

Creación de un almacén de datos para la empresa CITMATEL

**Trabajo de práctica profesional de 4er año de la carrera de ingeniería informática**

**Autor/es:**

Marcos Alejandro Fernández Piñero

**Tutores:**

Dra. C. Lisandra Bravo Ilisástigui

La Habana, Cuba

Diciembre, 2024

**Índice**

[**Resumen** 3](#_Toc183182407)

[**Abstract** 4](#_Toc183182408)

[**Introducción** 5](#_Toc183182409)

[**Desarrollo** 9](#_Toc183182410)

[**Capítulo 1: Análisis del marco teórico del problema** 9](#_Toc183182411)

[**1.1 Antecedentes del Estudio** 9](#_Toc183182412)

[**1.2 Análisis de la teoría asociada** 11](#_Toc183182413)

[**1.3 Opciones del mercado** 15](#_Toc183182414)

[**Capítulo 2: Análisis del negocio y Diseño de la solución** 18](#_Toc183182415)

[**2.1 Análisis y modelado del negocio** 18](#_Toc183182416)

[**2.2 Diseño de la solución** 23](#_Toc183182417)

[**Capítulo 3: Diseño y ejecución de pruebas** 24](#_Toc183182418)

[**Conclusiones** 25](#_Toc183182419)

[**Referencias bibliográficas** 26](#_Toc183182420)

# **Resumen**

# **Abstract**

# **Introducción**

Según (Calle García, Alvarado Choez, Anzules Cruz, & Menéndez Parrales, 2024), en la era digital actual, las empresas enfrentan un crecimiento exponencial en el volumen de datos que provienen de diversas fuentes. Este fenómeno es impulsado por la proliferación de dispositivos conectados, plataformas digitales y tecnologías emergentes que generan grandes cantidades de información en tiempo real. Según un estudio, herramientas tecnológicas avanzadas han incrementado la eficiencia y precisión en la recopilación de datos en un 85%, lo que ha permitido a empresas como IBM y FedEx mejorar significativamente sus procesos. Sin embargo, este aumento en la cantidad de datos también presenta desafíos significativos para las organizaciones.

Este fenómeno del flujo constante y variado de información, como refiere (Biroğul & Gültekin, 2016), ha obligado a las empresas a adoptar nuevas estrategias para su gestión y análisis. La diversidad en los formatos y las fuentes de estos datos complica su integración dentro de los sistemas empresariales existentes y para la toma de decisiones asociada al análisis de estos datos. Seguidamente en (Kumar, 2023) se plantea que el manejo inadecuado o la falta de integración eficiente pueden llevar a la creación de silos de datos, donde la información queda aislada dentro de diferentes departamentos o sistemas.

Todo esto desembocado en que la capacidad para tomar decisiones a partir de estos datos se vea afectada. En primer lugar, la calidad y consistencia de los datos son esenciales para obtener resultados precisos. Sin embargo, muchas organizaciones enfrentan problemas con la calidad de los datos debido a su dispersión y falta de estandarización (Robles Angeles, Robles Angeles, Ángeles Hernández, & Acosta Miranda, 2023). Además, las herramientas tradicionales a menudo no están equipadas para manejar el volumen y velocidad actuales del flujo de datos, lo que limita su utilidad en la toma de decisiones estratégicas(Dmitriyev, Mahmoud, & Marín-Ortega, 2022).

Para superar estos desafíos, las empresas están adoptando soluciones que centralizan la gestión y análisis de datos. La inteligencia de negocios (BI) se presenta como una solución efectiva al integrar todos los datos distribuidos dentro de un único almacén centralizado. Esta centralización permite una visión más completa del desempeño empresarial y facilita la toma rápida y precisa de decisiones(Biroğul & Gültekin, 2016). Finalmente, según (Robles Angeles et al., 2023), el uso del proceso ETL (extracción, transformación y carga) es crucial para consolidar los datos provenientes de diversas fuentes en un repositorio centralizado, mejorando así su integridad y accesibilidad.

En el contexto cubano actual, empresas como CITMATEL se han visto inmersas en esta situación. CITMATEL, una empresa cubana con presencia en el ámbito digital, enfrenta al desafiante obstáculo de la dispersión de sus datos en múltiples bases de datos correspondientes a sus diversas tiendas digitales. Esta fragmentación dificulta la realización de análisis complejos necesarios para procesos críticos como el márquetin, las finanzas y las estadísticas, se hace necesario en acceso frecuente y variante a estos datos.

La falta de un sistema centralizado y los mecanismos adecuados para el acceso, gestión y visualización de los datos impide a la empresa aprovechar plenamente su potencial de datos, lo que limita su capacidad para responder ágilmente a las demandas del mercado y optimizar sus operaciones internas. Inicialmente la empresa intento acceder a tecnologías de análisis de datos como:

* Zoho Analitycs, disponible en <https://to.softdoit.com/syjawq2h>
* Qlik, disponible en <https://to.softdoit.com/w5ctu7>
* Tableau, disponible en https://to.softdoit.com/shxo15
* Power BI, disponible en <https://pxl.to/knnpzro>

Durante las pruebas no se pudieron acceder a estas herramientas porque las tres últimas están bloqueadas para cuba y la primera es de pago.

Actualmente, Los departamentos solicitan al equipo de desarrollo los datos a través de correo electrónico, los cuales deben hacer consultas SQL directamente en las bases de datos en producción para obtener la información requerida, exigiendo la inversión de tiempo y esfuerzo en tareas que no deberían ser su responsabilidad y limitando la capacidad de gestión y procesamiento de la empresa.

Con este proyecto se busca diseñar e implementar una solución que permita centralizar los datos en un único repositorio accesible, facilitando así el acceso rápido y seguro a información confiable en tiempo real por personal no relacionado a la informática. Esta centralización no solo mejorará la integridad y seguridad de los datos, sino que también fomentará una mayor colaboración entre los departamentos al proporcionar una única fuente de verdad.

Con esta iniciativa, se espera transformar la manera en que la Entidad gestiona sus datos, permitiendo una toma de decisiones más ágil y fundamentada. El proyecto incluirá tanto el diseño e implementación técnica como la documentación exhaustiva del proceso, asegurando que todos los aspectos del sistema sean comprensibles y replicables en el futuro.

De ahí que se establezca la siguiente **situación problemática**: las consultas a los datos se deben hacer de forma manual directamente a las bases de datos debido a la falta de un repositorio de datos que permita la recopilación y manipulación de los datos, consumiendo demasiado tiempo y esfuerzo de parte del personal de desarrollo de software que es el único capacitado para llevar a cabo esta tarea. Esto limita a los equipos consumidores de estos datos de poder hacer consultas más complejas o especializadas. Adicionalmente las soluciones disponibles en el mercado o son de pago o no están disponibles para Cuba.

Esto plantea una contradicción entre los consumidores de estos datos que requieren un sistema capaz de proporcionárselos de forma clara y directa para su análisis y las estructuras de almacenaje y gestión de los mismos. Planteando la siguiente **pregunta científica**: Cómo se pueden centralizar todas las fuentes de datos en un mismo lugar que permita el acceso a los datos de forma óptima en función de las necesidades de los consumidores y permitirles a estos acceder a los datos de forma clara y entendible para ellos.

Para poder resolver este problema se ha establecido como **objetivo de la investigación**: Implementar un sistema de almacenamiento y gestión de todos los datos de la empresa que satisfaga la demanda de estos datos por parte de los diferentes departamentos de la empresa.

Para la culminación de estas prácticas laborales se acoto la solución hasta la implementación de la Base de Datos centralizada y el sistema de ETL.

Para esto se plantearon como objetivos específicos:

1. Revisión y análisis del estado del arte y la teoría asociada al tema.
2. Comprensión y documentación del negocio.
3. Implementación de la solución al problema identificado
4. Revisión de la factibilidad solución para el problema.

Para el cumplimiento de estos objetivos se proponen las siguientes tareas:

1. Documentación de las bases de datos existentes para la comprensión de las fuentes de datos.
2. Identificación y acercamiento a los *StakeHolders* para la obtención de requisitos del sistema
3. Revisión de estudio e investigaciones relacionadas con el tema para identificar conceptos y palabras claves.
4. Profundización de esos conceptos claves y la identificación de tecnologías y estrategias.
5. Identificación de las reglas del negocio y los requisitos del negocio y del sistema.
6. Modelado del negocio y del sistema.
7. Diseño de una solución general que se adapte a las necesidades específicas de la empresa.
8. Desnormalización de las bases de datos existentes y la renormalización de los datos en función de la demanda de los departamentos que requieren los datos.
9. Implementación de la solución propuesta.
10. Implementación de pruebas que validen la efectividad del sistema propuesto.
11. Validación de la capacidad del sistema para adaptase a las necesidades actuales y futuras de la empresa.
12. Despliegue final de la solución.

# **Desarrollo**

## **Capítulo 1: Análisis del marco teórico del problema**

El presente capítulo establece el marco teórico esencial para el desarrollo de un Sistema de Soporte a la Toma de decisiones (DSS en inglés) donde se unifiquen y pueden gestionar todos los datos de la entidad **CITMATEL** que gestiona cuatro tiendas en línea, cada una con una base de datos estructuralmente distinta. Este marco teórico proporciona la base conceptual y analítica necesaria para abordar los desafíos inherentes a la integración de múltiples sistemas de bases de datos.

Inicialmente, se presentarán los antecedentes del estudio, donde se revisarán investigaciones previas y estudios relacionados que han abordado problemáticas similares. Se identificarán las brechas existentes en el conocimiento actual que este proyecto busca llenar, destacando la necesidad de soluciones innovadoras para la integración de datos heterogéneos.

A continuación, se procederá con la definición de conceptos clave, donde se clarificarán los términos y conceptos fundamentales que serán recurrentes a lo largo del proyecto. Estas definiciones estarán respaldadas por fuentes académicas pertinentes, asegurando un entendimiento común y preciso.

Posteriormente, se expondrán las teorías relevantes que sustentan el trabajo. Se explicará cómo estas teorías se aplican específicamente al contexto del proyecto, proporcionando un marco conceptual robusto para guiar el desarrollo e implementación del sistema.

El estado del arte ofrecerá una visión general de los desarrollos actuales en el ámbito de la Inteligencia de Negocios y los Repositorios de Datos. Se incluirán innovaciones recientes y tendencias futuras que podrían influir en el diseño y ejecución del proyecto, asegurando que las soluciones propuestas sean contemporáneas y efectivas.

### **1.1 Antecedentes del Estudio**

Para realizar un análisis del estado del arte se precisa primero realizar un levantamiento exhaustivo de investigaciones previas sobre problemas similares y las soluciones aplicadas, es esencial considerar diversos estudios y enfoques que han abordado cuestiones análogas en el campo de interés. Para ello, se realizó una búsqueda en bases de datos académicas sobre términos como “*Business intelligence*”, “Repositorios de datos”, “Almacenes de datos” y “Sistemas de Apoyo a la toma de Decisiones”. De esta investigación se obtuvo una serie de documentos que serán analizados en esta sección.

En el contexto actual, donde las empresas enfrentan un aumento exponencial en el volumen y variedad de datos, se han identificado varios enfoques y soluciones que abordan problemas similares al que enfrenta CITMATEL.

El uso de Data Warehouses (DW) se ha consolidado como una herramienta esencial para integrar y analizar grandes volúmenes de datos provenientes de diversas fuentes para sistemas de soporte a la toma de decisiones (DSS). Estos sistemas permiten transformar datos brutos en información útil y accesible, facilitando así el análisis por parte de los usuarios finales sin necesidad de conocimientos técnicos avanzados(Vite, Villegas, & Pizarro, 2018). La metodología Hefesto, por ejemplo, ha sido ampliamente utilizada para guiar el diseño e implementación de DW, destacándose por su enfoque en la recolección de requerimientos y procesos ETL (Extracción, Transformación y Carga) eficientes (Romero-Chuquital & Melendres-Velasco, 2023).

En investigaciones previas, se ha demostrado que la implementación de un DW puede mejorar significativamente la eficiencia operativa y la capacidad de respuesta empresarial. Por ejemplo, en el ámbito bibliotecario, se ha utilizado un sistema de soporte a las decisiones basado en DW para integrar datos heterogéneos y mejorar la gestión bibliotecaria (Romero-Chuquital & Melendres-Velasco, 2023). Este enfoque ha permitido a los gestores generar informes personalizados y optimizar procesos transaccionales diarios.

Asimismo, y según (Valle, Rocano, Siguenza-Guzman, Cattrysse, & Galarza, 2014), en sistemas de gestión de proyectos, los repositorios de datos han facilitado la investigación al proporcionar una base estructurada para el análisis y la toma de decisiones. Estos repositorios permiten compartir y reutilizar datos, lo cual es crucial para el descubrimiento de nuevo conocimiento dentro de las organizaciones. En particular, se han desarrollado sistemas que utilizan técnicas avanzadas como la minería de datos para asegurar ingresos y evaluar proyectos bajo condiciones inciertas.

Por otro lado, el concepto de data lakes ha sido mencionado como una solución prometedora para manejar grandes volúmenes de datos estructurados y no estructurados por otros estudios como una alternativa más flexible. Este enfoque permite a las empresas almacenar datos en su formato nativo, ofreciendo flexibilidad para el análisis y experimentación por parte de científicos de datos y analistas(Llave, 2018). Los data lakes pueden servir como áreas de preparación para los DW, facilitando la integración de diversas fuentes de datos sin necesidad de transformaciones iniciales. Sin embargo, según (Sulova, 2019) presentan desafíos como la gobernanza de datos y la calidad de los mismos, aspectos que requieren atención cuidadosa durante su implementación.

De esta exhaustiva se definen como conceptos de interés temas como “Data WareHouse”, “Data lake”, “Data Marks”, “DSS”, “Business intelligence”, “ETL”, “Silos de datos” y “Data Hubs”.

### **1.2 Análisis de la teoría asociada**

La inteligencia de negocios (BI, por sus siglas en inglés) se define como un marco integral que abarca un conjunto de herramientas y técnicas destinadas a transformar datos brutos en información valiosa. Esta información es crucial para la toma de decisiones estratégicas dentro de una organización (Zhou, San, & Liu, 2023). A medida que las empresas generan cantidades masivas de datos, la inteligencia de negocios se ha convertido en una infraestructura crítica que permite a las organizaciones reaccionar de manera efectiva a las operaciones tanto operativas como estratégicas de sus clientes, mercados y otros grupos de interés (Arturo & Fernández, 2012).

Este proceso implica la extracción, análisis y presentación de datos mediante sistemas de información más entendibles por los clientes. Estos sistemas permiten a los gerentes obtener un conocimiento profundo sobre el comportamiento histórico del negocio y su entorno, lo que facilita decisiones informadas que mejoran el rendimiento organizacional y optimizan la utilidad económica mediante una reducción significativa de costos (Juárez & Francisco, 2019).

Como ya se mencionó, los sistemas de BI deben cumplir con la característica de permitir a los clientes entender estos datos y tomar decisiones en base a ellos, por lo que surge la necesidad de la creación de sistemas de apoyo a la toma de decisiones(DSS).

Los Sistemas de Apoyo a la Toma de Decisiones (DSS, por sus siglas en inglés) son aplicaciones prácticas de la inteligencia de negocios, diseñadas para asistir a los gestores en la toma de decisiones informadas basadas en datos analíticos. Estos sistemas integran diversas fuentes de información y aplican modelos analíticos avanzados para proporcionar recomendaciones y pronósticos que facilitan el proceso decisional en las organizaciones (Ordoñez Cuthbert & Sambola, 2023).

Los DSS se caracterizan por su capacidad para manejar grandes volúmenes de datos y transformarlos en información útil, permitiendo a los gestores evaluar diferentes escenarios y tomar decisiones estratégicas más acertadas. Al utilizar herramientas de análisis predictivo y modelos de simulación, los DSS ayudan a identificar tendencias, evaluar riesgos y oportunidades, y optimizar los recursos empresariales (Vásquez, Martinez, Aguilera, & Gil, 2018).

Según refiere (Torres & Velastegui, 2020) para logran una mayor eficiencia en el tratamiento de fuentes muy dispersas y heterogéneas, se hace necesario el empleo de herramientas ETL para la correcta recuperación y preparación de estos datos.

Los procesos de Extracción, Transformación y Carga (ETL) son fundamentales en la preparación de datos para su análisis en sistemas de inteligencia de negocios (BI) y sistemas de apoyo a la toma de decisiones (DSS). Estos procesos permiten integrar datos desde múltiples fuentes hacia un formato coherente y utilizable, lo que es esencial para convertir datos brutos en información valiosa.

Según las explicaciones de (Ratnaparkhi, 2019), estos sistemas, como su nombre lo indica, constan de 3 partes: Extracción, Transformación y Carga. La primera implica la recopilación de datos desde diversas fuentes, que pueden incluir bases de datos, archivos planos, servicios web, entre otros. Esta fase es crítica porque asegura que todos los datos relevantes sean capturados para el análisis posterior. Por otro lado, la Transformación es el proceso donde los datos extraídos se limpian y se convierten en un formato adecuado para el análisis. Esto puede incluir tareas como la depuración de datos duplicados, la normalización de formatos y la aplicación de reglas de negocio específicas. La transformación es vital para asegurar que los datos sean precisos y consistentes. Por último, la Carga es la fase final donde los datos transformados se almacenan en un sistema de destino, como un data warehouse o un data mart. La eficiencia en esta etapa garantiza que los datos estén disponibles para su análisis en tiempo real o casi real.

Para lograr la realización de un ETL es necesario entender la procedencia de los datos y su estructura en estos lugares. Usualmente, y según refiere la fuente (Hubspot, 2021), estos datos provienen de silos, o sea almacenes de datos que registran datos de un tema o departamento exclusivamente y que otros no necesitan, dejando la información inaccesible para otros grupos. En ocasiones la ocurrencia de silos es un fenómeno que se desea evitar, principalmente cuando los silos requieren ser consultados por varios departamentos.

Como vimos en los sistemas ETL, para poder llevarlo a cabo el proceso de Carga, es fundamental contar con un lugar bien estructurado para el almacenamiento de estos datos. Durante la anterior investigación inicial detectaron 4 “tipos de almacenes”: Data Hubs, Data Warehauseses, Data lakes y Dama Marts.

Un **Data Warehouse** es un sistema centralizado diseñado para almacenar grandes volúmenes de datos estructurados provenientes de diversas fuentes. Su principal objetivo es facilitar el análisis y la generación de informes. Este organiza los datos en esquemas predefinidos, como el esquema estrella o copo de nieve, lo que permite realizar consultas complejas de manera eficiente. Este tipo de almacenamiento es ideal para el análisis histórico y la toma de decisiones estratégicas, ya que ofrece una visión consolidada y limpia de los datos (Aissi et al., 2021).

Por su parte un **Data Mart** es una versión más pequeña y especializada de un data warehouse. Está diseñado para atender las necesidades específicas de un departamento o área dentro de una organización, como ventas o finanzas. Los data marts suelen contener subconjuntos de datos relevantes para ese departamento, permitiendo un acceso más rápido y eficiente a la información necesaria para análisis específicos. Esto reduce la complejidad y el tiempo de respuesta en comparación con un data warehouse completo (Jaleel & Abbas, 2020).

Por su parte, Un **Data Lake** es un repositorio de almacenamiento que permite guardar grandes cantidades de datos en su formato nativo, ya sean estructurados, semiestructurados o no estructurados. A diferencia del data warehouse, los data lakes no requieren una estructura predefinida para los datos, lo que los hace más flexibles y escalables. Son ideales para el análisis exploratorio y el procesamiento de big data, ya que permiten almacenar datos brutos que pueden ser transformados y analizados posteriormente según las necesidades del negocio (Aissi et al., 2021)

Finalmente, **un Data Hub** se define como una plataforma que facilita la integración de datos en tiempo real para aplicaciones colaborativas. Aunque no existen estándares fijos para su definición, generalmente se caracteriza por su capacidad para manejar datos estructurados, ofreciendo una baja flexibilidad, pero alta accesibilidad. Este tipo de sistema permite a las organizaciones integrar eficientemente información de diversas fuentes, mejorando la colaboración y el intercambio de datos entre diferentes sistemas y departamentos(Amid et al., 2019).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Característica | Data Hub | Data Warehouse | Data Mart | Data Lake |
| Formato de Datos | Estructurado | Estructurado | Estructurado | Estructurado, semiestructurado y no estructurado |
| Escalabilidad | Limitada | Limitada | Limitada | Alta |
| Flexibilidad | Baja | Baja | Media | Alta |
| Objetivo Principal | Integración de datos en tiempo real para aplicaciones colaborativas. | Análisis histórico y generación de informes | Análisis específico por departamento | Almacenamiento y análisis exploratorio de big data. |

Gracias a esta tabla que resume las características de cada uno, y teniendo en cuenta las características del negocio, resulta más conveniente la implementación de un Data Lake sobre los otros sistemas de almacenamiento, pues permite el manejo de datos más heterogéneos y permite mantener estos datos en su forma original sin necesidad de transformarlos.

### **1.3 Opciones del mercado**

Realizando una exhaustiva revisión en internet se identificaron varias tecnologías que son ampliamente empleadas en el mundo del BI como SAS Visual Analytics, Zoho Analytics y Google Data Studio. Adicionalmente, revisando estudios y trabajos publicados también se encontraron las tecnologías Power BI, Tableau, IBM Cognos Analytics, Qlik Sense (Guerrero Garcia & Rodas-Silva, 2022),(Ayala, Ortíz, Guevara, & Maya, 2018).

**Power BI** es una plataforma de análisis de datos desarrollada por Microsoft que permite a los usuarios conectar, modelar y visualizar datos desde diversas fuentes. Ofrece una versión en la nube llamada Power BI Service, que facilita la creación de dashboards interactivos y la colaboración en tiempo real. Su integración con otros productos de Microsoft, como Excel y Azure, lo convierte en una opción atractiva para empresas que ya utilizan el ecosistema de Microsoft. Además, Power BI incluye capacidades avanzadas como el uso de inteligencia artificial para identificar patrones en los datos. Es ampliamente utilizado por su facilidad de uso y su capacidad para manejar grandes volúmenes de información (Microsoft, 2015).

**Tableau** es una herramienta líder en visualización de datos que permite a los usuarios crear gráficos y dashboards interactivos sin necesidad de conocimientos avanzados en programación. Tableau se destaca por su capacidad para manejar grandes conjuntos de datos y su facilidad para integrarse con múltiples fuentes, incluidas bases de datos SQL y servicios en la nube. La versión Tableau Online permite a los usuarios acceder a sus informes desde cualquier lugar con conexión a internet. Tableau es conocido por su enfoque en el diseño visual, lo que permite crear informes altamente personalizables y estéticamente atractivos. Además, ofrece funciones avanzadas como análisis predictivo y soporte para big data (Software, 2019).

**IBM Cognos Analytics** es una solución integral de Business Intelligence que combina análisis avanzado con capacidades robustas de generación de informes. La plataforma permite a los usuarios crear dashboards interactivos y realizar análisis predictivos utilizando técnicas como el machine learning. Cognos Analytics ofrece una versión en la nube que facilita la colaboración entre equipos distribuidos geográficamente. Su arquitectura escalable lo convierte en una opción ideal para grandes organizaciones que requieren un alto grado de personalización y seguridad. Además, se integra fácilmente con otras soluciones empresariales, lo que permite un flujo continuo de datos entre diferentes sistemas (Ibm, 2024).

**Qlik Sense** es una herramienta diseñada para permitir a los usuarios explorar sus datos sin restricciones mediante un motor asociativo único. Este motor permite descubrir relaciones ocultas entre los datos que no son evidentes con las herramientas tradicionales basadas en consultas SQL. Qlik Sense ofrece una versión en la nube llamada Qlik Sense Cloud, que facilita la colaboración entre equipos al permitir el acceso remoto a los dashboards. La plataforma también incluye capacidades avanzadas como análisis predictivo e integración con big data. Su enfoque centrado en el usuario final lo hace adecuado tanto para analistas técnicos como para usuarios empresariales (Qlik, 2022).

**SAS Visual Analytics** es una plataforma avanzada que combina visualización de datos con capacidades analíticas robustas, incluyendo análisis predictivo y minería de datos. Esta herramienta está diseñada para manejar grandes volúmenes de información y realizar análisis complejos utilizando técnicas estadísticas avanzadas. SAS Visual Analytics ofrece opciones tanto locales como basadas en la nube, permitiendo a las empresas elegir la infraestructura que mejor se adapte a sus necesidades. La plataforma también incluye herramientas colaborativas que facilitan el intercambio de insights entre equipos multidisciplinarios. Es ampliamente utilizada en sectores como finanzas, salud y manufactura debido a su capacidad para procesar datos complejos (Statistical Analysis, 2023).

**Zoho Analytics** es una solución asequible y fácil de usar diseñada principalmente para pequeñas y medianas empresas (PYMEs). Ofrece tanto versiones locales como basadas en la nube, permitiendo a las empresas elegir cómo desean gestionar sus datos. Zoho Analytics destaca por su interfaz intuitiva que facilita la creación rápida de informes interactivos sin necesidad de conocimientos técnicos avanzados. La plataforma también incluye capacidades básicas de análisis predictivo y soporte para big data, lo que permite a las PYMEs obtener insights valiosos sin incurrir en altos costos (Zoho, 2019).

**Google Data Studio** es una herramienta gratuita basada completamente en la web que permite crear informes interactivos conectando diversas fuentes de datos, como Google Analytics, Google Sheets o bases de datos SQL. Data Studio se destaca por su simplicidad y facilidad para compartir dashboards interactivos con otros usuarios mediante enlaces públicos o privados. Aunque no ofrece algunas funcionalidades avanzadas presentes en otras plataformas comerciales, su integración nativa con productos Google lo convierte en una opción atractiva para empresas pequeñas o equipos internos que ya utilizan el ecosistema Google Workspace (Nedelcu, 2024).

Estas herramientas al ser probadas presentaron en general 3 problemas críticos. El primero es que eran software propietario, lo cual limita la capacidad de escalar en funcionalidades especificas del cliente, además son de pago o tienen versiones gratuitas muy limitadas y algunas estaban bloqueadas para Cuba. Esto llevo a la necesidad del desarrollo de un nuevo sistema capaz de adaptarse a las necesidades actuales y futuras que sea propio de la entidad y de pueda escalar y personalizar.

## **Capítulo 2: Análisis del negocio y Diseño de la solución**

### **2.1 Análisis y modelado del negocio**

En el proceso de captura y modelado de requerimientos para un sistema de ingeniería de software, se distinguen varias fases indispensables. La primera es la captura de requisitos y reglas del negocio, donde se recopilan las necesidades funcionales y no funcionales del sistema, así como las reglas que rigen los procesos del negocio. Este proceso implica identificar las expectativas y restricciones que se tienen sobre el sistema, asegurando que los requisitos sean claros, completos y verificables. Además, se documentan las reglas del negocio, que son fundamentales para garantizar que el sistema se alinee con los objetivos estratégicos de la organización. Estas reglas definen cómo se deben ejecutar los procesos dentro del sistema y establecen las condiciones bajo las cuales se toman decisiones automatizadas o manuales (Sumerville, referencia faltante).

Posteriormente, se procede a la fase de modelado del negocio y modelado del sistema, donde se construyen representaciones abstractas tanto del entorno empresarial como del sistema en desarrollo. El modelado del negocio permite visualizar y analizar los procesos actuales, identificando oportunidades de mejora y asegurando que el diseño del sistema refleje fielmente la realidad operativa. En paralelo, el modelado del sistema traduce los requisitos capturados en una representación técnica mediante diagramas permitiendo a los desarrolladores entender cómo deben implementarse las funcionalidades dentro del software (Sumerville, referencia faltante).

#### **2.1.1 Requisitos y reglas del negocio**

Para una correcta captura de los requisitos lo primero es la correcta identificación de los stakeholders. Los stakeholders son todas aquellas personas o entidades que tienen un interés directo o indirecto en el desarrollo del sistema, ya sea porque lo utilizarán, lo afectarán o se verán impactados por su implementación. En este caso, se identificaron cuatro grupos principales de stakeholders:

* Elena (Lider del Equipo de Desarrollo)
* Equipo de Desarrollo de software
* Equipo de Marketing y Publicidad
* Osvaldo (Analista de datos)

Luego es una correcta identificación y aplicación de técnicas de captura de la información. Estas son métodos estructurados que permiten recolectar datos relevantes sobre los requisitos y las reglas del negocio de un sistema. Estas técnicas facilitan la interacción entre los analistas y los stakeholders, asegurando que se obtenga una visión clara y precisa de las necesidades del proyecto. En esta investigación, se emplearon tres técnicas principales: entrevistas, sesiones de trabajo conjunto y cuestionarios. Las entrevistas consistieron en reuniones individuales con los stakeholders clave, donde se profundizó en sus expectativas y requerimientos específicos. Las sesiones de trabajo conjunto permitieron a varios actores involucrados colaborar simultáneamente para identificar problemas y soluciones en tiempo real, fomentando el consenso sobre los requisitos. Finalmente, los cuestionarios se utilizaron para recopilar información estructurada de un grupo más amplio de participantes, permitiendo obtener datos cuantitativos y cualitativos que complementaron las otras técnicas(Pressman, referencia faltante).

De estas técnicas se obtuvieron las siguientes reglas de negocio:

* + El sistema maneja información de pedidos, comprador, receptor, pagos, transportistas y proveedores.
  + Los departamentos de Marketing, Ventas y Estadísticas requieren para sus operaciones el consumo de datos alojados en las bases de datos.
  + Los datos requeridos por estos departamentos son heterogéneos, lo que implica que la combinación, tiempo o rango de los datos en un pedido no se repite con frecuencia.
  + El acceso a los datos debe estar restringido al personal autorizado

Una vez identificadas las reglas de negocios se procede a la identificación de los requisitos funcionales y no funcionales de la misma.

**Requerimientos funcionales**

* El sistema debe permitir la recuperación de datos procedentes de las bases de datos de las plataformas de ventas.
* El sistema debe permitir la actualización diaria de los datos.
* Si un dato que ya se encontraba en el repositorio fue modificado en la fuente, el sistema debe ser capaz de reflejar esa actualización
* El sistema debe ser capaz de eliminar información que sobrepaso los 5 años de antigüedad
* Si un dato que ya fue eliminado del repositorio es solicitado, el sistema debe ser capaz de recuperarlo
* El sistema debe ser capaz de obtener toda la información referente a: pedidos, comprador, receptor, pagos, transportistas y proveedores.
* El sistema debe implementar lógica de autentificación para evitar el acceso a los datos de personal no autorizado.
* El sistema debe implementar mecanismos de seguridad que impidan el acceso externo a la empresa a los datos.
* El sistema debe permitir al usuario la creación de consultas dinámicas de los datos.
* El sistema debe permitir la visualización de los datos en formatos entendibles para los consumidores.
* El sistema debe permitir la creación de diferentes tipos de gráficos y tablas.
* El sistema debe ser capaz de realizar resúmenes estadísticos sobre los datos.
* El sistema debe de permitir la creación de reportes de los datos obtenidos en las consultas.
* El sistema debe permitir la exportación de estos reportes a formatos entendibles por herramientas de hojas de cálculos
* El sistema debe comunicarse en solo lectura con las bases de datos de las plataformas.

En cuanto a los **requerimientos no funcionales** solamente se solicitó el empleo de tecnología OpenSource.

#### **2.1.2 Casos de uso y actores**

Luego de haber identificado las requisitos y reglas del negocio podemos empezar a analizar cómo debería comportarse el negocio. Según la fuente Summerville una forma de lograr este objetivo es a través de la identificación de actores y casos de uso y la creación de un diagrama de casos de uso.

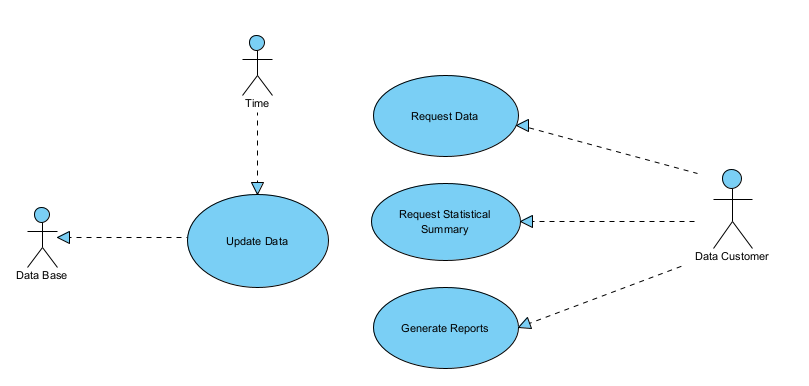
En este caso de identificaron dos actores principales: El tiempo y los consumidores de estos datos y como actor secundario se identificó a las fuentes de datos (bases de datos de las plataformas). En cuanto a los casos de uso se identificaron fundamentalmente 4: Solicitar datos, Solicitar resumen estadístico, Generar reportes y actualizar datos.

El caso de uso Solicitar datos inicia cuando un actor Consumidor de datos hace una consulta predefinida o diseñada al sistema solicitando un conjunto de datos. Si la consulta no está predefinida, el usuario tiene que crearla. Una vez lista la consulta el sistema recopila los datos solicitados. Luego el sistema prepara los datos para la visualización. El caso de uso concluye con la entrega de los datos solicitados.

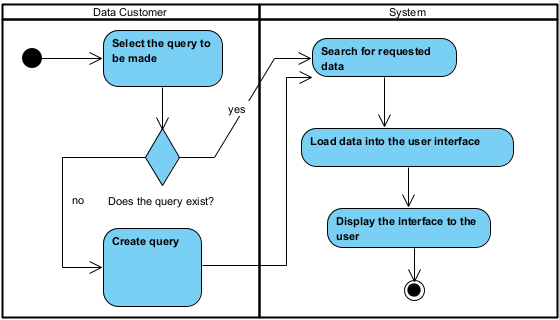
El segundo caso de uso, Solicitar Resumen Estadístico, inicia cuando el actor Consumidor de datos Solicita un resumen de diferentes métricas estadísticas predefinidas. El sistema realiza las operaciones estadísticas y las prepara para ser mostradas. El caso de uso culmina con la entrega del resumen al Consumidor de Datos.

El tercer caso de uso, Generar Reportes, inicia cuando el actor data Customer solicita la creación de un reporte al sistema. Este genera un archivo de hoja de cálculo con la extensión xlxs e introduce los datos en la misma. El caso de uso concluye con el aviso de la conclusión de la generación del reporte.

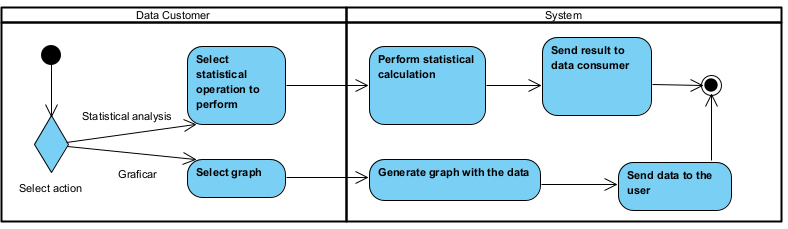
El último caso de uso, Actualizar Datos, inicia cuando ha pasado un día desde su última actualización. El sistema solicita a las bases de datos de las plataformas los datos alojados en ella, luego el sistema supervisa si se ha modificado algún campo de las filas viejas o se ha agregado alguna fila. Si algún dato fue modificado, se actualiza en el sistema. El sistema finaliza con la revisión de todas las filas de las bases de datos.

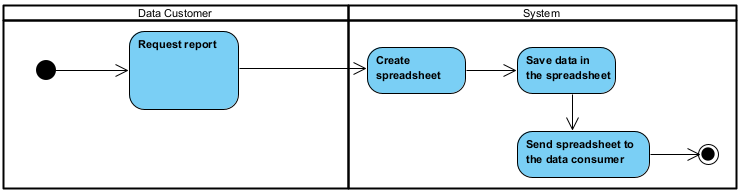
****

**Figura número 1: Diagrama de Casos de Uso**

****

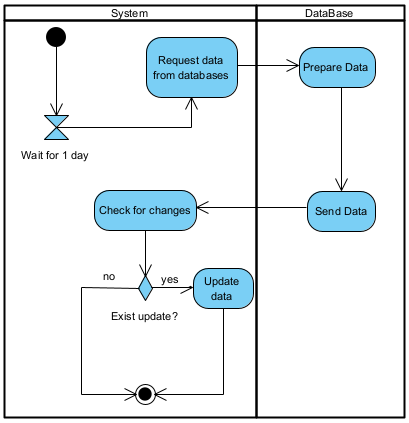
**Figura número 2: Diagrama de actividades del caso de uso Solicitar datos**

**Figura número 3: Diagrama de actividades del caso de uso Solicitar Resumen Estadístico**

****

**Figura número 3: Diagrama de actividades del caso de uso Generar Reportes**

**Figura número 3: Diagrama de actividades del caso de uso Generar Reportes**

****

### **2.2 Diseño de la solución**

## **Capítulo 3: Diseño y ejecución de pruebas**

# **Conclusiones**

# **Referencias bibliográficas**

Aissi, M. E. M. E., Benjelloun, S., Loukili, Y., Lakhrissi, Y., Boushaki, A. E., Chougrad, H., & Ali, S. E. B. (2021). Data Lake Versus Data Warehouse Architecture: A Comparative Study. *Lecture Notes in Electrical Engineering*.

Amid, C., Pakseresht, N., Silvester, N., Jayathilaka, S., Lund, O., Dynovski, L. D., . . . Alako, B. T. (2019). The COMPARE data hubs. *Database, 2019*, baz136.

Arturo, H., & Fernández, F. (2012). *Inteligencia de negocios como apoyo a la toma de decisiones en la gerencia Business Intelligence as Support of Decisions Making in Management*.

Ayala, J. L., Ortíz, J., Guevara, C. P., & Maya, E. (2018). *Herramientas de Business Intelligence (BI) modernas, basadas en memoria y con lógica asociativa*.

Biroğul, S., & Gültekin, H. (2016). Importance of Business Intelligence Solution on Decision-Making Process of Companies. *International Journal of Applied Mathematics, Electronics and Computers, 4*, 86-89.

Calle García, A. J., Alvarado Choez, K. J., Anzules Cruz, J. J., & Menéndez Parrales, A. G. (2024). NUEVAS HERRAMIENTAS TECNOLÓGICAS EN LA RECOPILACIÓN DE DATOS DE MERCADO. *Ciencia y Desarrollo*.

Dmitriyev, V., Mahmoud, T., & Marín-Ortega, P. M. (2022). SOA enabled ELTA: approach in designing business intelligence solutions in Era of Big Data. *International Journal of Information Systems and Project Management*.

Guerrero Garcia, M. J., & Rodas-Silva, J. (2022). Análisis comparativo de metodologías y herramientas tecnológicas para procesos de Business Intelligence orientado a la toma de decisiones. *Informática y Sistemas: Revista de Tecnologías de la Informática y las Comunicaciones*.

Hubspot. (2021). Silos de datos: qué son, qué los causa y cómo solucionarlos. *Hubspot.es*. Retrieved from <https://blog.hubspot.es/marketing/silos-de-datos>

Ibm. (2024). IBM Cognos Analytics. [*www.ibm.com*](www.ibm.com). Retrieved from <https://www.ibm.com/es-es/products/cognos-analytics>

Jaleel, R. A., & Abbas, T. M. J. (2020). Design and Implementation of Efficient Decision Support System Using Data Mart Architecture. *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, 1-6.

Juárez, A., & Francisco, J. A. (2019). *La inteligencia de negocios como una herramienta estratégica y de apoyo a la toma de decisiones en el área de ventas de la empresa Benites SRL*.

Kumar, S. (2023). Data Silos A Roadblock for AIOps. *ArXiv, abs/2312.10039*.

Llave, M. R. (2018). Data lakes in business intelligence: reporting from the trenches. *Procedia computer science, 138*, 516-524.

Microsoft. (2015). Power BI: visualización de datos | Microsoft Power Platform. [*www.microsoft.com*](www.microsoft.com). Retrieved from <https://www.microsoft.com/es-es/power-platform/products/power-bi>

Nedelcu, L. (2024). Exploring Google Analytics 4: Unveiling Insights through Data Analysis.

Ordoñez Cuthbert, D. K., & Sambola, D.-M. (2023). Herramienta basada en Inteligencia de Negocios y Analíticas para la toma de decisiones académicas. Caso de Bluefields Indian & Caribbean University. *Revista Científica de FAREM-Estelí*.

Qlik. (2022). Qlik | Analytics & Data Integration Platform. *Qlik*. Retrieved from <https://www.qlik.com>

Ratnaparkhi, R. S. (2019). The Extract Transform Load (ETL) Process and Optimization using Ab Initio. *International Journal for Research in Applied Science and Engineering Technology*.

Robles Angeles, L. J., Robles Angeles, M. B., Ángeles Hernández, L., & Acosta Miranda, M. L. (2023). El poder de los datos: Cómo ETL reformuló la gestión de datos. *Programación Matemática y Software*.

Romero-Chuquital, A., & Melendres-Velasco, J. J. (2023). Uso de data Warehouse para la toma de decisiones empresariales: una revisión literaria. *Revista Científica de Sistemas e Informática, 3*(2), e543-e543.

Software, T. (2019). Tableau Public. *Tableau Public*. Retrieved from <https://public.tableau.com>

Statistical Analysis, S. (2023). SAS: Analítica, Inteligencia Artificial y Gestión de Datos. [*www.sas.com*](www.sas.com). Retrieved from <https://www.sas.com/es_mx/home.html>

Sulova, S. (2019). *The Usage of Data Lake for Business Intelligence Data Analysis.* Paper presented at the International Conference Information and communication technologies in business and education.

Torres, L., & Velastegui, E. A. S. (2020). *Creación de una solución de inteligencia de negocios para la generación de reportes con información relevante del proceso de titulación de los estudiantes de la Carrera de Ingeniería de Sistemas de la Universidad Politécnica Salesiana Sede Quito*.

Valle, V. A. H., Rocano, W. R. P., Siguenza-Guzman, L., Cattrysse, D., & Galarza, V. H. S. (2014). *Diseño e Implementación de un Sistema de Soporte de Decisiones para el Centro de Documentación Regional “Juan Bautista Vázquez”*.

Vásquez, A. C., Martinez, G., Aguilera, F. J. E., & Gil, M. V. L. (2018). *EVOLUCIÓN DE LOS SISTEMAS DE SOPORTE A LA TOMA DE DECISIONES*.

Vite, V. J., Villegas, S. C., & Pizarro, G. (2018). Aplicación de un Spatial Data Warehouse en la gestión de proyectos de vinculación: Caso de Estudio. *Journal of Science and Research: Revista Ciencia e Investigación, 3*(9), 19-24.

Zhou, J., San, O. T., & Liu, Y. (2023). Design and Implementation of Enterprise Financial Decision Support System Based on Business Intelligence. *International Journal of Professional Business Review*.

Zoho. (2019). Data Analytics Platform for Organizations - Zoho Analytics. *Zoho*. Retrieved from <https://www.zoho.com/analytics/data-analytics-organization.html>