

---

# Diseño de un ODS para optimizar el rendimiento de consultas y reportes

---

Autores  
Sebastián Bejarano Zarate

Director  
Andres Julian Aristizabal Cardona

Co-Director  
Mauricio Garcés Restrepo



**Universidad de Bogotá Jorge Tadeo Lozano**  
Facultad de Ciencias Naturales e Ingeniería  
*Especialización en Desarrollo de Bases de Datos*

Bogotá - Colombia, noviembre de 2025

# Índice

	Página
<b>Resumen</b>	<b>v</b>
<b>Abstract</b>	<b>vi</b>
<b>Glosario</b>	<b>vii</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. Descripción del Problema</b>	<b>2</b>
<b>3. Objetivos</b>	<b>3</b>
3.1. Objetivo General . . . . .	3
3.2. Objetivos Específicos . . . . .	3
<b>4. Requerimientos</b>	<b>4</b>
4.1. Requerimiento Funcional . . . . .	4
4.2. Requerimientos de Calidad . . . . .	4
<b>5. Estado del Arte</b>	<b>5</b>
<b>6. Marco Teórico</b>	<b>7</b>
6.1. Bodega de Datos . . . . .	7
6.2. ODS . . . . .	7
6.3. ETL . . . . .	8
<b>7. Solución propuesta</b>	<b>11</b>
7.1. Descripción general de la solución . . . . .	11
7.2. Modelo conceptual . . . . .	12
7.2.1. Fuentes . . . . .	12
7.2.2. Staging . . . . .	12
7.2.3. Bodega de datos . . . . .	14
7.2.4. Extracción . . . . .	14
7.2.5. Transformación . . . . .	14
7.2.6. Carga . . . . .	14
7.3. Estándares de la solución . . . . .	14
7.3.1. Calidad de Datos . . . . .	14
7.3.2. Procesos ETL (Extraer, transformar y cargar) . . . . .	14
7.3.3. Visualización . . . . .	14
7.3.4. Documentación y Mantenimiento . . . . .	15
7.3.5. Rendimiento y Optimización . . . . .	15
7.4. Condiciones de diseño, propuesta de implementación y evaluación . . . . .	15

<b>8. Planeación del Trabajo</b>	<b>17</b>
8.1. Descomposición de actividades WBS . . . . .	17
8.1.1. Cumplimiento del Objetivo Específico 1: Diseño de la Arquitectura ODS	17
8.1.2. Cumplimiento de los Objetivos Específicos 2 y 3: Diseño de los Procesos ETL y Calidad de Datos . . . . .	17
8.1.3. Cumplimiento del Objetivo Específico 4: Operación y Automatización .	18
8.2. Diagrama de Gantt . . . . .	19
<b>9. Presupuesto</b>	<b>21</b>
<b>10. Conclusiones</b>	<b>22</b>
<b>Bibliografía</b>	<b>24</b>

## Índice de figuras

1.	Arquitectura proceso - ETL. Imagen tomada de [18] . . . . .	8
2.	Transformación y agregación de información. Imagen tomada de [21] . . . . .	9
3.	Proceso de carga de información. Imagen tomada de [23] . . . . .	10
4.	Diagrama de arquitectura. Fuente el autor . . . . .	11
5.	Diseño ETL. Fuente el autor . . . . .	11
6.	Cuadrante de Gartner herramientas de integracion de datos. Imagen tomada de [24] . . . . .	13
7.	Diagrama de descomposición de actividades. Fuente el autor . . . . .	18
8.	Diagrama de Gantt. Fuente el autor . . . . .	20

## Índice de tablas

1.	Tabla Comparativa de Arquitecturas de Almacenamiento de Datos. Fuente el autor . . . . .	7
2.	Condiciones de Diseño para el ODS. Fuente el autor . . . . .	15
3.	Condiciones de implementación para el ODS. Fuente el autor . . . . .	16
4.	Condiciones de evaluación para el ODS. Fuente el autor . . . . .	16
5.	Detalles de las actividades de las tareas 1. Fuente el autor . . . . .	19
6.	Detalles de las actividades de las tareas 2. Fuente el autor . . . . .	20
7.	Presupuesto. Fuente el autor. . . . .	21

## Resumen

---

Este proyecto propone el diseño de una bodega de datos tipo ODS, orientada a dar solución a la problemática de consultas eficientes que no afecten el sistema transaccional del ente financiero. Se plantea una solución que integre procesos ETL (extracción, transformación y carga de datos) sobre una arquitectura utilizando Oracle.

La propuesta aborda las principales fuentes de datos transaccionales en una entidad financiera, principalmente transacciones y procesos que conlleven a una gestión contable, además, se establece una estructura que facilita la generación de reportes mediante ETL para las áreas contables para la toma de decisiones.

## Abstract

---

This capstone project addresses the critical issue of performance degradation in transactional systems within the financial sector, which is frequently caused by running complex analytical queries and generating operational reports directly from the Online Transaction Processing (OLTP) database. This resource contention compromises the stability of the business CORE and introduces high data latency, undermining the reliability of information crucial for tactical decision-making. The General Objective of this study was to design an Operational Data Store (ODS) architecture optimized for reporting, implementing a strategic query offloading approach. The solution includes the development of a Logical Data Model (3NF) that consolidates one year of historical data. The proposed methodology details the design of the Extract, Transform, and Load (ETL) processes, which not only ensure secure data migration but also incorporate cleaning and validation mechanisms during the transformation phase to guarantee the quality and consistency of the accounting information. The resulting design establishes an architecture capable of maintaining low data latency, defining criteria for periodicity and automation of data loading, and providing a dedicated repository that significantly enhances the operational agility of the financial entity.

## Glosario

- B-TREE** Árbol B (B-Tree). Es la estructura de índice más común y estándar utilizada en bases de datos relacionales (OLTP y ODS). Está optimizada para búsquedas rápidas, inserciones y eliminaciones, siendo eficiente en columnas con alta cardinalidad (muchos valores únicos). 15
- BITMAP** Índice de Mapa de Bits (Bitmap Index). Es un tipo de índice especializado en bases de datos que utiliza una matriz de bits para identificar las filas que contienen un valor específico en una columna. Es altamente eficiente para columnas con un bajo número de valores únicos (baja cardinalidad), común en Data Warehouses. 15
- CORE** Sistema Central o Núcleo. Se refiere a las aplicaciones y sistemas de información esenciales y críticos que soportan las operaciones fundamentales de una empresa (ej. procesamiento de transacciones financieras, contabilidad, gestión de inventarios). Estos sistemas suelen operar bajo una arquitectura OLTP. 2
- DATAWAREHOUSE** Bodega de Datos o Almacén de Datos (DW). Es un sistema de almacenamiento diseñado para la consulta y el análisis de grandes volúmenes de datos históricos. Su propósito principal es soportar el Business Intelligence (BI) y la toma de decisiones estratégicas, utilizando modelos dimensionales (OLAP). 2
- ETL** Proceso de gestión de datos que consta de tres pasos: Extracción de datos de los sistemas fuente, Transformación de datos para aplicar reglas de negocio y asegurar la calidad, y Carga de los datos en el sistema de destino (Data Warehouse o ODS). 3
- LATENCIA** Retraso o demora en la transmisión o procesamiento de datos. En el contexto de datos, mide el tiempo que transcurre desde que se genera un dato en el sistema fuente hasta que está disponible para su uso en el sistema de destino (ODS o DW). La baja latencia indica alta frescura del dato. 5
- ODS** Almacén de Datos Operacional (Operational Data Store). Es una base de datos centralizada diseñada para integrar datos recientes y detallados de múltiples sistemas transaccionales, con el propósito principal de soportar el reporting operativo y la toma de decisiones tácticas en tiempo real o casi real. 1
- OLAP** Procesamiento Analítico en Línea (Online Analytical Processing). Tecnología de bases de datos optimizada para el análisis complejo de datos multidimensionales, utilizada típicamente en Data Warehouses (DW) para consultas de agregación, segmentación y perforación (drill-down). 6
- OLTP** Procesamiento de Transacciones en Línea (Online Transaction Processing). Clase de sistemas de bases de datos optimizados para manejar un gran volumen de transacciones cortas y concurrentes (ej. inserción, actualización y eliminación de registros), cruciales para las operaciones diarias del negocio (CORE). 5

## 1. Introducción

EL siguiente documento es construido en el marco de proyecto de grado como habilitador para el título de especialista en desarrollo de base de datos, el cual usara los requerimientos construidos para dar solución a la optimización de las consultas y reportes obtenidos del ODS, por lo cual se realizó el diseño lógico y físico para consolidar la información histórica de la base de datos transaccional, con un período de un año de antigüedad.

Lo anteriormente descrito se realizó bajo la metodología Kimball para el desarrollo de las etapas del ciclo de vida del proyecto, enfocándose en lo siguiente:

- Datos: Diseño del modelo dimensional y el proceso de Extracción, Transformación y Carga (Extract, Transformation, and Load - ETL) para cargar el ODS.

## 2. Descripción del Problema

El rendimiento de la base de datos transaccional se ve afectado cuando se utiliza para generar reportes y consultas históricas. La arquitectura de los sistemas de procesamiento transaccional está diseñada para ser eficiente en alta transaccionalidad, mas no en la generación de reportes bajo consultas complejas. La ejecución de estas operaciones compite por recursos con las actividades del CORE de la empresa, como el procesamiento de transacciones, lo que resulta en una ralentización del sistema y una experiencia de usuario deficiente. Como señalan Romney y Steinbart (2003), los DATAWAREHOUSE (DW) suelen ser sistemas separados de los sistemas de información contable (AIS) utilizados para apoyar las actividades comerciales diarias de la organización, de modo que el rendimiento de esas actividades no se vea afectado negativamente por los recursos de procesamiento informático consumidos en el análisis de los datos [1].

Además, la falta de una fuente de datos consolidada dificulta la obtención rápida de la información para generar reportería a tiempo para las áreas financieras, que deben tomar decisiones para el negocio. Esta problemática subraya la necesidad de una solución que optimice el acceso a la información histórica sin comprometer la operación, un concepto que el Dr. Jim Goodnight, CEO de SAS, considera más relevante que nunca ante las crecientes necesidades informativas de la organización [2]. La principal estrategia para incrementar la competitividad de una empresa radica en la organización inteligente de sus datos históricos (transacciones u operaciones diarias), usualmente residiendo en Data Warehouse corporativos o Data Marts departamentales [3] .

El diseño de un Almacén de Datos Operacional (ODS) se presenta como la estrategia para abordar la problemática. Los almacenes de datos tienen como propósito proporcionar información, respaldar la toma de decisiones y alertar a la empresa sobre los patrones detectados[4]. Al consolidar datos transaccionales con un año de antigüedad, el ODS servirá como un repositorio dedicado para fines de reporting, liberando la base de datos principal de cargas innecesarias y proporcionando una plataforma eficiente para el análisis y la toma de decisiones estratégicas en el área contable y operativa. A diferencia del Data Warehouse tradicional, donde la información se encuentra altamente desnormalizada y modelada en esquemas de estrella o copo de nieve [5], el ODS mantiene un nivel de detalle operativo y una estructura que facilita la presentación de reportes con información reciente.

### **3. Objetivos**

#### **3.1. Objetivo General**

Diseñar una arquitectura de Almacén de Datos Operacional (ODS) que optimice el rendimiento de las consultas y la generación de reportes, utilizando un año de datos históricos de la base de datos transaccional y con un modelo de carga bajo ETL para la toma de decisiones contables y operativas.

#### **3.2. Objetivos Específicos**

- Objetivo específico 1: Diseñar la arquitectura de un Almacén de Datos Operacional (ODS) que optimice la generación de reportes contables y operativos, modelo de datos lógico y requisitos de información a partir de fuentes de datos clave.
- Objetivo específico 2: Diseñar los procesos de Extracción, Transformación y Carga (ETL) para migrar de forma segura y eficiente los datos históricos (un año) desde la base de datos transaccional al ODS.
- Objetivo específico 3: Proponer mecanismos de limpieza y validación de datos durante la fase de transformación para garantizar la calidad y la consistencia de la información en el ODS.
- Objetivo específico 4: Establecer la periodicidad y la automatización de los procesos ETL para mantener el ODS actualizado.

## 4. Requerimientos

### 4.1. Requerimiento Funcional

- **Reporting Contable y Operativo Consolidado:** El sistema debe permitir a los usuarios del área financiera y de operaciones generar reportes que consoliden datos históricos de hasta un año para identificar tendencias, patrones de ventas, costos, sin afectar el rendimiento del sistema transaccional.

### 4.2. Requerimientos de Calidad

- **Integridad y Precisión de Datos:** Los datos cargados en el ODS deben tener una precisión del 99 % en comparación con la fuente de origen, y los errores de transformación no deben superar el 0.1 %.
- **Escalabilidad del Almacén:** La arquitectura debe ser capaz de soportar un crecimiento anual del 20 % en el volumen de datos durante los próximos 3 años sin degradar el rendimiento.

## 5. Estado del Arte

La gestión de los datos en el sector financiero ha sido un tema de constante evolución. Tradicionalmente, se ha dependido de sistemas transaccionales OLTP (Online Transaction Processing) para la operación diaria, lo cual ha generado ineficiencias al intentar obtener información histórica y consolidada para la toma de decisiones [6]. Esta dependencia genera desventajas significativas, ya que usar el sistema OLTP principal sobrecarga la base de datos, lo que podría resultar en degradación del rendimiento o tiempo de inactividad para este sistema crítico [7]. Por naturaleza, los sistemas OLTP no fueron diseñados para gestionar consultas complejas [7].

En la actualidad, las organizaciones buscan optimizar sus procesos de toma de decisiones utilizando sistemas integrados de almacenamiento y procesamiento de datos operacionales. Uno de los componentes clave en este tipo de infraestructuras es el Operational Data Store (ODS). A diferencia de una bodega de datos tradicional (Data Warehouse), que se enfoca en el análisis histórico y estratégico, el ODS tiene como principal característica la actualización continua y la integración de datos provenientes de sistemas transaccionales, permitiendo decisiones más ágiles y oportunas [8]. Además, los almacenes de datos analíticos, aunque adecuados para cargas de trabajo analíticas, no cubren casos de uso donde la frescura de los datos se mide en milisegundos [7].

El diseño de un ODS surge para abordar esta deficiencia. El propósito fundamental de un ODS es integrar datos corporativos de diferentes fuentes heterogéneas para facilitar la generación de informes operativos en tiempo real o casi real [9]. El ODS funciona como una alternativa a una aplicación de sistema de apoyo a la toma de decisiones que accede a los datos directamente desde un sistema de procesamiento de transacciones en línea (OLTP) [10]. El ODS permite una estrategia de descarga OLTP, esencial para la estabilidad operativa, al permitir que los sistemas OLTP se centren en atender consultas simples de lectura/escritura a gran escala, mientras el ODS gestiona cargas de trabajo complejas con uso intensivo de lectura [7].

La eficacia del ODS se mide por su capacidad de reducir la LATENCIA de datos, que se refiere al tiempo que transcurre entre el envío de datos para su uso y el momento en que se obtiene el resultado deseado [11]. La latencia puede deberse a limitaciones del hardware y los tiempos de procesamiento del software [11].

Durante la construcción del ODS, se asegura que los datos tengan una estructura similar a la de los sistemas de origen, aunque durante la integración se pueden depurar, desnormalizar y aplicar reglas de negocio para garantizar su integridad [9].

Esta integración se realiza al nivel más bajo de detalle y con bastante frecuencia a lo largo del día [9]. Esta mejora en la calidad del dato es crítica, dado que los Sistemas de Información Contable (SIC) se han convertido en herramientas clave para la administración empresarial, permitiendo la integración de datos financieros y operativos en la toma de decisiones estratégicas [12].

Desde la perspectiva del diseño, la construcción de un ODS comienza con el diseño lógico, que consiste en obtener un esquema que depende del tipo de base de datos elegido, aunque es independiente de la implementación concreta del sistema de gestión de bases de datos (SGBD) [13].

En cuanto a las plataformas, la implementación de almacenes de datos ha evolucionado. Las organizaciones tienen diversas opciones, que incluyen software de sistemas de gestión de bases de datos (DBMS) convencional, DBMS analíticos especializados, appliances de almacenamiento de datos, y almacenes de datos en la nube [14]. Específicamente, el uso de la nube se ha convertido en una opción más viable, con tecnologías de almacenamiento de datos como servicio (DWaaS) [14].

Un estudio de Ferreira, Almeida y Monteiro (2017) validó esta premisa al implementar un Data Warehouse con cubos OLAP para optimizar la reportería financiera. Sus resultados mostraron una mejora “extremadamente significativa” en el rendimiento, con ahorros de tiempo considerables en la generación de reportes recurrentes [15]. Dicha investigación concluyó que el uso de estas arquitecturas no solo reduce los tiempos de consulta, sino que también automatiza y robustece el proceso de análisis, ofreciendo una solución más flexible y menos propensa a errores que los métodos manuales.

Finalmente, es importante destacar que, a diferencia de los datos de un almacén de datos maestros, los datos ODS no se devuelven a los sistemas operativos, aunque pueden transferirse a otras operaciones y al Data Warehouse (DW) para la generación de informes [10].

Característica	OLTP (Online Transaction Processing)	ODS (Operational Data Store)	DW (Data Warehouse)
Propósito Primario	Registrar transacciones diarias y operaciones [6].	Soporte para reporting operativo y toma de decisiones tácticas [9].	Soporte para análisis estratégico e histórico a largo plazo.
Volumen de Consultas	Muchas consultas cortas (lectura/escritura) tácticas [7].	Consultas medianas, intensivas en lectura.	Pocas consultas, pero muy complejas y largas.
Actualidad de la información	En tiempo real (instante de la transacción).	Casi en tiempo real o baja latencia (minutos u horas) [9].	Histórico (generalmente, con un retraso de horas o días).
Volatilidad	Alta (los datos cambian constantemente con las transacciones).	Media/Alta (se actualizan para reflejar el estado actual) [10].	Baja (los datos son no volátiles; solo se añaden).
Impacto en el Core	Crítico (consultas complejas causan degradación del rendimiento) [7].	Nulo (permite el OLTP Offload de consultas).	Nulo (permite el OLTP Offload de análisis).
Modelo de Datos Típico	Altamente Normalizado (3NF) [13].	Normalizado (similar a 3NF) para detalle atómico tácticas [9].	Desnormalizado (Esquema Estrella/Copo de Nieve).
Ventana de Datos	Datos actuales (días o semanas, lo necesario para la operación).	Datos recientes (ej. un año, según tu proyecto).	Datos históricos (años o décadas).

Tabla 1: Tabla Comparativa de Arquitecturas de Almacenamiento de Datos. Fuente el autor

## 6. Marco Teórico

### 6.1. Bodega de Datos

Un Data Warehouse es un sistema de gestión de información cuya arquitectura está diseñada para el almacenamiento no volátil, integrado y orientado por temas de datos históricos. Su propósito principal es servir como base de conocimiento para el soporte de decisiones de la organización [8]. Este almacenamiento es útil para la toma de decisiones de cualquier organización.

### 6.2. ODS

Un ODS es un tipo de almacén de datos que proporciona el valor de los datos y es admisible un retraso de los datos [16]. Además, no suele contener información temporal [17]. Esto es útil porque funciona como fuente de datos integrada para generación de reporting.

### 6.3. ETL

“La migración de datos de las fuentes de datos al datawarehouse requiere un proceso para extraer, transformar y cargar dichos datos. Este proceso se conoce como ETL” [18].

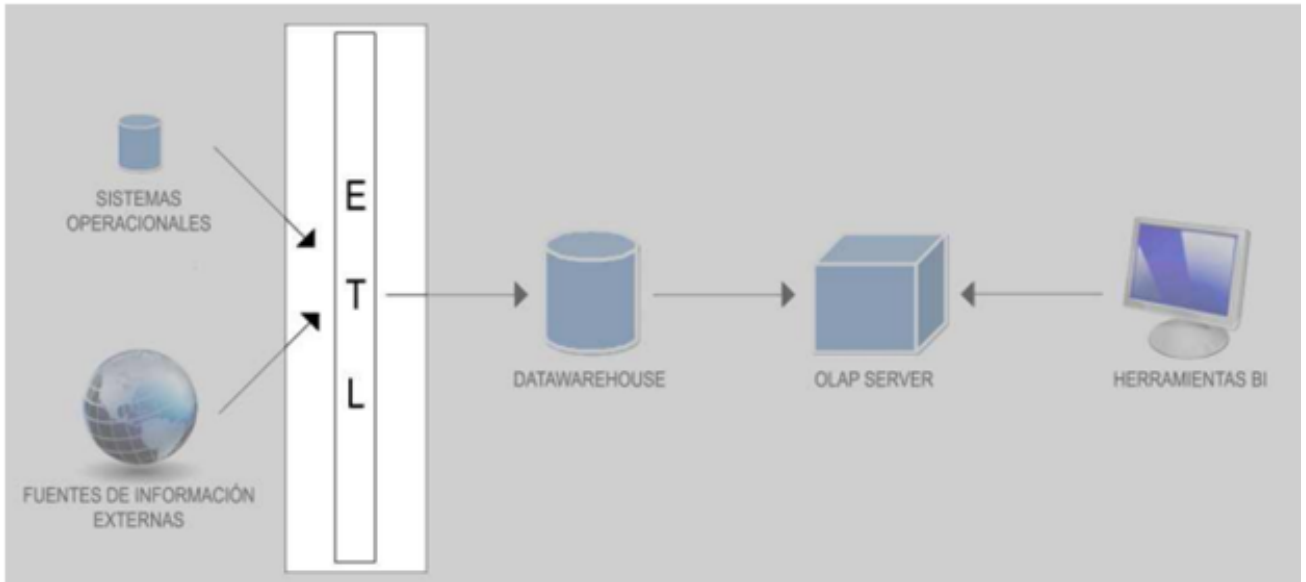


Figura 1: Arquitectura proceso - ETL. Imagen tomada de [18]

La importancia de la ETL en el proceso de cargar un DATAWAREHOUSE es fundamental y es el más costoso y complejo del proceso, ya que es usado para realizar la limpieza y/o transformación de los datos [18], el cual se realiza continuamente, según necesidades, las cuales pueden llegar a ser incluso diarias en su gran mayoría.

#### **Extracción**

Es un subproceso de la ETL, este es el encargado de traer los datos de la o las fuentes de información [19].

#### **Transformación**

Este paso esta destinado a la integración, derivación, agregación y totalización de datos. En donde dos de los más importantes son:

Integración: se refiere al nombramiento de campos de acuerdo con estándares del datawarehouse [20]

Derivación: se basa en el calculo de nuevos campos derivados de un campo original, como subdividir una fecha en día, año, minutos y segundos [21]

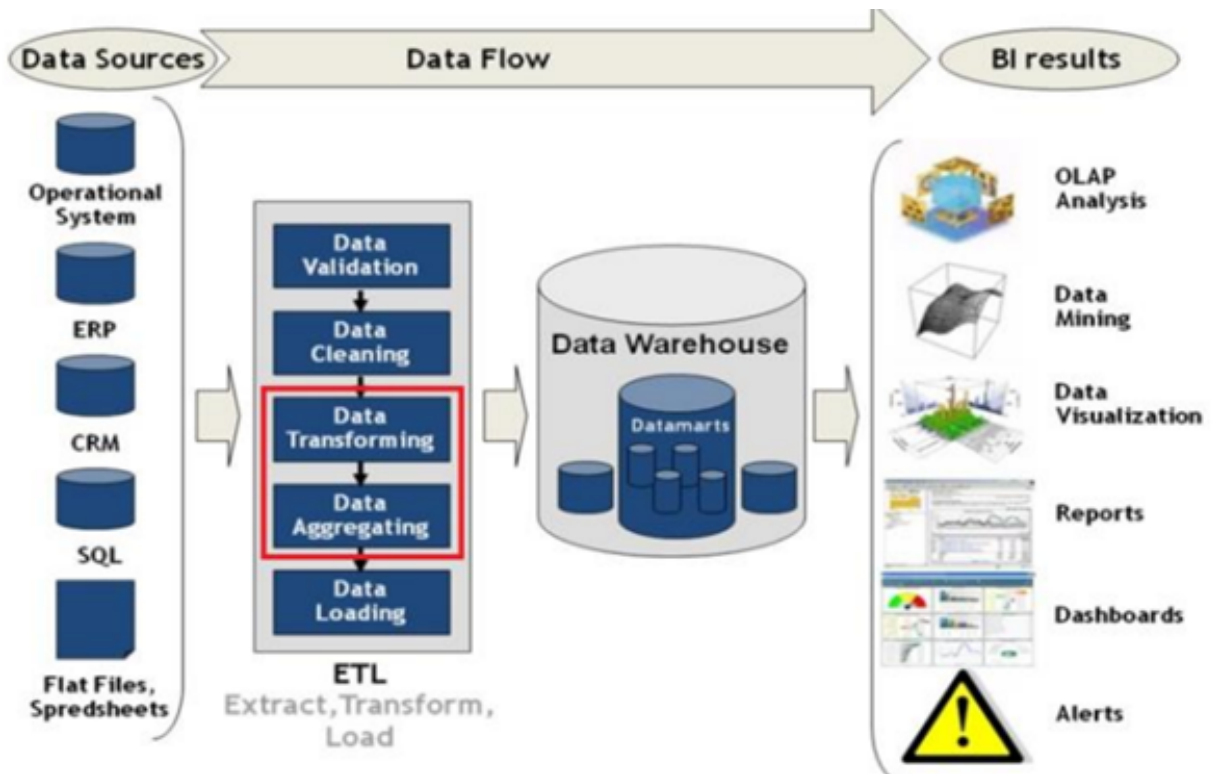


Figura 2: Transformación y agregación de información. Imagen tomada de [21]

El proceso de transformación se realiza sobre un área llamada staging, la cual consiste en un área intermedia de almacenamiento, con menos recursos que el área final, esto debido a que está orientada específicamente a realizar la transformación antes descrita, por ende, suele ser un área volátil [22].

## Carga

Este proceso consiste en el paso final de la ETL, donde se realiza la carga de los datos transformados en el datawarehouse [23].

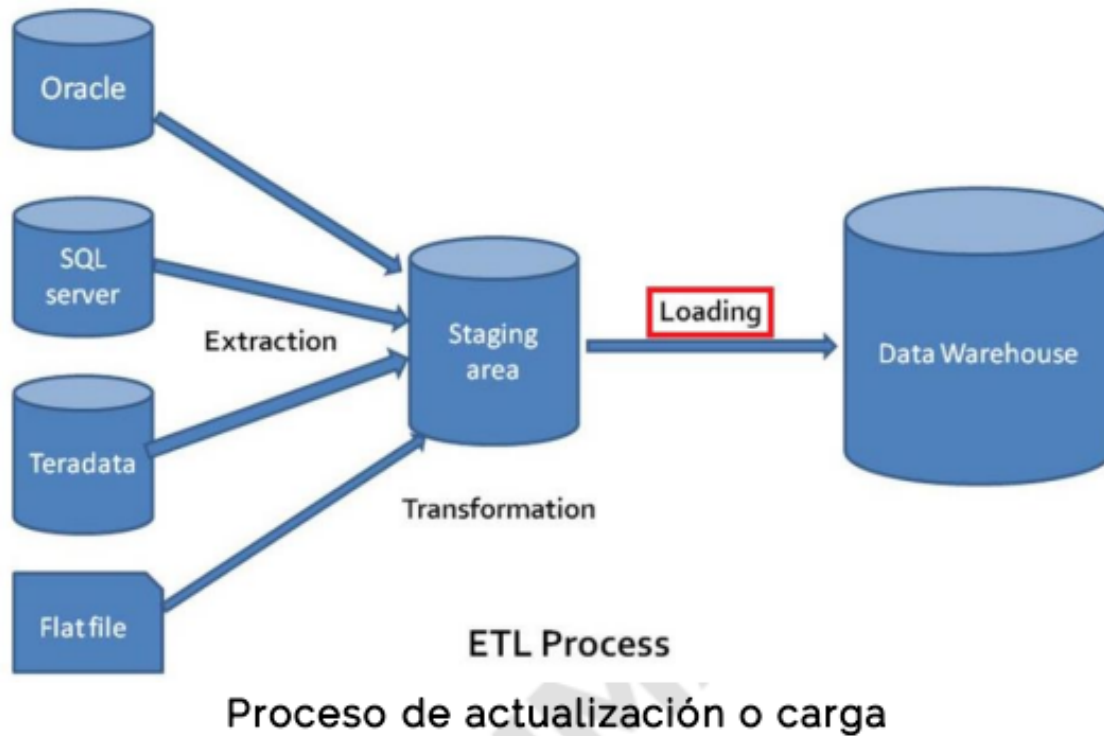


Figura 3: Proceso de carga de información. Imagen tomada de [23]

## 7. Solución propuesta

### 7.1. Descripción general de la solución

Con base en la necesidad actual de la entidad financiera, se plantea el diseño de una bodega de datos tipo ODS para el almacenamiento de la transaccionalidad del Core del negocio. Esto surge de la necesidad de tomar decisiones para futuras acciones de la compañía.

Se contempla un proceso con la estructura necesaria para realizar las fases esenciales para una bodega tipo ODS, esto según distintas investigaciones descritas en el marco teórico, logrando así, asegurar la integración de los datos, y posteriormente su visualización.

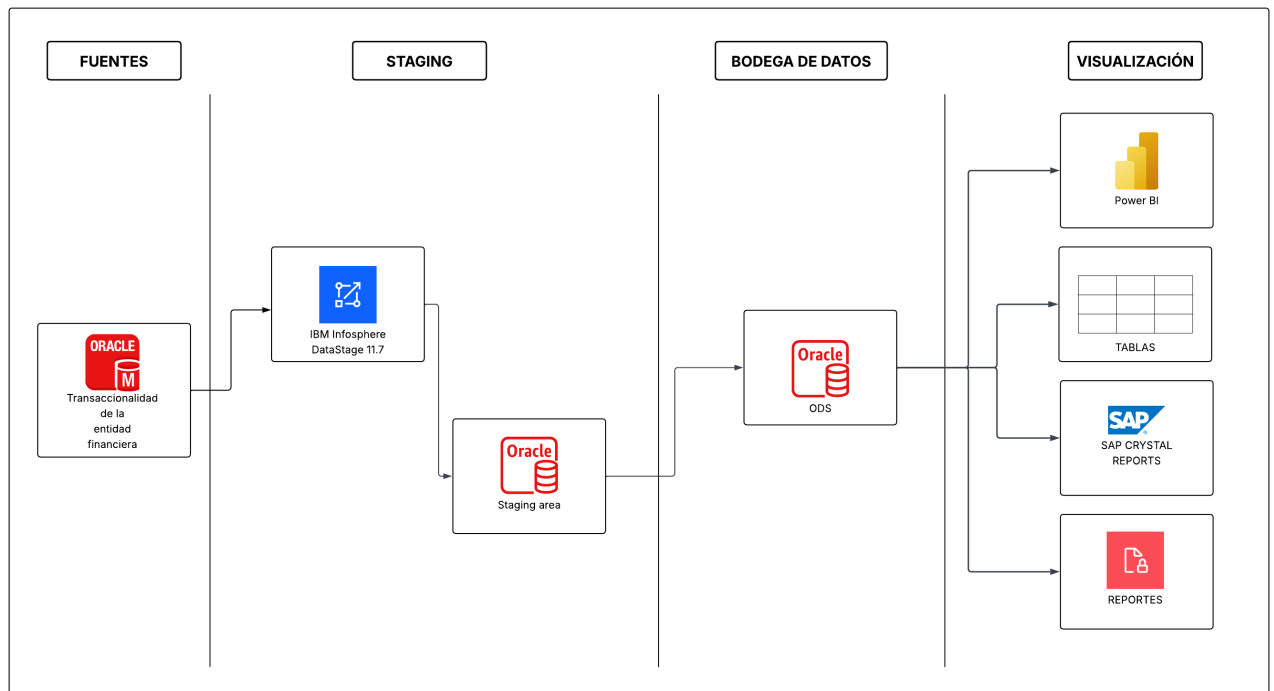


Figura 4: Diagrama de arquitectura. Fuente el autor

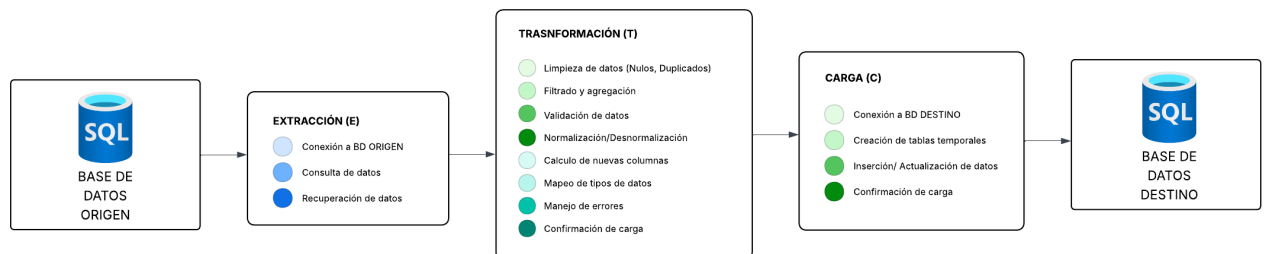


Figura 5: Diseño ETL. Fuente el autor

## **7.2. Modelo conceptual**

### **7.2.1. Fuentes**

Corresponde a la transaccionalidad diaria de la entidad financiera, la cual se guarda originalmente en una base de datos de Oracle, específicamente Flexcube, orientada al universo bancario por el funcionamiento específico que posee. Además, contiene una amplia gama de productos bancarios, como cuentas corrientes, cuentas de ahorro, depósitos fijos, depósitos recurrentes y pequeños ahorros.

### **7.2.2. Staging**

Para el proceso de ETL se sugiere usar IBM INFOSPHERE DATASTAGE, por su amplio contenido de conexiones a diferentes BD y herramientas múltiples para transformación, depuración y tratamiento general de los datos antes de ser cargados, así mismo, se encuentra dentro de las compañías líderes para herramientas de integración de datos.

Figure 1: Magic Quadrant for Data Integration Tools



Gartner

Figura 6: Cuadrante de Gartner herramientas de integracion de datos. Imagen tomada de [24]

### 7.2.3. Bodega de datos

Una vez transformados los datos, se guardarán en una instancia de Oracle Flexcube, creada con el nombre ODS y ubicada en un nodo diferente para su consulta más eficaz, aunque a pesar de estar en Oracle, las tablas de ODS no estarán relacionadas de acuerdo con el tipo de DATA WAREHOUSE que se espera tener.

### 7.2.4. Extracción

Para el diseño de la ETL, en su extracción se debe componer de la conexión a la base de datos para consultar y recuperar los datos para almacenar en el DATA WAREHOUSE.

### 7.2.5. Transformación

Para el diseño de la ETL, en su transformación se deben tener en cuenta los procesos necesarios para normalizar la información la cual debe ser cargada, esto mediante normalización, agregación, filtrado, mapeos o cálculos de campos.

### 7.2.6. Carga

Para el diseño de la ETL, en su carga se deben tener en cuenta la carga en tablas temporales para pasar a las tablas permanentes en la base de datos destino, esto una vez transformados los datos.

## 7.3. Estándares de la solución

### 7.3.1. Calidad de Datos

- Integridad: Los datos almacenados deben ser consistentes y coherentes.
- Actualización: La información debe actualizarse diariamente.
- Depuración: Los datos extraídos deben pasar por tratamientos para eliminar duplicados, errores y registros incompletos.

### 7.3.2. Procesos ETL (Extraer, transformar y cargar)

- Extracción: Extraer exclusivamente los datos necesarios para asegurar la optimización en el almacenamiento.
- Transformación: Según las reglas de negocio, se debe estandarizar la agregación, limpieza, filtrados, cálculos, etc. de los datos.
- Carga Programada: Automatizar la carga diaria de datos en horarios de baja transaccionalidad.

### 7.3.3. Visualización

- Herramientas BI Estándar: Utilizar Power BI o SAP CRYSTAL REPORTS para dashboards dinámicos, filtros en tiempo real y exportación de informes.

#### 7.3.4. Documentación y Mantenimiento

- Documentación: Registrar el diseño y ejecución de ETL, esquemas de bases de datos y flujos de datos.
- Versionado y Actualización: Realizar control de versiones para cambios y procesos en el cargue de información.

#### 7.3.5. Rendimiento y Optimización

- Indexación: Aplicar índices B-TREE o BITMAP a campos clave para mejorar tiempos de respuesta en informes.
- Particionamiento: Dividir históricamente por año o períodos para optimizar consultas.

### 7.4. Condiciones de diseño, propuesta de implementación y evaluación

En la siguiente tabla, se especificaran las condiciones de diseño, propuesta de implementación y evaluación.

Tabla 2: Condiciones de Diseño para el ODS. Fuente el autor

<b>DISEÑO</b>	
<b>Condición de diseño</b>	<b>Descripción</b>
Auditoría y Trazabilidad	Debe incluir campos para registrar la fecha y hora de la carga.
Retención de Datos Limitada	Se debe definir un periodo de tiempo específico para la retención de datos(1 año).
Consolidación e Integración	El ODS debe unificar datos de múltiples tablas de origen en un modelo coherente, eliminando inconsistencias y duplicados.

Tabla 3: Condiciones de implementación para el ODS. Fuente el autor

<b>IMPLEMENTACIÓN</b>	
<b>Fase</b>	<b>Descripción</b>
Análisis y Modelado	Definir las dimensiones y hechos para el modelado físico, basado en los requisitos iniciales.
Configuración de la Infraestructura	Instalar y configurar el servidor de bases de datos que alojará el ODS. Crear los esquemas, tablas e índices diseñados.
Desarrollo de los Procesos ETL	Crear los scripts o flujos de trabajo que extraerán datos de los sistemas de origen, aplicarán las transformaciones necesarias para el ODS, y cargarlos de forma incremental.
Desarrollo de Capa de Reporte	Crear vistas SQL optimizadas sobre el ODS que sirvan como fuente de datos directa para las herramientas de Business Intelligence (BI).
Pruebas Iniciales	Realizar pruebas de carga y funcionalidad para asegurar que el ODS maneja el volumen de datos esperado y que los reportes generados son precisos.

Tabla 4: Condiciones de evaluación para el ODS. Fuente el autor

<b>EVALUACIÓN</b>	
<b>Métrica</b>	<b>Descripción</b>
Tiempo de Respuesta de Consultas	Medir el tiempo que tarda un conjunto de consultas críticas en ejecutarse antes (en el sistema de origen) y después (en el ODS).
Tasa de Errores de Datos	Evaluar la precisión de los datos cargados en el ODS, verificando que no se introdujeron inconsistencias durante el proceso ETL.

## 8. Planeación del Trabajo

### 8.1. Descomposición de actividades WBS

El presente diseño se estructura en fases secuenciales y lógicas que buscan alcanzar el Objetivo General: la implementación exitosa de un Almacén de Datos Operacional (ODS) con fines de reporting contable y operativo. La metodología de trabajo se alinea directamente con los objetivos específicos (OE) definidos.

#### 8.1.1. Cumplimiento del Objetivo Específico 1: Diseño de la Arquitectura ODS

El Objetivo Específico 1 se centra en el diseño fundamental y arquitectónico del ODS. Se cumplirá principalmente a través de la fase 01-01 Análisis y la subfase 01-02-100 Definición y Diseño de la Arquitectura ODS.

- **Recolección y Mapeo:** Se iniciará con la Recolección de Requerimientos de Información de las áreas contable y operativa para definir exactamente qué reporting es necesario. Esto se traduce en la Identificación de Fuentes de datos clave en el sistema transaccional (OLTP).
- **Modelado Lógico:** A partir de los requisitos, se procederá al Diseño del Modelo de Datos Lógico del ODS. Dado que el ODS prioriza el detalle y la frescura, este modelo se centrará en una estructura más normalizada (3NF) para facilitar el reporting operativo y la consolidación de entidades.
- **Arquitectura Física:** Finalmente, el Diseño del Modelo Físico y el Diseño de la Arquitectura Técnica establecerán la infraestructura (servidor, software de base de datos) y las especificaciones técnicas para soportar la carga histórica y las consultas de baja latencia.

#### 8.1.2. Cumplimiento de los Objetivos Específicos 2 y 3: Diseño de los Procesos ETL y Calidad de Datos

Estos dos objetivos están ligados a la subfase 01-02-200 Diseño de los Procesos de Carga (ETL).

- **Extracción (E):** El Diseño del Proceso de Extracción se enfocará en la migración inicial de un año de datos históricos desde la base de datos transaccional, definiendo la estrategia de acceso más segura y eficiente para no degradar el rendimiento del CORE del negocio.
- **Transformación y Calidad (T):** El Diseño del Proceso de Transformación será la clave del Objetivo Específico 3. Se propone el Diseño de Mecanismos de Calidad de Datos, estableciendo reglas para la limpieza de datos y la validación para garantizar la confiabilidad de la información contable.
- **Carga (L):** El Diseño del Proceso de Carga definirá las reglas para insertar los datos transformados en el Modelo Físico del ODS, priorizando la optimización de los índices para la velocidad de consulta del reporting.

### 8.1.3. Cumplimiento del Objetivo Específico 4: Operación y Automatización

El Objetivo Específico 4 asegura que el ODS se mantenga actualizado y operativo en el tiempo, lo cual se logra con la subfase 01-02-300 Diseño de la Operación y Mantenimiento.

- Periodicidad y Frecuencia: El Diseño de la Periodicidad de Carga establecerá el ciclo de vida del dato en el ODS. Esto es para asegurar la actualidad de la información requerida para la toma de decisiones.
- Automatización: El Diseño de la Automatización del ETL definirá las herramientas de orquestación que ejecutarán los procesos ETL, generalmente en una ventana de tiempo nocturna para garantizar la mínima latencia sin impactar las operaciones.
- Documentación: la Documentación del Proceso de Mantenimiento servirá como manual de referencia para monitorear el rendimiento del ODS, gestionar errores de calidad de datos y asegurar la continuidad operativa del sistema.

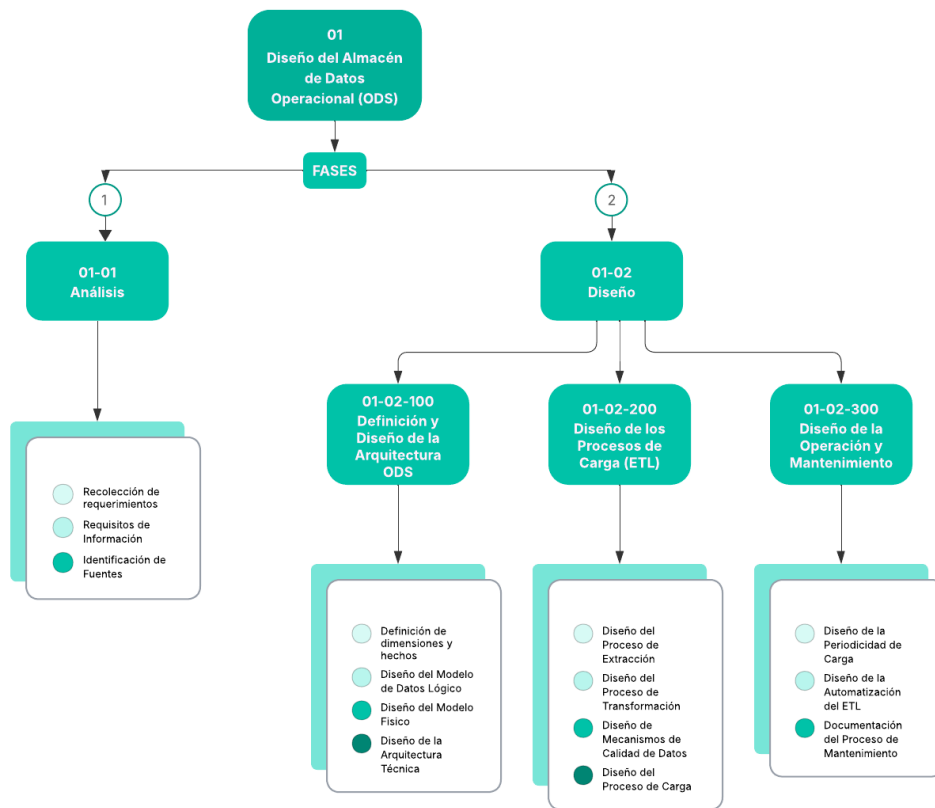


Figura 7: Diagrama de descomposición de actividades. Fuente el autor

## 8.2. Diagrama de Gantt

La Tabla 5 y 6 muestra la descripción de las tareas en la planificación del proyecto, compuestas por su recurso necesario, duración y riesgo.

TAREAS			
Actividad	Recursos	Tiempo (horas)	Riesgo
Recolección de requerimientos	Acceso a reunion con area funcional	20	Retraso en la identificación de necesidades
Requisitos de Información	Acceso a reunion con area funcional	20	Retraso en la identificación de reportes
Identificación de Fuentes	Acceso a bases de datos	35	Retraso de inicio de diseños de dimensiones
Definición de dimensiones y hechos	Herramienta de diagramacion	35	Retraso de inicio de diseño de modelos
Diseño del Modelo de Datos Lógico	Herramienta de diagramacion	8	Retraso de inicio de diseño de modelo fisico
Diseño del Modelo Fisico	Herramienta de diagramacion	8	Retraso de inicio de diseño de arquitectura
Diseño de la Arquitectura Técnica	Herramienta de diagramacion	8	Retraso de inicio de diseño de ETL

Tabla 5: Detalles de las actividades de las tareas 1. Fuente el autor

TAREAS 2			
Diseño del Proceso de Extracción	Herramienta de diagramación	10	Retraso de inicio de procesos de transformación
Diseño del Proceso de Transformación	Herramienta de diagramación	10	Retraso de inicio de procesos de calidad
Diseño de Mecanismos de Calidad de Datos	Herramienta de diagramación	10	Retraso de inicio de procesos de carga
Diseño del Proceso de Carga	Herramienta de diagramación	10	Retraso del diseño de automatización
Diseño de la Periodicidad de Carga	Herramienta de diagramación	5	Retraso del diseño de automatización
Diseño de la Automatización del ETL	Herramienta de diagramación	5	Retraso de la documentación final
Documentación del Proceso de Mantenimiento	Herramienta de documentación	8	

Tabla 6: Detalles de las actividades de las tareas 2. Fuente el autor

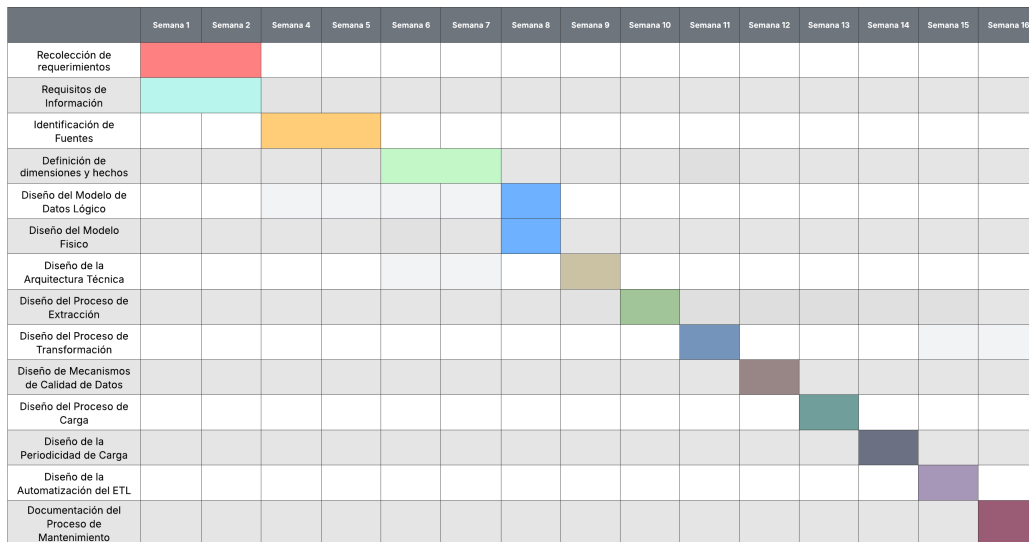


Figura 8: Diagrama de Gantt. Fuente el autor

## 9. Presupuesto

En la tabla 7 se observa un estimado de presupuesto, el cual tiene un valor de \$77,956,000(COP) para realización del proyecto, esto bajo una estimación de duración de 16 semanas, aquí se detallan los principales factores a cubrir, como lo son personal esencial y herramientas para cumplir sus funciones, cada ítem se mide bajo cantidad de horas o meses, con el respectivo precio unitario y total por ítem.

<b>Presupuesto</b>			
<b>Item</b>	<b>Cantidad/Mes</b>	<b>Precio Unitario(COP)</b>	<b>Total(COP)</b>
Horas de trabajo de los estudiantes	288	\$20.000	\$5.760.000
Coordinador académico	1	\$3.500.000	\$3.500.000
Arquitecto de Datos	2	\$5.500.000	\$11.000.000
Lucidchart	6	\$116.000	\$696.000
Analista de Requisitos	1	\$3.000.000	\$3.000.000
Especialista ETL	2	\$5.000.000	\$10.000.000
Herramienta ETL	2	\$10.000.000	\$20.000.000
Herramienta de BI/Reporting	2	\$12.000.000	\$24.000.000
<b>TOTAL ESTIMADO</b>			<b>\$77.956.000</b>

Tabla 7: Presupuesto. Fuente el autor.

El rubro de Horas de trabajo de los estudiantes \$5,760,000(COP) corresponde a un total de 288 horas estimadas, calculadas sobre la duración total de 16 semanas del proyecto. Esto incluye actividades de investigación, documentación y elaboración de entregables.

Los costos de personal se calcularon con base en la distribución de las actividades del WBS 7. El costo del Arquitecto de Datos (2 meses) y el Especialista ETL (2 meses) reflejan la duración de las fases críticas de Diseño de Arquitectura (OE 1) y Diseño de Procesos (OE 2, OE 3, OE 4), respectivamente. El costo del Analista de Requisitos (1 mes) se concentra en la fase inicial de Análisis (WBS 01-01), esencial para la correcta definición de los requisitos de reporting.

El valor presupuestado para las herramientas de desarrollo, como lo son Herramienta ETL y BI, se basan bajo el supuesto único de desarrollo (2 meses), no se contempla pago de licencias para correcciones o soporte, adicional, para la visualización de reporteria tampoco está contemplado el pago para la visualización de los mismos por un tiempo indefinido.

## 10. Conclusiones

El desarrollo de este proyecto de grado logró satisfactoriamente el Objetivo General de diseñar una Arquitectura de Almacén de Datos Operacional (ODS) para optimizar el rendimiento del reporting contable y operativo, demostrando la viabilidad de la solución a través de los siguientes puntos clave:

- **Arquitectura Operacional Diseñada (OE 1):** Se entregó una arquitectura de ODS validada y un Modelo de Datos Lógico de alta granularidad que garantiza la descarga de consultas (OLTP Offload). Este logro es fundamental, ya que protege el rendimiento y la disponibilidad del sistema transaccional (CORE), resolviendo la ineficiencia histórica de ejecutar consultas complejas directamente en la base de datos operativa.
- **Procesos ETL Definidos y Orientados a Calidad (OE 2 y OE 3):** Se diseñaron los procesos ETL para la migración segura y eficiente del histórico de un año. Fundamentalmente, se propusieron mecanismos específicos de limpieza y validación durante la fase de Transformación (OE 3), asegurando que los datos contables y operativos que alimentan el ODS cumplan con los criterios de consistencia y calidad, mitigando la principal causa de desconfianza en la reportería manual.
- **Diseño para la Continuidad Operacional (OE 4):** Se establecieron los criterios de periodicidad y automatización del ciclo de carga, confirmando que el ODS puede mantenerse con baja latencia para entregar información con la actualidad requerida para la toma de decisiones tácticas, garantizando la continuidad operativa del reporting.
- **Optimización de Recursos y Estabilidad (Mitigación de Riesgos):** La implementación del ODS representa una estrategia de optimización de recursos operacionales. Al eliminar la contienda por recursos entre transacciones y consultas, se incrementa la disponibilidad y el tiempo de respuesta del sistema transaccional crítico, lo que a su vez extiende la vida útil del hardware OLTP.
- **Estandarización y Fiabilidad del Reporting:** El diseño del ODS establece una única fuente de verdad para los reportes operativos, ya que centraliza la limpieza y aplicación de reglas de negocio durante el ETL. Esto elimina errores y discrepancias inherentes a la extracción manual de datos, proporcionando una base sólida para la toma de decisiones.
- **Viabilidad de Implementación y Escala:** El diseño propuesto asegura la viabilidad técnica de la implementación inicial, ya que el Modelo Lógico del ODS está diseñado para funcionar como un repositorio intermedio de alta calidad. Esto posiciona al ODS como la fuente de alimentación óptima y limpia para futuros proyectos de Data Warehouse (DW) o de análisis avanzados.

Dado que el proyecto se centró en la fase de diseño, se recomienda la fase de construcción e implementación real del ODS. Esto debe incluir una evaluación comparativa de rendimiento (benchmarking) entre el ODS y el sistema OLTP original, para cuantificar la mejora en los tiempos de consulta, además, se sugiere explorar la integración de herramientas de machine learning (ML) o análisis predictivo sobre los datos limpios y de baja latencia del ODS para desarrollar modelos de alerta temprana o detección de anomalías financieras, maximizando el valor estratégico de la arquitectura diseñada.

## Referencias

- [1] D. C. Prodromos, V. Pavlos y C. Sofoklis, *Knowledge Management Practices in the Financial Sector*. 2005, pág. 4. dirección: <https://research.ebsco.com/c/utjqkg/viewer/pdf/ipjmizgjmn?modal=cite>.
- [2] P. Crosman, *Far from Outdated, Data Warehouse Is Ready to Shine, SAS CEO Says*. 2014, pág. 1. dirección: <https://research.ebsco.com/c/utjqkg/viewer/html/xk5xmfkvvr>.
- [3] ITMadrid, *Qué es Inteligencia de Negocios (Business Intelligence)*. 2008, pág. 1. dirección: <https://www.itmadrid.com/que-es-inteligencia-de-negocios-business-intelligence/#>.
- [4] P. Sanjiv, «Handbook of Data Management 1999 Edition,» pág. 21, 2011. DOI: <https://doi.org/10.1201/9780429114878>.
- [5] R. Salvador, *Microsoft Business Intelligence: vea el cubo medio lleno*. 2016, pág. 11, ISBN: 978-84-936417-8-8. dirección: [https://www.amazon.com/-/es/Salvador-Ramos-ebook/dp/B01END4EYC#detailBullets\\_feature\\_div](https://www.amazon.com/-/es/Salvador-Ramos-ebook/dp/B01END4EYC#detailBullets_feature_div).
- [6] R. Kimball y M. Ross, *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3<sup>ª</sup> ed.* Kimball Group, 2013, pág. 601, ISBN: 978-1-118-53080-1.
- [7] B. Kevin, *5 Methods for OLTP Offload*. 2024, pág. 1. dirección: <https://materialize.com/guides/5-methods-oltp-offload>.
- [8] A. Abello, J. Curto, A. Rius, M. Serra, J. Samos y J. Vidal, *Diseño y construcción de un almacén de datos*. 2016, pág. 1. dirección: <https://hdl.handle.net/10609/136246>.
- [9] V. B. Henry Fernando y A. W. Lissette Sara, «Los sistemas de información contable en la administración. Una revisión sistemática,» pág. 269, 2024, ISSN: 2588-073X. DOI: [https://doi.org/10.26820/recimundo/8.\(3\).julio.2024.269-286](https://doi.org/10.26820/recimundo/8.(3).julio.2024.269-286).
- [10] K. Jason, *Browsing the Operational Data Store (ODS)*. 2018, pág. 1. dirección: <https://www.iri.com/blog/data-transformation2/browsing-the-operational-data-store-ods>.
- [11] S. James, *Operational Data Store (ODS) Defined*. 2015, pág. 1. dirección: <https://www.jamesserra.com/archive/2015/02/operational-data-store-ods-defined/#:~:text=An%20ODS%20is%20meant%20for,the%20entire%20history%20of%20data>.
- [12] C. Michael, *What Is Data Latency? An Expert Guide*. 2024, pág. 1. dirección: [https://www-oracle-com.translate.google/data-latency/?\\_x\\_tr\\_sl=en&\\_x\\_tr\\_tl=es&\\_x\\_tr\\_hl=es&\\_x\\_tr\\_pto=tc](https://www-oracle-com.translate.google/data-latency/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=tc).
- [13] B. I. Xavier y C. O. Josep, *Diseño lógico de bases de datos*. 2020, pág. 1. dirección: <https://openaccess.uoc.edu/server/api/core/bitstreams/bf7eed3b-b758-4d1c-8c2a-54d586525a3e/content>.
- [14] M. Craig, *Evaluating data warehouse deployment options and use cases*. 2022, pág. 1. dirección: <https://www.techtarget.com/searchdatamanagement/feature/Evaluating-your-need-for-a-data-warehouse-platform>.

- [15] J. Ferreira, F. Almeida y J. Monteiro, «Building an Effective Data Warehousing for Financial Sector,» *Automatic Control and Information Sciences*, pág. 10, 2017. DOI: 10.12691/acis-3-1-4.
- [16] A. Abello, J. Curto, C. Llorach y J. Samos, *Explotación y administración de sistemas de data warehouse*. 2016, pág. 23. dirección: <https://hdl.handle.net/10609/136206>.
- [17] A. Abello, J. Curto, A. Rius, M. Serra, J. Samos y J. Vidal, *Diseño y construcción de un almacén de datos*. 2016, pág. 23. dirección: <https://hdl.handle.net/10609/136246>.
- [18] C. U. Asturias, *Almacenamiento y Procesamiento de la Información*. 2020, pág. 7. dirección: [https://www.centro-virtual.com/recursos/biblioteca/pdf/herramientas\\_big\\_data\\_gobierno\\_dato\\_esp/unidad1\\_pdf1.pdf](https://www.centro-virtual.com/recursos/biblioteca/pdf/herramientas_big_data_gobierno_dato_esp/unidad1_pdf1.pdf).
- [19] C. U. Asturias, *Almacenamiento y Procesamiento de la Información*. 2020, pág. 8. dirección: [https://www.centro-virtual.com/recursos/biblioteca/pdf/herramientas\\_big\\_data\\_gobierno\\_dato\\_esp/unidad1\\_pdf1.pdf](https://www.centro-virtual.com/recursos/biblioteca/pdf/herramientas_big_data_gobierno_dato_esp/unidad1_pdf1.pdf).
- [20] C. U. Asturias, *Almacenamiento y Procesamiento de la Información*. 2020, pág. 9. dirección: [https://www.centro-virtual.com/recursos/biblioteca/pdf/herramientas\\_big\\_data\\_gobierno\\_dato\\_esp/unidad1\\_pdf1.pdf](https://www.centro-virtual.com/recursos/biblioteca/pdf/herramientas_big_data_gobierno_dato_esp/unidad1_pdf1.pdf).
- [21] C. U. Asturias, *Almacenamiento y Procesamiento de la Información*. 2020, pág. 10. dirección: [https://www.centro-virtual.com/recursos/biblioteca/pdf/herramientas\\_big\\_data\\_gobierno\\_dato\\_esp/unidad1\\_pdf1.pdf](https://www.centro-virtual.com/recursos/biblioteca/pdf/herramientas_big_data_gobierno_dato_esp/unidad1_pdf1.pdf).
- [22] C. U. Asturias, *Almacenamiento y Procesamiento de la Información*. 2020, pág. 11. dirección: [https://www.centro-virtual.com/recursos/biblioteca/pdf/herramientas\\_big\\_data\\_gobierno\\_dato\\_esp/unidad1\\_pdf1.pdf](https://www.centro-virtual.com/recursos/biblioteca/pdf/herramientas_big_data_gobierno_dato_esp/unidad1_pdf1.pdf).
- [23] C. U. Asturias, *Almacenamiento y Procesamiento de la Información*. 2020, pág. 12. dirección: [https://www.centro-virtual.com/recursos/biblioteca/pdf/herramientas\\_big\\_data\\_gobierno\\_dato\\_esp/unidad1\\_pdf1.pdf](https://www.centro-virtual.com/recursos/biblioteca/pdf/herramientas_big_data_gobierno_dato_esp/unidad1_pdf1.pdf).
- [24] C. Thornton, M. Sharat, T. Robert, L. Michele y S. Nina, *Magic Quadrant for Data Integration Tools*. 2024. dirección: <https://www.gartner.com/doc/reprints?id=1-2JARDD5F&ct=241107&st=sb>.
- [25] W.H.Inmon, *Building the Data Warehouse, 3a ed*, Wiley Publishing. Wiley, 2002, pág. 288, ISBN: 978-1-118-53080-1.