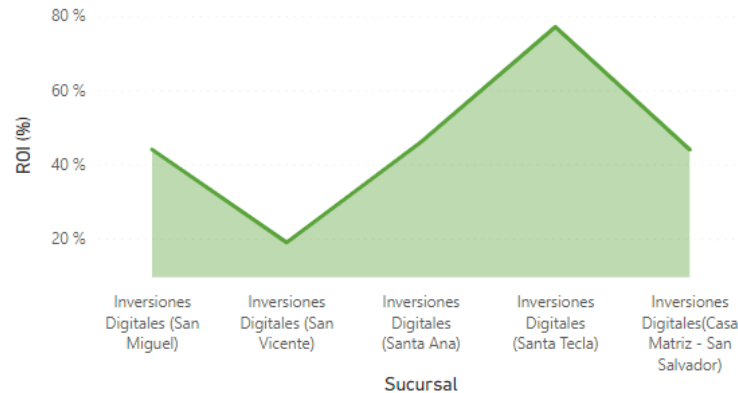
Descripción: El retorno de inversión le permite a la empresa medir la capacidad que tiene para generar ingresos por cada unidad de capital accionario. En el informe se listan las sucursales, con sus respectivos índices de retorno de inversión para un rango de fecha específica. Se presenta además de la gráfica de este por sucursal, la sucursal con mayor índice, y el filtro para el tiempo en la parte inferior.

Inversiones Digitales - Reporte Financiero Índice de Retorno de Inversión (ROI)

Sucursal	ROI %
Inversiones Digitales (San Miguel)	44,00 %
Inversiones Digitales (San Vicente)	19,00 %
Inversiones Digitales (Santa Ana)	46,00 %
Inversiones Digitales (Santa Tecla)	77,00 %
Inversiones Digitales(Casa Matriz - San Salvador)	44,00 %

El ROI es el rendimiento obtenido sobre las inversiones realizadas, en porcentaje.

ROI (%) por Sucursal



Representación los porcentajes de ROI para cada sucursal.

Sucursal con mayor índice de rendimiento sobre inversión.

Inversiones Digitales (Santa Tecla)
77,00 %
Suma de ROI %

- Año, Semestre, Trimestre, Mes... ▾
- ▾ 2020
 - ▾ 2021
 - ▾ 2022
 - ▾ 2023

Figura 64 Reporte financiero sobre el índice de retorno de inversión (ROI)

Reporte 3: Reporte financiero sobre índice de liquidez

Descripción: En el informe se listan las sucursales, con sus respectivos valores de activos y pasivos corrientes, su liquidez y respectiva ratio para un rango de fecha específica. Se presenta la gráfica de los activos corrientes vs pasivos corrientes por sucursal, un gráfico de líneas del Ratio, y el filtro para el tiempo en la parte inferior y dado que el índice de liquidez está evaluando la capacidad de la empresa para cumplir con sus obligaciones financieras a corto plazo utilizando sus activos corrientes, y como se puede observar los valores son positivos nos indica que la solvencia y la salud financiera de la empresa a corto plazo es buena.

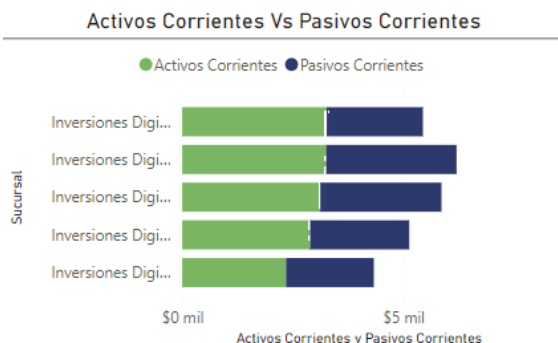
Inversiones Digitales - Reporte Financiero Índice de Liquidez

Sucursal	Activos Corrientes	Pasivos Corrientes	Liquidez	Ratio de Liquidez
Inversiones Digitales (San Miguel)	\$2.350,00	\$1.985,00	\$365,00	1,18
Inversiones Digitales (San Vicente)	\$3.270,00	\$2.175,00	\$1.095,00	1,50
Inversiones Digitales (Santa Ana)	\$2.895,00	\$2.241,00	\$654,00	1,29
Inversiones Digitales (Santa Tecla)	\$3.125,00	\$2.740,00	\$385,00	1,14
Inversiones Digitales(Casa Matriz - San Salvador)	\$3.257,00	\$2.950,00	\$307,00	1,10

La liquidez hace referencia a la capacidad de convertirse en efectivo y no perder su valor.

En términos monetarios se obtiene restando los pasivos corrientes de los activos corrientes.

Su ratio es la división de los activos corrientes con los pasivos corrientes.



Gráficos de Activos Corrientes vs Pasivos Corrientes, por sucursal



Gráficos ratio de liquidez, por sucursal.

Año, Semestre, Trimestre, Mes... ▾

2020
 2021
 2022
 2023

Figura 65 Reporte financiero sobre el índice de liquidez

Reporte 4: Reporte financiero sobre tasa de crecimiento en utilidades por período

Descripción: En el reporte se pueden observar el listado de sucursales donde se detallan las utilidades para el período uno contra las utilidades del período dos, además una tabla donde se muestra las utilidades de cada una de las sucursales, la métrica de tasa de crecimiento en utilidades le permite a la empresa saber si ha experimentado un aumento en sus beneficios entre los dos periodos comparados, para este caso podemos ver que 2 de las sucursales tienen valores negativos lo que indica una baja en las ganancias, una sucursales con valores positivos lo que indica un aumento en las ganancias y las otras dos sucursales con valores en cero que indica que las ganancias se han mantenido iguales.

Inversiones Digitales - Reporte Financiero Tasa de crecimiento en utilidades por período

Sucursal	Año	Mes	Día	Utilidades período 1	Año	Mes	Día	Utilidades período 2
Inversiones Digitales(Casa Matriz - San Salvador)	2023	diciembre	4	\$675,00	2023	diciembre	5	\$798,00
Inversiones Digitales (Santa Tecla)	2023	diciembre	4	\$789,00	2023	diciembre	5	\$920,00
Inversiones Digitales (Santa Ana)	2023	diciembre	4	\$1.002,00	2023	diciembre	5	\$740,00
Inversiones Digitales (San Vicente)	2023	diciembre	4	\$740,00	2023	diciembre	5	\$741,00
Inversiones Digitales (San Miguel)	2023	diciembre	4	\$851,00	2023	diciembre	5	\$850,00
Total				\$4.057,00				\$4.049,00

La tasa de crecimiento indica el comportamiento de las utilidades de una sucursal en un período con respecto al anterior.

Una tasa negativa, indica bajas en las ganancias.

Una tasa en 0, indica que las utilidades se mantienen

Una tasa positiva, indica un incremento de utilidades

Sucursal	Tasa de crecimiento por período
Inversiones Digitales(Casa Matriz - San Salvador)	-18,00
Inversiones Digitales (Santa Tecla)	-17,00
Inversiones Digitales (Santa Ana)	26,00
Inversiones Digitales (San Vicente)	0,00
Inversiones Digitales (San Miguel)	0,00

Tasa de crecimiento entre el período filtrado y su antecesor.

Tasa de crecimiento por período por Sucursal



Figura 66 Reporte financiero sobre tasa de crecimiento en utilidades por período

Reporte 5: Reporte financiero sobre índice de eficiencia de cobros

Descripción: El reporte presenta el índice de eficiencia de cobros, la cual es una métrica que evalúa la eficacia y rapidez con la que una empresa puede recuperar los fondos adeudados por sus clientes, se puede observar una tabla por sucursales donde se tiene el índice de eficiencia de cobros para cada una, siendo la sucursal de Santa Tecla la que mayor índice obtuvo para el período de tiempo mostrado.

Inversiones Digitales - Reporte Financiero Índice de Eficiencia de cobros

El índice de efectividad de cobros, refleja el manejo de las operaciones de cobro.

Sucursal	Índice de eficiencia de cuentas por cobrar
Inversiones Digitales (San Miguel)	1,24
Inversiones Digitales (San Vicente)	1,47
Inversiones Digitales (Santa Ana)	1,03
Inversiones Digitales (Santa Tecla)	1,51
Inversiones Digitales(Casa Matriz - San Salvador)	1,15

Sucursal con mayor eficiencia de cobros.

Inversiones Digitales (Santa Tecla)
1,51

Eficiencia de cuentas por cobrar por Sucursal

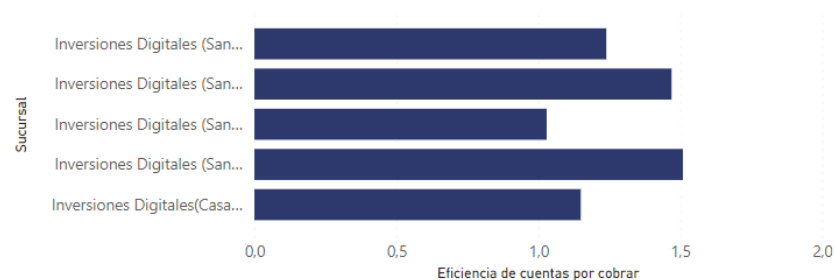


Gráfico de índice de efectividad de cobro por sucursal.

Figura 67 Reporte financiero sobre índice de eficiencia de cobros

Año, Semestre, Trimestre, Mes...
 2020
 2021
 2022
 2023

7. Conclusiones

Se logró realizar un análisis exhaustivo de la base de datos transaccional de Inversiones Digitales gracias a lo cual se identificaron las tablas, relaciones y entidades relevantes vinculadas a los procesos clave que impactan la información financiera de la empresa, logrando así generar un data set acorde a la solución que se requería presentar asegurando una estructura óptima que sirve como insumo fundamental para la construcción del Data Warehouse, además de un diccionario de datos y un data profiling que proporcionaron información valiosa sobre cada uno de los datos a utilizar en la solución financiera.

La construcción del Data Warehouse se llevó a cabo de manera efectiva, brindando una solución integral a la problemática presentada por Inversiones Digitales contribuyendo así a una gestión más eficiente de los datos financieros.

Se estableció una estructura de carpetas idónea en Talend Open Studio para generar los ETL necesarios. Esto incluye procesos para carga, transformación, construcción del Data Warehouse y descarga de datos, facilitando la gestión y mantenimiento de los flujos de datos, gracias a esto se logró diseñar una estructura de carpetas eficiente en Amazon S3 para almacenar los datos obtenidos de los diferentes procesos ETL generados. Esto asegura una organización ordenada y accesible de los conjuntos de datos.

El modelo dimensional se implementó con éxito en el servicio Redshift de AWS, aprovechando su capacidad de procesamiento masivo y eficiencia en consultas para respaldar el análisis de datos financiero y debido a que ya se contaba con todos los datos cargados en el servicio de S3 se pudo cargar todos esos datos en las dimensiones y tabla de hechos, cuya estructura ya estaba creada en Redshift, además se logró realizar una configuración exitosa para conectar PowerBI con Reshift para cargar los datos y así poder realizar las operaciones necesarias para obtener las métricas que fueron establecidas para dar solución a la problemática de la empresa, es así que se crearon 5 tableros con el análisis de las métricas.

8. Recomendaciones

- Implementar un plan sistemático de mantenimiento periódico del modelo dimensional es fundamental para asegurar su continua relevancia y precisión en consonancia con las cambiantes necesidades del negocio. La actualización constante se vuelve esencial para reflejar de manera efectiva los cambios tanto internos en la empresa como aquellos derivados de ajustes en el entorno operativo. Este enfoque proactivo garantiza que el modelo dimensional siga siendo una herramienta estratégica y alineada con los objetivos y evolución dinámica del negocio.
- En la implementación actual, se empleó Talend Open Studio para llevar a cabo los procesos ETL, aprovechando sus componentes integrados. Sin embargo, se sugiere considerar la posibilidad de explorar otras herramientas de interfaz gráfica similares o evaluar la utilización de lenguajes de programación como Java, Python o Scala, que ofrecen robustos paquetes y librerías especializadas para la gestión de procesos ETL. La elección de la herramienta más adecuada debe basarse en la naturaleza específica de las tareas a realizar y en el volumen de datos que se tiene previsto procesar. Esto permitirá optimizar la eficiencia y flexibilidad de los procesos ETL, adaptándolos de manera más precisa a los requisitos y metas del proyecto.
- Una consideración adicional que podría explorarse es la elaboración de una cotización detallada para los servicios de almacenamiento, ya sea en la nube o en un entorno local, basándose en el volumen de datos que se gestionará. Sería beneficioso evaluar la frecuencia con la que estos datos se actualizarán y se accederá a ellos. Esto permitirá obtener una estimación precisa de los costos asociados, facilitando así una planificación financiera más precisa y asegurando que los recursos de almacenamiento se alineen adecuadamente con las necesidades operativas del proyecto.

9. Bibliografía

- Chavez, J. (2022, 8 julio). Ceupe. Ceupe. <https://www.ceupe.com/blog/que-es-una-empresa.html>
- ¿Qué es el big data? (s. f.). <https://www.oracle.com/es/big-data/what-is-big-data/>
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: the next frontier for innovation, competition, and productivity. En *McKinsey & Company*. <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation>
- ¿Qué es OLAP? | IBM. (s. f.). <https://www.ibm.com/mx-es/topics/olap>
- Bases de datos en AWS: la herramienta correcta para el trabajo adecuado. (s. f.). [Vídeo]. Amazon Web Services, Inc. <https://aws.amazon.com/es/nosql/>
- ¿Qué es una base de datos relacional? | IBM. (s. f.). <https://www.ibm.com/mx-es/topics/relational-databases>
- ¿Qué es la automatización? | IBM. (s. f.). <https://www.ibm.com/es-es/topics/automation>
- Quiles, J. M. (2023, 19 junio). ¿Qué es un ERP? ¿Para qué sirve y qué beneficios aporta? Clavei | Software solutions for business. <https://www.clavei.es/blog/erp-que-es/>
- Daniela. (2022, 1 noviembre). ¿Que son los inventarios? características y consejos clave - Corposuite | Consultoría del ERP en la nube. Corposuite | Consultoría del ERP en la nube Oracle NetSuite México. <https://corposuite.com.mx/2022/10/25/que-son-los-inventarios/>
- IBM documentation. (s. f.). <https://www.ibm.com/docs/es/i/7.3?topic=platform-java-development-kit>
- A, D., & A, D. (2023, 10 enero). ¿Qué es JSON? Tutoriales Hostinger. <https://www.hostinger.es/tutoriales/que-es-json>
- Facturama, T. (2021, 29 abril). ¿Qué es una nota de crédito? concepto y ejemplo. Facturama Blog - Actualidad en información Contable y Fiscal. <https://facturama.mx/blog/que-significa/nota-de-credito/>
- Facturama, T. (2022, 18 abril). ¿Qué es la nota de débito de los documentos relacionados? Facturama Blog - Actualidad en información Contable y Fiscal. <https://facturama.mx/blog/que-significa/nota-de-debito-de-los-documentos-relacionados/>
- Morales, F. C. (2022, 24 noviembre). Punto de venta. Economipedia. <https://economipedia.com/definiciones/punto-de-venta.html>
- Stock: qué es Y tipología. (s. f.). <https://blog.toyota-forklifts.es/stock-que-es>
- Equipo editorial de IONOS. (2020, 17 septiembre). XML-RPC: Llamadas a procedimientos remotos en formato XML. IONOS Digital Guide. <https://www.ionos.es/digitalguide/paginas-web/desarrollo-web/que-es-xml-rpc/>
- ¿Qué es Hadoop? (s. f.). SAS. https://www.sas.com/es_mx/insights/big-data/hadoop.html
- ¿Qué es el Internet de las cosas (IoT) y cómo funciona? (s. f.). <https://www.redhat.com/es/topics/internet-of-things/what-is-iot>

10. Glosario

- **Automatización:** busca gestionar procesos a través de herramientas, en su mayoría digitales, que agilizan su ejecución, aplicando un mínimo de esfuerzo humano.
- **Base de datos no relacional:** información resumida por medio de filas y columnas, estructurada en tablas, las cuáles pueden mantener relaciones entre sí.
- **Base de datos relacional:** tipo de base de datos que contiene distintos formatos y modelos de datos que pueden no tener relaciones entre sí, se utilizan en su mayoría como contenedores de grandes volúmenes de datos, que posteriormente serán procesados.
- **Cotización:** documento oficial donde se detallan las características, condiciones y precio de un bien o producto, orientado a un proceso de venta en el futuro.
- **CRM (Customer Relationship Management):** estrategia para manejar la conexión y relaciones con los clientes y clientes potenciales.
- **ERP (Enterprise Resource Planning):** es una o varias aplicaciones de software que permiten a una empresa administrar y automatizar procesos principalmente vinculados a áreas operativas y administrativas.
- **Hadoop:** software de código abierto que permite almacenar datos y procesar conjuntos grandes de datos.
- **Internet de las cosas (IoT):** dispositivos físicos cotidianos que se pueden conectar a internet o a una red e interactuar con otros dispositivos o usuarios.
- **Inventario:** instrumento que permite llevar un registro de existencias y actividades de los bienes y recursos de una entidad.
- **JDK (Java Development Kit):** software que permite escribir y ejecutar aplicaciones desarrolladas bajo el lenguaje de programación Java.
- **JSON (Javascript Object Notation):** documento de texto cuyo formato y estructura permite almacenar información y es compatible en numerosas aplicaciones y lenguaje de programación.
- **Machine learning:** uso de datos y algoritmos para que máquinas o dispositivos simulen aprendan de la información, simulando el aprendizaje humano.
- **Nota de crédito:** documento fiscal que sirve como comprobante que una factura fue anulada de forma total o parcial, debido a un reembolso.
- **Nota de débito:** documento fiscal dirigido al cliente que indica que el monto de pago, deuda o compra tiene un incremento de cargos.
- **Onboarding:** prácticas de capacitación que realiza una empresa para integrar al nuevo personal.
- **OnPremise:** término que hace referencia a cuando la infraestructura tecnológica de una entidad está instalada en la empresa misma.
- **Punto de venta:** lugar físico o virtual donde un negocio expone sus productos a los consumidores con la finalidad de comercializarlos.

- **Stock:** cantidad de mercancías guardadas en un almacén en un determinado período de tiempo.
- **TinyERP:** software ERP en línea, que permite a micro y pequeñas empresas gestionar procesos financieros y de comercio.
- **XML-RPC (Extensible Markup Language Remote Procedure Call):** protocolo que permite el acceso remoto a ordenadores que son parte de distintas redes o se encuentran distribuidos.

11. Anexos

Anexo 1: Cronograma de actividades

Establecer un cronograma de actividades permite llevar a cabo un proyecto exitoso y gestionar los recursos de manera más efectiva, es por eso que en la figura 3 se presenta el cronograma que se siguió para el desarrollo del prototipo de diseño e implementación de un Data Warehouse financiero para la empresa Inversiones Digitales.

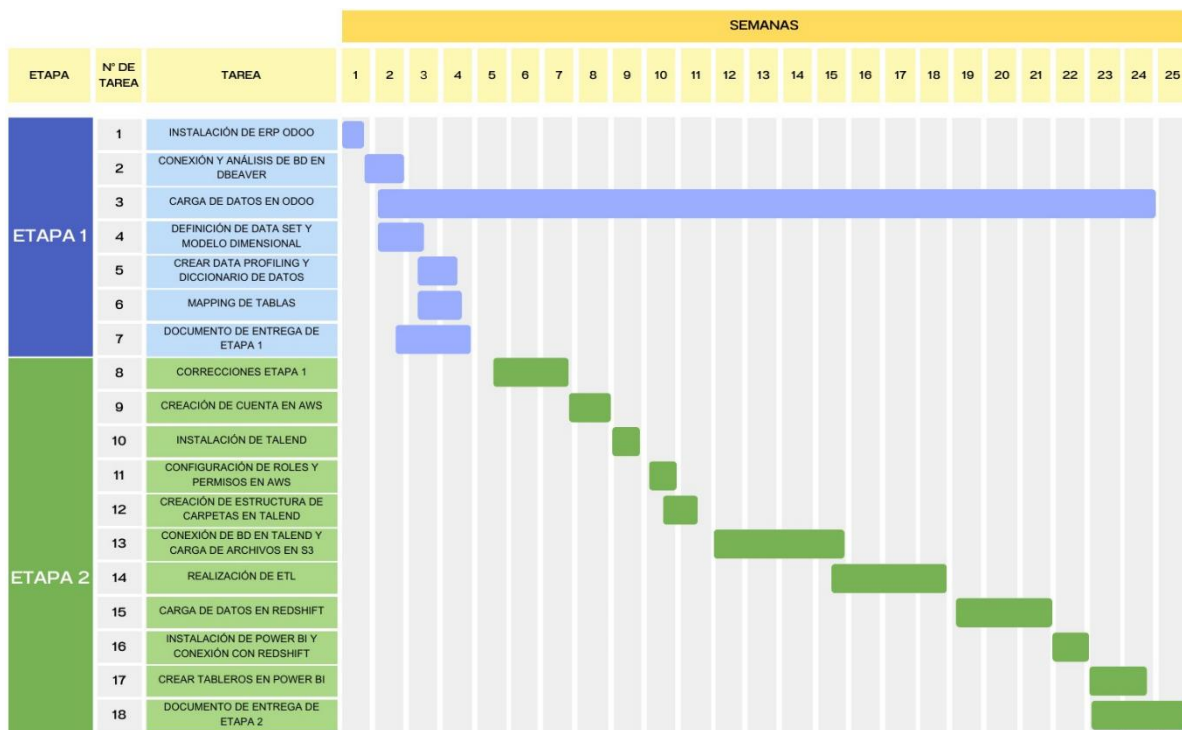


Figura 6368 Cronograma de

Anexo 2: Presupuesto

El costo de la realización de la solución planteada se divide en los siguientes elementos:

Los salarios se toman en base con la misma referencia del salario para un ingeniero de datos en El Salvador, por lo cual se tomará el promedio.

Recurso humano

Se muestra a continuación

Cargo	Funciones para realizar	Salario mensual	Cantidad requerida	Costo
Ingeniero(a)/Analista de datos	<ul style="list-style-type: none"> • Análisis de la base de datos transaccional. • Diseño del modelo dimensional. • Diseño e implementación de los ETL en el software Talend Open Studio. • Manejo de los servicios Amazon S3, Amazon RedShift y Amazon IAM. • Comunicación con la API de Amazon para consumir datos desde Amazon RedShift a Power BI. • Creación de reportes en Power BI. • Análisis de resultados del proceso de negocios. 	\$1,450	2	\$ 2,900
Total, costo mensual				\$ 2,900

Tabla 37 Presupuesto de implementación solución planteada, Recurso humano.

Equipo informático, Hardware

El presente presupuesto se realizó estimando los valores promedio actuales del mercado en el país.

Componente	Descripción	Precio	Cantidad	Total
Disco duro	SSD nVME 1 TB	\$172,32	2	\$344,64
Memoria RAM	16 GB RAM DDR4	\$95,48	2	\$190,96
Procesador	i5 12Th	\$295,00	2	\$590,00
Fuente de poder	80 Plus, 600 W, certificada	\$79,00	2	\$158,00
UPS	APC 1000VA	\$144,00	2	\$288,00
Otros componentes	Teclado, mouse, case, motherboard...	\$350,00	2	\$700,00
Total				\$2.271,60

Tabla 38 Presupuesto de implementación solución planteada, Equipo informático

Licencias

Se utilizaron licencias gratuitas para las herramientas utilizadas (Talend Open Studio y Microsoft Power BI).

Servicios AWS

Se utilizó la capa gratuita que ofrece AWS, pero cabe mencionar que es para un entorno controlado donde se limita las transacciones y tiempo de disponibilidad para RedShift

Servicio	Descripción	Costo	Subtotal
S3 Free Tier	<ul style="list-style-type: none">Almacenamiento 5 GB20,000 solicitudes Get2,000 solicitudes PutTiempo 12 meses	\$0.00	\$0.00
Amazon RedShift	<ul style="list-style-type: none">1 nodos dc2.large, con disponibilidad del 100% del tiempo, condicionada.Limite de crédito \$300, 3 meses.	\$0.01	\$0.01
Total mensual		\$0.01	

Tabla 39 Presupuesto implementación solución planteada, servicios AWS

Los servicios de AWS pueden variar según vayan cambiando las necesidades de la empresa, y las políticas por parte de Amazon, por lo cual es un estimado para la actualidad y es sujeta a cambios.

Elemento	Costo
Recurso humano	\$2,900
Equipo informático, hardware	\$2,271.60
Licencias	\$0.00
Servicios Amazon AWS	\$0.01
Total	\$5,171.61

Tabla 40 Presupuesto de implementación solución planteada.

Nota: los costos estimados son mensuales a excepción del equipo informático que es un costo único.

Anexo 3: Script de generación de tablas en RedShift

--dimension de dimsucursal

```
CREATE TABLE dimsucursal(  
    sucursalKey int PRIMARY KEY,  
    sucursalId int distkey,  
    nombre varchar(150)  
)  
sortKey(sucursalKey);
```

--dimension de cuenta

```
CREATE TABLE dimcuenta(  
    cuentaKey int PRIMARY KEY,  
    cuentaId int,  
    codigoCuenta varchar(15),  
    nombreCuenta varchar(150),  
    tipoCuenta varchar(150),  
    grupoCuenta varchar(150),  
    descripcionCuenta varchar(150),  
    tieneCuentaPadre varchar(3),  
    codigoCuentaPadre varchar(15),  
    nombreCuentaPadre varchar(150)  
)  
sortKey(cuentaKey);
```

--dimension de fecha

```
CREATE TABLE dimfecha(  
    fechaKey int PRIMARY KEY distkey, --formato iso  
    fecha varchar(10),  
    dia varchar(25),  
    diaNumeroSemana int,  
    diaNumeroMes int,  
    diaNumeroAnio int,
```

```

mes varchar(25),
mesNumero int,
numeroSemanaMes int,
numeroSemanaAnio int,
trimestre varchar(25),
trimestreNumero int,
semestre varchar(25),
semestreNumero int,
anio int,
diaLaboral varchar(10)
)sortKey(fechaKey, mes, anio);
--fact table de transacciones financieras
CREATE TABLE facttransaccionfinanciera(
    fechaKey int,
    cuentaKey int,
    sucursalKey int,
    montoCargo numeric(8,2),
    montoAbono numeric(8,2),
    balance numeric(8,2),
    saldoInicial numeric(8,2),
    saldoFinal numeric(8,2),
    FOREIGN KEY(cuentaKey) REFERENCES dimcuenta(cuentaKey),
        FOREIGN KEY(sucursalKey) REFERENCES dimsucursal(sucursalKey),
        FOREIGN KEY(fechaKey) REFERENCES dimfecha(fechaKey)
)sortKey(cuentaKey,sucursalKey, fechaKey);

```