

Diseño y Arquitectura de un Data Lake para la Integración de Datos Empresariales  
en Sutex

Julieth Paola Pachón López  
Mauricio Mateus  
Edda Acened Cárdenas Cristancho  
Luis Carlos Quiñones Díaz

Directores:

Ernesto Cadena

UNIVERSIDAD SANTO TOMÁS  
FACULTAD DE INGENIERÍA DE TELECOMUNICACIONES  
ESPECIALIZACIÓN EN GESTIÓN DE SERVICIOS DE TECNOLOGÍAS DE  
LA INFORMACIÓN  
BOGOTÁ  
2024

## Tabla de contenido

1.1. ÁRBOL DE PROBLEMAS.....	4
1.2. ¿QUÉ SE QUIERE SOLUCIONAR?.....	4
2.1 POR QUÉ SE PLANTEA AHORA LA SOLUCIÓN.....	6
2.2 SECTOR OBJETIVO.....	10
2.3 TENDENCIAS DEL SECTOR.....	12
2.4 ANÁLISIS DE MERCADO.....	13
2.5 ÁRBOL DE OBJETIVOS.....	14
2.6 CUÁL ES LA SITUACIÓN DESEADA.....	15
2.7 INTRODUCCIÓN A LA SITUACIÓN DESEADA.....	15
2.8 PROPUESTA DE VALOR.....	17
3. ANÁLISIS DE LAS ALTERNATIVAS TÉCNICAS PARA SOLUCIONAR EL PROBLEMA .....	18
4. MODELO DE NEGOCIO .....	20
4.1 PROPUESTA DE MODELO DE NEGOCIO.....	21
5. PROPUESTA DE LA SOLUCIÓN TECNOLÓGICA .....	27
6. ANÁLISIS DEL PROCESO DE TRANSFORMACIÓN DIGITAL.....	31
7. ASPECTOS LEGALES Y CONTRATACIÓN.....	33

## INTRODUCCIÓN

La presente monografía, desarrollada como trabajo final de la Especialización en Gestión de Servicios de Tecnologías de la Información, tiene como objetivo evidenciar el cumplimiento de las competencias adquiridas durante el programa. Estas competencias incluyen la capacidad de establecer modelos de negocio dentro de marcos legales para servicios Cloud e IoT, seleccionar tecnologías adecuadas basadas en análisis técnico, económico y ambiental, dimensionar arquitecturas TIC que aporten valor a la humanidad y proponer soluciones colaborativas e interdisciplinarias para resolver problemáticas organizacionales y sociales.

En este contexto, se presenta un caso teórico enfocado en la empresa Sutex, cuyo desafío principal es el diseño y la arquitectura de un Data Lake para la integración de datos empresariales generados por los diferentes aplicativos dentro de la empresa. El objetivo de este proyecto es desarrollar una solución en la nube que permite centralizar todas las fuentes de datos de la organización que serán mencionadas en este documento, teniendo en cuenta su origen o tipo, utilizando una arquitectura multi -cloud enfocada en la temática desarrollada en la clase de arquitectura de servicios informáticos. Esto facilitará el análisis profundo de la información y respaldará la toma de decisiones estratégicas basadas en datos generados en la empresa en tiempo real.

El diseño inicial de esta solución se proyecta para ser completado en seis meses, seguido de un desarrollo e implementación que abarca aproximadamente un año. A través de este proyecto, se busca aplicar los conocimientos adquiridos durante la especialización, integrando tecnologías avanzadas, técnicas de diseño de arquitecturas modernas y un enfoque de valor centrado en las necesidades empresariales de Sutex y en el impacto positivo en su gestión operativa y estratégica.

# 1. PROBLEMA

## 1.1. ÁRBOL DE PROBLEMAS

Actualmente, en Sutex SAS se ha identificado una necesidad urgente de integrar la data de los diversos sistemas como ERP, CRM, sistema de diseño y smartsheet de la compañía. Las causas principales de esta situación incluyen un crecimiento tecnológico que no ha avanzado al mismo ritmo que las necesidades de la empresa. Esto ha resultado en sistemas operando de manera independiente, lo que genera múltiples problemas, especialmente en el manejo de la información para realizar un análisis eficiente sobre la información obtenida de estos aplicativos. Esta fragmentación dificulta la proyección de la empresa y la toma de decisiones estratégicas que definen la visión de Sutex, lo cual impacta negativamente en su competitividad en el mercado.

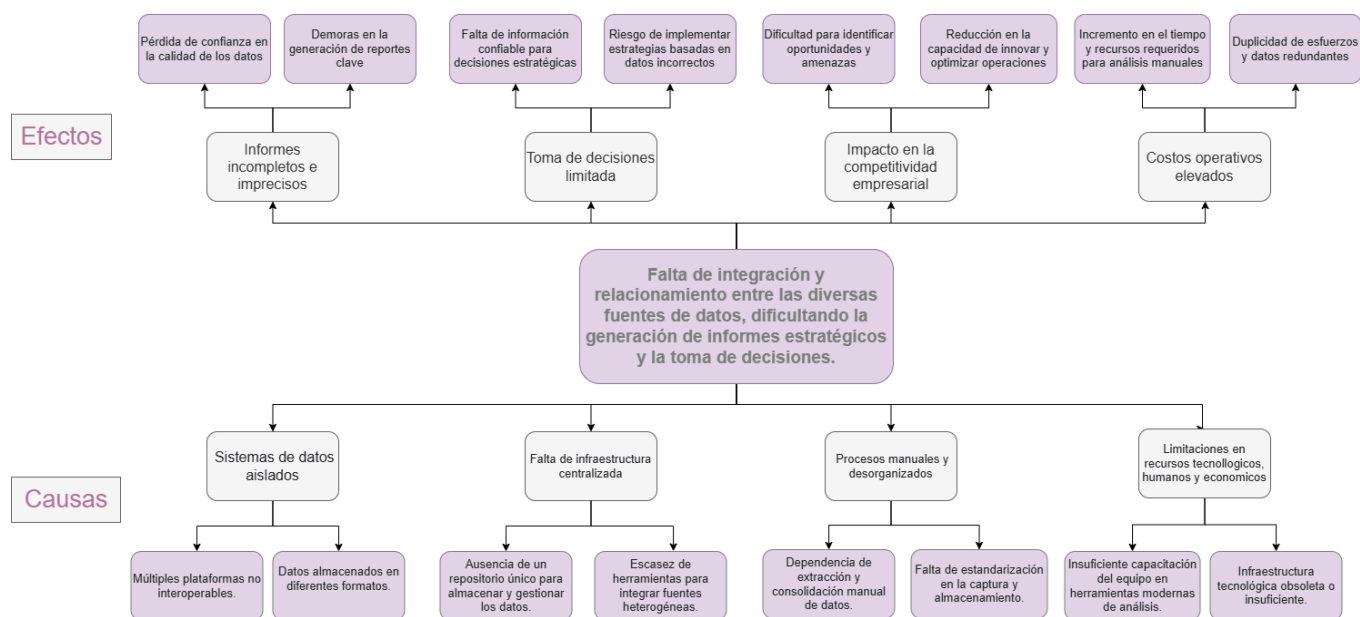


Figura 1. Árbol del problema

## 1.2. ¿QUÉ SE QUIERE SOLUCIONAR?

La empresa Sutex ha logrado un notable crecimiento en el sector textil, destacándose por su capacidad de expansión y diversificación de servicios. Este aumento en su tamaño y oferta ha conllevado también a la adopción y desarrollo de

múltiples aplicativos diseñados para mejorar y automatizar sus procesos internos. Sin embargo, este crecimiento ha resultado en una infraestructura tecnológica fragmentada donde cada solución applicativa opera de manera aislada, abordando solo segmentos específicos del flujo de trabajo general, lo cual ha complicado la integración y el manejo eficiente de los datos.

A medida que Sutex implementa nuevos sistemas, cada uno viene con su propio repositorio de datos, generando silos de información que dificultan la visibilidad global. Los datos generados en estos sistemas varían no solo en tipo, sino también en formato, lo que representa un desafío adicional para la consolidación de la información. Esta falta de integración impide que la empresa realice análisis exhaustivos y obtenga insights valiosos que podrían impulsar aún más su crecimiento y eficiencia.

Ante este panorama, es imperativo para Sutex considerar la adopción de una plataforma unificada que pueda integrar todos estos sistemas dispares. Una solución de este tipo permitiría una gestión de datos centralizada, facilitando consultas y análisis que abarque toda la información de la empresa. Esto no solo mejoraría la toma de decisiones basada en datos, sino que también optimizaría las operaciones al reducir redundancias y mejorar la precisión de la información.

Para abordar estas necesidades, se sugiere la implementación de un sistema de gestión unificado que consolide toda la data de la compañía en una única plataforma en la nube. Esta solución promovería una mayor accesibilidad, escalabilidad y seguridad del sistema, optimizando la gestión y el flujo de información en toda la organización.

## **2. IDEACIÓN DE LA SOLUCIÓN**

Para solucionar la falta de integración entre los sistemas y la descentralización de la información en Sutex, se propone implementar un Data Lake basado en una arquitectura multicloud que tenga las capacidades de Amazon Web Services (AWS) y Google Cloud Platform (GCP). Esta idea de solución permitirá centralizar datos provenientes de diversas fuentes y en múltiples formatos enfocados al problema principal de la empresa, optimizando la gestión y el acceso a la información en un único repositorio.

Utilizando servicios de almacenamiento como Amazon Redshift, es posible integrar y centralizar la información de los diferentes sistemas de la compañía. Esta

integración no solo facilita el acceso a los datos, sino que también mejora la eficiencia en su procesamiento y análisis. Este enfoque centralizado es crucial para superar los desafíos relacionados con la dispersión de datos.

Además, se planea aprovechar herramientas avanzadas como AWS Glue para la orquestación e integración de datos, y BigQuery en GCP para realizar análisis avanzados. Estas tecnologías no sólo facilitarán el procesamiento y análisis de la información, sino que también permitirán extraer insights clave que apoyen la toma de decisiones estratégicas en tiempo real de Sutex.

## 2.1 POR QUÉ SE PLANTEA AHORA LA SOLUCIÓN

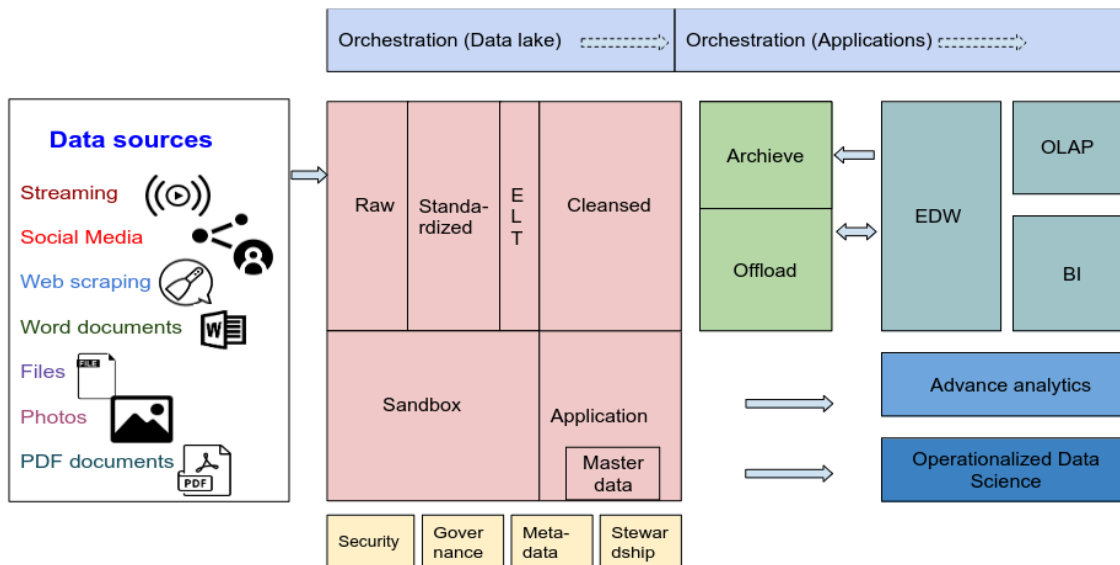
Una vez identificada la problemática de la falta de relacionamiento entre los datos, que dificulta la elaboración de informes estratégicos para la toma de decisiones en la compañía, se consideraron varias tendencias tecnológicas para establecer una posible solución. Estas incluyen el uso de Data Lakes, el análisis de Big Data, el Machine Learning, la integración en la nube, los procesos ETL y la gobernanza de datos. (KYOCERA Document Solutions, 2020)

Una de las primeras tendencias que utilizaremos para resolver nuestra problemática es el Data Lake. Este es un repositorio de almacenamiento centralizado que permite a los usuarios almacenar datos en su formato original, ya sean estructurados, semiestructurados o no estructurados, a gran escala. A nivel de compañía, su implementación implica la creación de una arquitectura en capas que incluye:

- **Capa de datos en bruto:** para la ingesta rápida y eficiente de datos.
- **Capa estandarizada:** para mejorar la transferencia de datos.
- **Capa depurada:** donde los datos se transforman y organizan para su consumo.
- **Capa de aplicaciones:** donde se aplica la lógica empresarial.

Además, se incorporan medidas de seguridad, gobernanza y metadatos para garantizar la integridad, seguridad y facilidad de acceso a los datos. Las tecnologías

utilizadas suelen incluir soluciones en la nube como Azure Data Lake, AWS Lake Formation o Google Big Lake, que proporcionan escalabilidad y durabilidad. Esto optimiza la toma de decisiones basada en datos y facilita el análisis avanzado y en tiempo real. (Nambiar and Mundra 2022)



**Figura 2: Estructura Data lake**

*Fuente: [1] J. Pérez, "Aplicaciones de Data Lake y Data Warehouse en Big Data," Revista de Tecnología y Gestión, vol. 5, no. 2, pp. 123-130, marzo 2023. [En línea].*

Para implementar un sistema tan complejo como un Data Lake, es fundamental agregar la integración en la nube. Este enfoque permite interconectar infraestructuras para intercambiar información y sincronizar datos, independientemente de si se encuentran localmente o alojados en la nube. Esto minimiza las limitaciones que pueden poner en peligro la productividad de las empresas, creando un ecosistema tecnológico unificado y estandarizado. La integración en la nube facilita el movimiento de grandes cantidades de datos, una tarea que usualmente presenta alta complejidad. (www.incentro.com, 2024).

Comparación entre los data lakes y los almacenes de datos: Un data lake también se puede definir en función de lo que no es. No es solo almacenamiento y no es lo mismo que un almacén de datos. (Blockchain, Bitcoin y Criptomonedas, s. f.)

Para poder lograr una solución robusta con lo anterior se debe tener presente que si bien en cierto modo los data lakes y los almacenes de datos almacenan datos, cada uno está optimizado para distintos usos. Considerarlos herramientas complementarias, en lugar de rivales. Además, en algunos casos, es posible que las empresas necesiten usar ambas. Como punto de comparación, los almacenes de datos suelen ser ideales para el tipo de generación de informes y análisis repetible que resulta común en las prácticas comerciales, como los informes de ventas mensuales, el seguimiento de las ventas por región o el tráfico del sitio web. (*¿Qué es un Data Lake? | Google Cloud, s. f.*)

Algunos de los principales beneficios que deberíamos esperar de un data lake en la nube, son:

- Rentabilidad: los proveedores en la nube ofrecen muchas opciones de almacenamiento y precio.
- Autoescala: los servicios en la nube están diseñados para brindar una funcionalidad de escalado que les permita a las empresas computar y aprovechar la capacidad de almacenamiento on-demand.
- Repositorio de datos central: un data lake en la nube reúne la información, y sirve como única fuente de verdad con un acceso gobernado a los datos que habilita la eficiencia de los procesos entre equipos.
- Seguridad de los datos: los proveedores de almacenamiento en la nube garantizan la seguridad de los datos mediante un modelo de responsabilidad compartida.
- Herramientas: los proveedores de almacenamiento en la nube, entre otros, brindan herramientas de ETL que rastrean datos, los catalogan y realizan la preparación, transformación e ingesta para que puedan ser consultados.
- Analíticas mejoradas para obtener nueva información estratégica y mejores resultados de negocio: un data lake en la nube puede combinar los datos de nuevas maneras. Por ejemplo, el CRM y las analíticas de redes sociales pueden brindar nuevos datos sobre la causa de insatisfacción de un cliente, o mostrar qué promociones aumentan la fidelidad. Además, se puede mejorar la eficiencia operativa mediante el análisis de datos de IoT

(*¿Qué Es un Data Lake? | SAP, s. f.-b*)

También es importante tener presente que se tienen desafíos, el principal desafío de una arquitectura de lago de datos es que los datos sin procesar se almacenen sin supervisión de los contenidos. Para que un lago de datos habilite el uso de los datos, debe contar con mecanismos definidos para catalogar y proteger los datos. Sin esos elementos, no se pueden encontrar los datos ni se puede confiar en ellos, lo que resulta en un “pantano de datos”. Para satisfacer las necesidades de audiencias más amplias, los lagos de datos deben tener gobernanza, coherencia semántica y controles de acceso. (*¿Qué Es un Lago de Datos? - Introducción A los Lagos de Datos y los Análisis - AWS, s. f.*)

Dado el gran volumen de información, aplicaremos análisis de Big Data. Este se define como el proceso de examinar grandes conjuntos de datos para descubrir patrones ocultos, correlaciones y otros insights valiosos. Implica extraer y analizar conjuntos de datos diversos, transformándolos en información significativa. Los datos generados en grandes volúmenes, como los de transacciones comerciales, streaming en tiempo real, redes sociales, analítica de video y minería de texto, requieren una integración y análisis eficientes para proporcionar una visión completa y coherente de la base de datos. (Nambiar and Mundra 2022).

Este tipo de análisis es esencial para mejorar la inteligencia empresarial y la toma de decisiones basadas en datos. En la actualidad, existen múltiples herramientas para el análisis de datos, como Hadoop, Spark, bases de datos NoSQL y técnicas de Machine Learning. Estas herramientas permiten manejar y procesar grandes volúmenes de datos de manera eficiente, proporcionando insights valiosos que pueden guiar las estrategias y operaciones empresariales. (Alonso-Zárate, 2022)

Una vez que contemos con toda la información centralizada, necesitaremos una tecnología que nos permita organizarla adecuadamente. Para ello, utilizaremos ETL (Extracción, Transformación y Carga), una herramienta esencial para la integración de datos. ETL se basa en la extracción de datos de diversas fuentes, su transformación mediante procesos de limpieza y estandarización, y finalmente, su carga en el sistema de destino. Este proceso mejora la calidad y consistencia de los datos integrados, facilitando un análisis más preciso y eficiente.

Al aplicar el proceso de extracción, transformación y carga (ETL), los conjuntos de datos en bruto individuales se pueden preparar en un formato y una estructura que son más consumibles para fines analíticos, lo que da como resultado información más significativa. Por ejemplo, los minoristas en línea pueden analizar los datos de los puntos de venta para pronosticar la demanda y administrar el inventario. Los equipos de marketing pueden integrar los datos de CRM con los comentarios de los clientes en las redes sociales para estudiar el comportamiento del consumidor. (www.incentro.com, 2024).

Otras de las tecnologías que se establecerá dentro de la solución es el Machine Learning el cual consiste en una disciplina del campo de la Inteligencia Artificial que, a través de algoritmos, dota a los ordenadores de la capacidad de identificar patrones en datos masivos y elaborar predicciones (análisis predictivo). Este aprendizaje permite a los computadores realizar tareas específicas de forma autónoma, es decir, sin necesidad de ser programados. El término se utilizó por primera vez en 1959. Sin embargo, ha ganado relevancia en los últimos años debido al aumento de la capacidad de computación y al boom de los datos. Las técnicas de aprendizaje automático son, de hecho, una parte fundamental del Big Data. (iberdrola, 2023)

En conclusión, la adopción de estas tendencias tecnológicas aborda de manera efectiva la problemática de la falta de integración de la información en la compañía. Los Data Lakes, combinados con técnicas de Machine Learning, herramientas de integración en la nube, ETL y gobernanza de datos, permiten una gestión y análisis más integrados y eficientes de los datos. Esto facilita la elaboración de informes estratégicos y mejora la toma de decisiones, proporcionando una base sólida para el crecimiento y la competitividad empresarial.

## **2.2 SECTOR OBJETIVO**

El sector textil es una industria clave en la economía global, dedicada a la producción, diseño y comercialización de tejidos, prendas y accesorios. Se caracteriza por su dinamismo, impulsado por la innovación tecnológica y las demandas cambiantes del mercado. Sin embargo, enfrenta constantes desafíos relacionados con la optimización de procesos operativos y la mejora en la toma de

decisiones estratégicas, especialmente en un entorno cada vez más competitivo, donde el análisis y gestión eficiente de datos se han convertido en factores determinantes para su sostenibilidad y crecimiento.

Los cambios en los costos de materias primas y las demandas de sostenibilidad que son utilizados para la creación en dichas empresas como Sutex tienen a tener cambios no solo tecnológicos sino también en ambiente de productividad de telas. En Colombia, la industria textil se destaca por su resiliencia, aprovechando estrategias de diversificación y exploración de mercados emergentes para garantizar su competitividad y estabilidad operativa y esto mismo lo realiza Sutex todo el tiempo.

En esta oportunidad se hará énfasis sobre las características de los procesos del sector textil, entendiendo que se requieren proceso de automatización en la producción de tejidos y confección; agilidad y precisión de los diseños e identificación de las métricas de la producción; planeación y programación en las operaciones, preferencias en los consumidores, reducción de costos de almacenamiento y mejora en la disponibilidad, reducción de errores humanos y eficiencia en los consumos, sostenibilidad y trazabilidad en las cadenas de suministro.

Una de las necesidades fundamentales es la integración de datos, un proceso que busca asegurar el acceso coherente y la entrega de todos los tipos de datos de la empresa. En una organización, todos los departamentos recopilan grandes volúmenes de datos con diferentes estructuras, formatos y funciones. La integración de datos emplea técnicas de arquitectura, herramientas y prácticas para unificar estos datos heterogéneos, permitiendo llevar a cabo análisis exhaustivos y obtener insights valiosos.

Dentro de las aplicaciones que tenemos para la integración de datos en la industria textil, podemos resaltar:

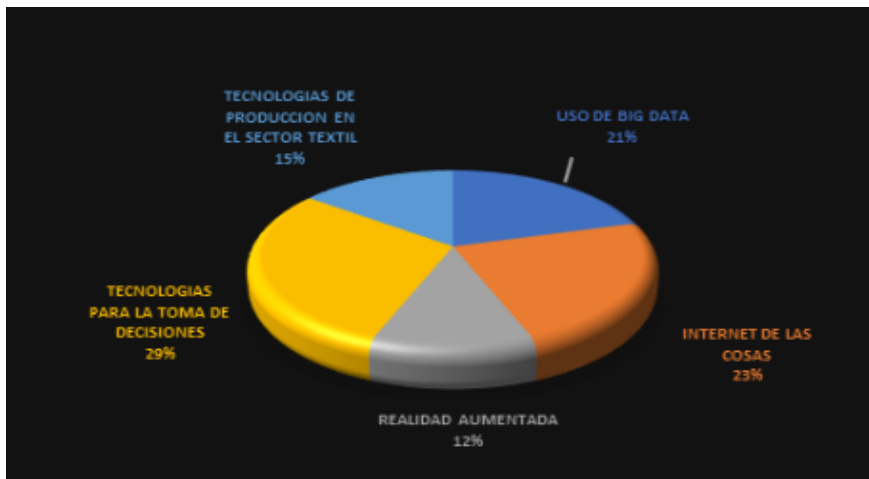
- Optimización de la cadena de suministro, permitiendo la predicción de la demanda de los productos.
- Mejora de la calidad del producto, ayudando a la predicción de defectos y monitorear la calidad de los productos

- Reducción de costos, optimizando la producción y recursos de las compañías.

Basado en las aplicaciones que se pueden dar en este sector, la relación que se establece en la propuesta es que gracias a una centralización de los datos de todos los eventos y resultados del proceso se pueden mejorar de manera continua, llegando a desarrollar productos de mejor calidad, además se tendría un aumento de las ventas si se tiene una predicción hacia donde se mueve el mercado del sector.

### 2.3 TENDENCIAS DEL SECTOR

Dentro del marco de la revolución industrial actual, la llamada Industria 4.0 presenta como tendencias herramientas como Big data, Internet de las cosas (IoT), Realidad aumentada, tecnologías de producción y tecnologías para la toma de decisiones.



**Figura 3.** Tendencias tecnológicas sector textil

Con frecuencia, observamos que las empresas del sector textil tienen la oportunidad de implementar nuevas tecnologías que pueden ayudar a mejorar sus procesos de producción, haciéndolos más rápidos y rentables, optimizando los costos y mejorando la inspección de calidad. A continuación, se presentan algunas

innovaciones en la fabricación de textiles, muchas de las cuales están impulsadas por la robótica industrial. (gruposim.eu, 2024)

### Big Data e Inteligencia Artificial

El uso y análisis de datos masivos y los sistemas de automatización con Inteligencia Artificial están permitiendo a las industrias del sector textil y el tapizado mejorar determinados procesos como la previsión de la demanda, influyendo en la estrategia de producción para optimizar los costes y aumentar la rentabilidad de la industria. (grupo sim.eu, 2024)

#### Gestión de inventario inteligente

Muchas empresas están implementando soluciones que les ayudan a optimizar el uso de materias primas del ámbito textil, como es el caso de las empresas en el sector del descanso, donde es primordial hacer una gestión eficaz de las telas, armazones y rellenos para optimizar los costes, evitar el exceso de materiales sobrantes y mejorar la eficacia en los procesos de producción. Esto es posible gracias a la implementación de soluciones de tecnología de software como ERP para el sector del descanso. (gruposim.eu, 2024)

## 2.4 ANÁLISIS DE MERCADO

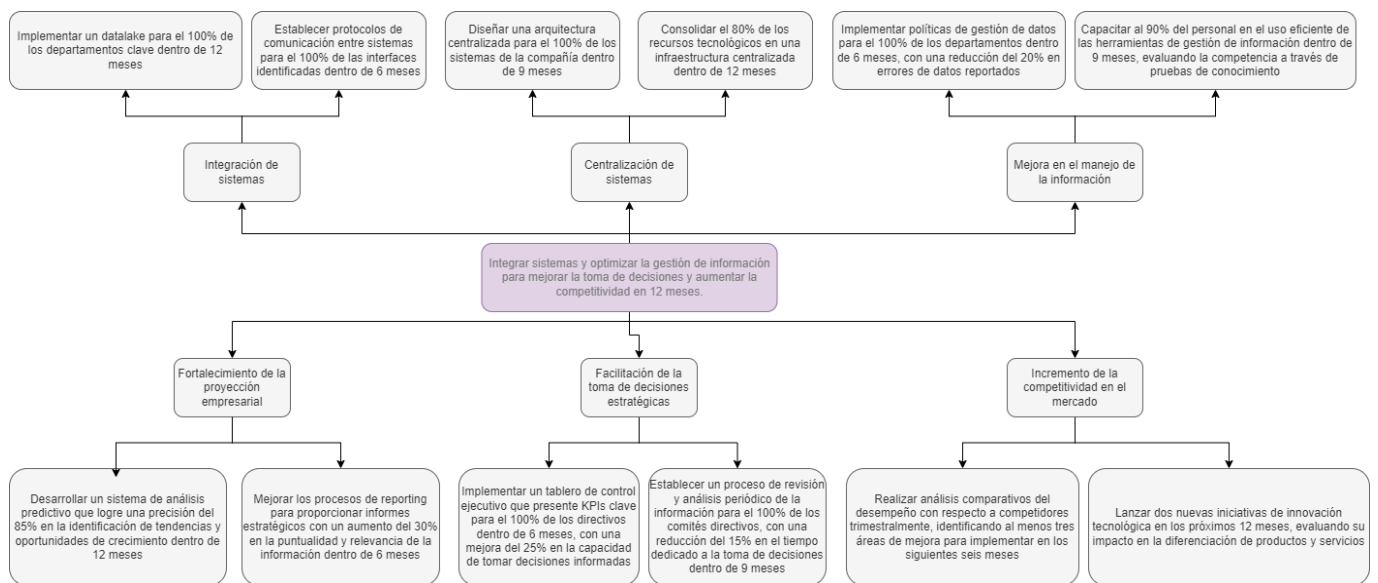
La aplicación de modelos de TI dentro de una organización debe tener un enfoque en las necesidades y las estrategias hacia dónde se orienta la compañía. Para el proyecto de TI que estamos analizando, se considera un enfoque centrado en la creación de un modelo sostenible que tenga recursos tecnológicos, procesos operativos y estrategias de mercado que puedan solventar a Sutex. Esto incluye la definición de un modelo de ingresos claro, basado en la optimización de costos mediante tecnologías escalables y en la generación de valor a través del análisis de datos para la toma de decisiones que hoy en día se está cubriendo con un Data Lake o una base de datos muy grande.

La implementación y puesta en marcha de un data lake multcloud, por ejemplo, se piensa no solo como una herramienta tecnológica de gran ayuda, sino como una pieza fundamental para transformar datos en activos estratégicos aportando a los objetivos de las empresas. En este modelo se consideran los costos operativos asociados, las oportunidades de monetización de los datos procesados y

las proyecciones de retorno de inversión al mejorar la eficiencia y competitividad de la organización. De esta manera, el proyecto se enmarca en una visión que integra la tecnología con el modelo de negocio, generando un impacto positivo tanto en los resultados financieros como en la capacidad de innovación de la empresa.

## 2.5 ÁRBOL DE OBJETIVOS

En la siguiente figura se muestra una identificación de la problemática, donde se expone las causas y las consecuencias que se obtienen del problema expresado.



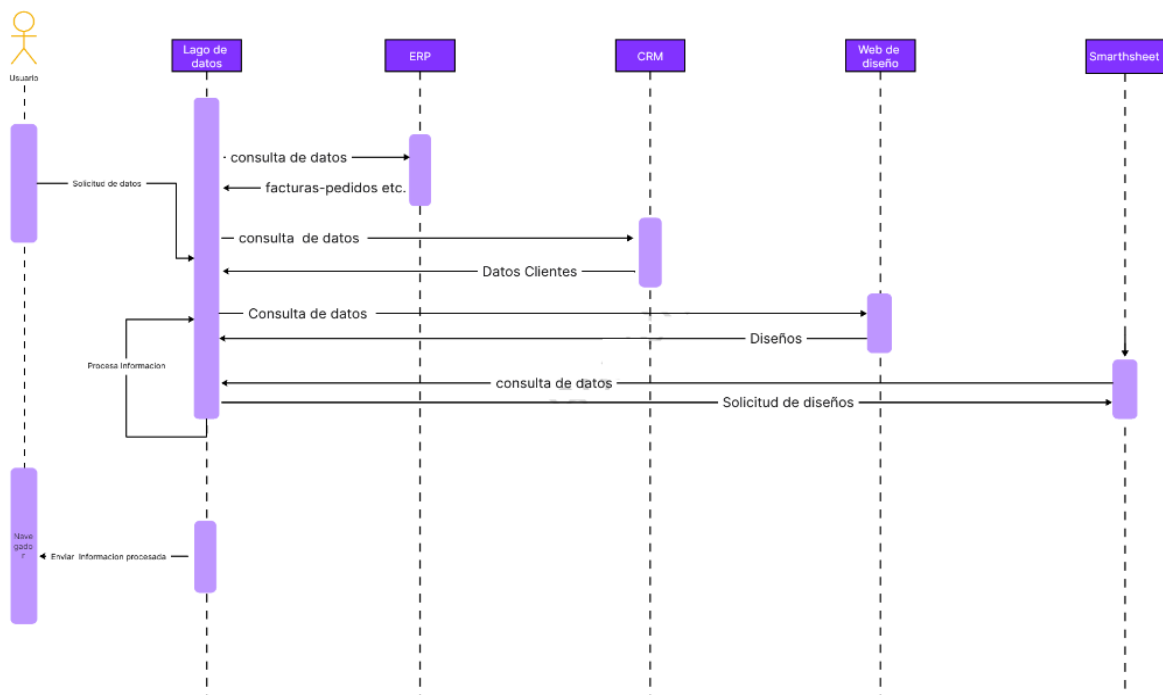
**Figura 4.** Árbol de objetivos

- Diseñar un esquema de arquitectura para integrar diferentes aplicativos, mostrando relaciones y flujos de datos, con diagramas y especificaciones técnicas. Utilizando los recursos disponibles y colaborando con los equipos para mejorar la eficiencia y alinearse con los objetivos organizacionales, completando el esquema en 6 meses.
- Asegurar que los datos integrados sean precisos, consistentes y actualizados, eliminando duplicados y corrigiendo errores, para facilitar análisis avanzados y decisiones informadas en cada una de las áreas de la compañía cada 6 meses.
- Diseñar la solución de integración de datos para que pueda crecer y adaptarse a medida que aumenten las necesidades de la organización.

## 2.6 CUÁL ES LA SITUACIÓN DESEADA

La situación deseada para Sutex con este proyecto es la implementación de una solución tecnológica escalable y eficiente que permita centralizar todas las fuentes de datos de la compañía en una arquitectura multicloud, utilizando las plataformas de AWS y Google Cloud (GCP). Esto garantizará la integración de datos estructurados y no estructurados provenientes de diversas áreas de la empresa que están en sistemas como ERP y CRM, eliminando los procesos de información que actualmente dificultan su análisis ya que son procesos manuales. La solución optimizará la gestión de datos y su visualización, sino que también podrá disponibilizar para todas las personas con roles dentro de Sutex, teniendo en cuenta la seguridad y precisión en ellos, creando una herramienta que centralice la información del módulo de ventas para la toma de decisiones informadas y estratégicas.

Con esta infraestructura centralizada que se propone, Sutex podrá acceder a herramientas avanzadas de análisis y visualización de datos que se van a describir en este documento que potenciarán la generación de insights valiosos para los altos directivos y las personas que lo quieran ver. Esto impulsará la capacidad de la empresa para adaptarse rápidamente a las dinámicas del mercado textil, mejorando la eficiencia operativa.

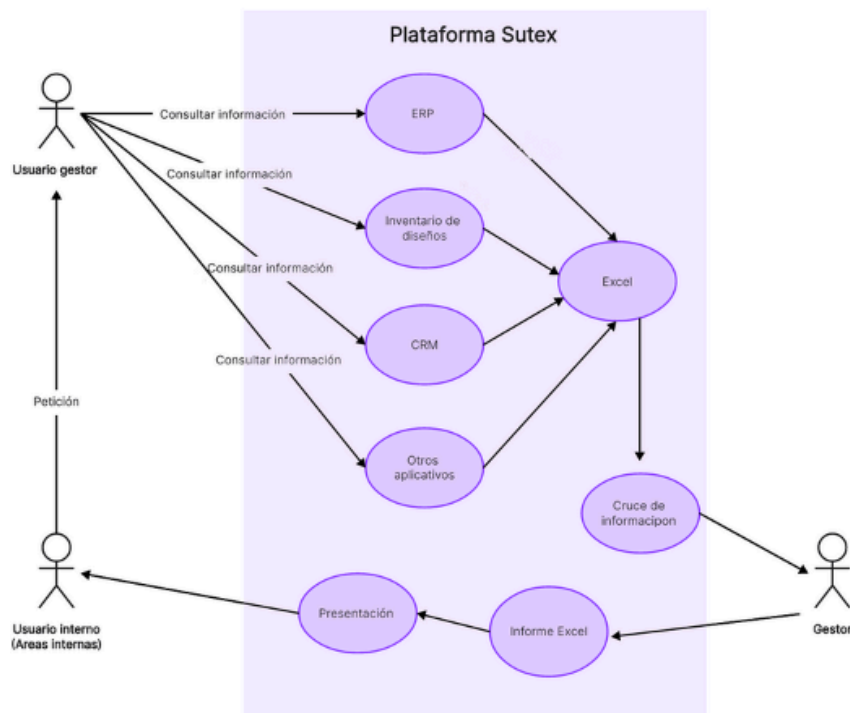


**Figura 5.** Diagrama de secuencia situación esperada

## 2.7 INTRODUCCIÓN A LA SITUACIÓN DESEADA

Situación actual:

El siguiente diagrama refleja el proceso actual en Sutex, que integra diversos sistemas y aplicaciones clave para optimizar los procesos operativos y de gestión. En este modelo, se destacan tres componentes principales: el ERP (Enterprise Resource Planning), el Inventario de Diseño y el CRM (Customer Relationship Management), junto con otros aplicativos relacionados. Cada uno de estos sistemas juega un rol fundamental en la recopilación, procesamiento y análisis de datos provenientes de distintas áreas.



**Figura 6.** Diagrama situación actual.

Situación esperada:

En la solución implementada tenemos los orígenes de datos con fuentes como el ERP y CRM, estas fuentes se conectan a la nube de AWS para llegar a un Data lake el cual almacena la información en Amazon S3 para luego pasar a la etapa de transformación y después de esto es analizada con la herramienta de Bigquery para luego pasar a ser representada por medio de reportes o dashboard.

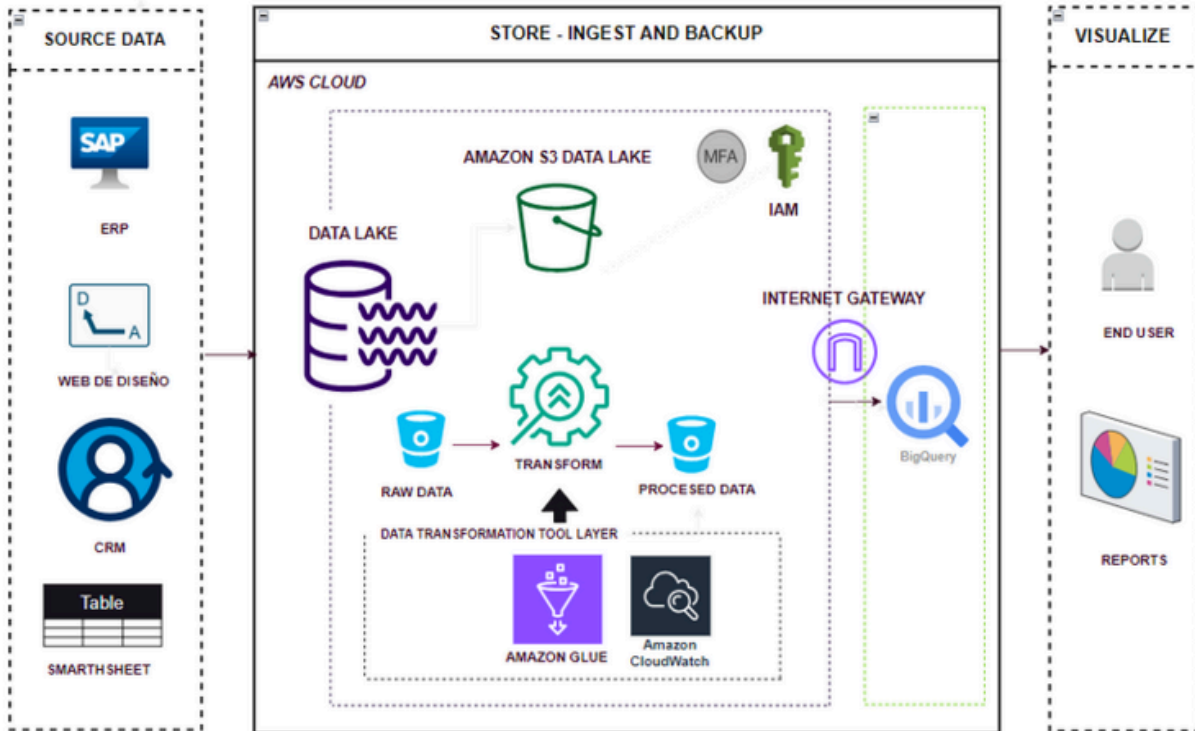


Figura 7. Diagrama situación esperada

## 2.8 PROPUESTA DE VALOR

### 2.8.1 PERFIL DEL CLIENTE

Como se especifica en la situación deseada, el proyecto se enfoca en las compañías pequeñas y mediana empresa, las cuales no cuentan con sistema de análisis de datos, ya sea por su limitación de recursos o por su falta de innovación. Tener acceso a estas nuevas tecnologías le aportaría una visión más general del estado actual de su compañía y cómo se podría mejorar, basándose en un análisis de la información.

### 2.8.2 MAPA DE VALOR

Implementar un Data Lake multcloud con AWS y GCP que le permitirá a Sutex solventar la falta de integración y la descentralización de los datos. Este enfoque proporciona una solución escalable que permite mejorar la accesibilidad, optimizar la gestión de datos para así realizar un análisis avanzado para tomar decisiones más rápidas y acertadas.

### **3. ANÁLISIS DE LAS ALTERNATIVAS TÉCNICAS PARA SOLUCIONAR EL PROBLEMA**

La solución propuesta es la implementación con servidores Cloud los cuales permiten tener diferentes beneficios como escalabilidad, disponibilidad entre otros, los proveedores con los cuales se realizaron las cotizaciones con Google Cloud y Amazon Web Service.

Primera solución propuesta es Google Cloud:

Cuenta con las siguientes características:

- Redundancia Data Lake Storage Gen2, Estándar y LRS, Acceso frecuente Nivel de acceso, Estructura de archivos Espacio de nombres plano, Capacidad: 100 TB - Pago por uso, operaciones de escritura: 4 MB x 10 operaciones, operaciones de lectura: 4 MB x 10 operaciones, 10 operaciones de lectura iterativas, 100.000 operaciones de lectura de prioridad alta de Archive Storage, 10 operaciones de escritura iterativas, 10 operaciones de otro tipo. 1000 GB Recuperación de datos, 1000GB de recuperación de prioridad alta de Archive Storage, 1000 GB Escritura de datos, costo mensual 2196,54 US\$.

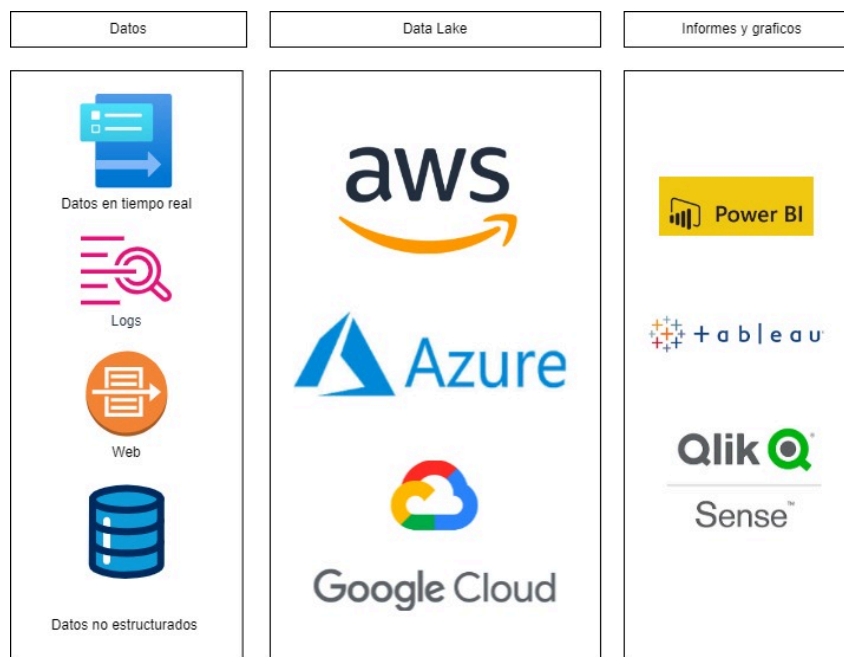
Segunda solución propuesta de Amazon Web Service:

Cuenta con las siguientes características:

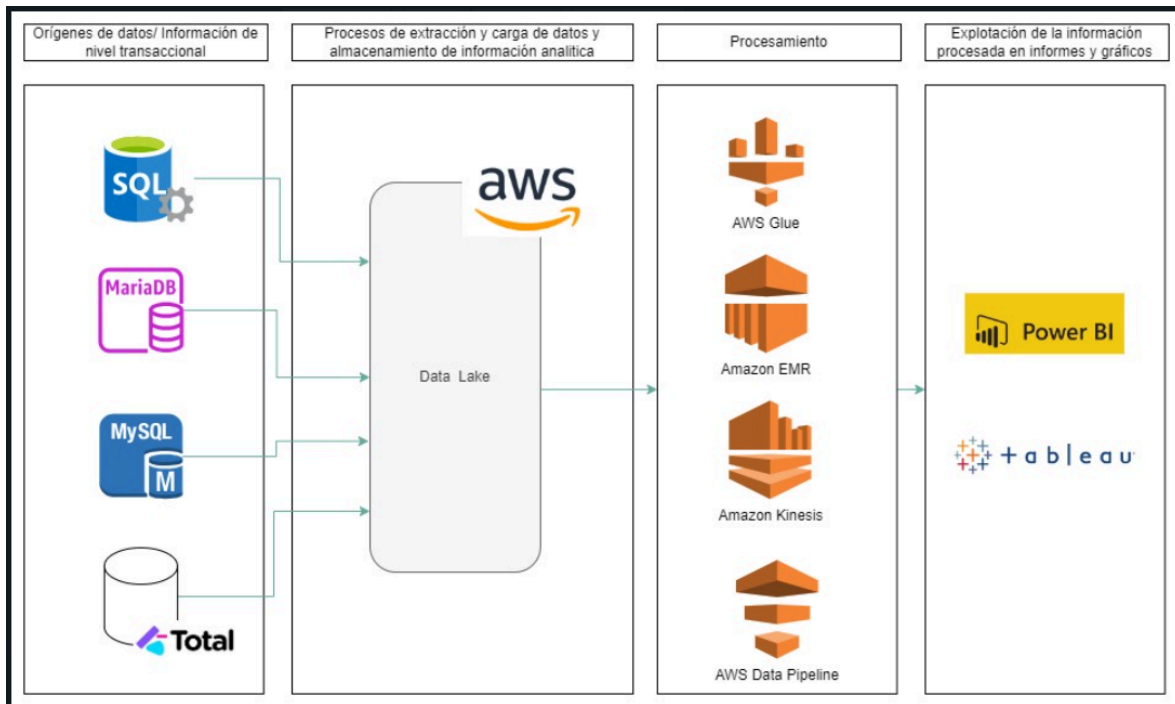
- Config summary: Nodes(1). Instance type (dc2.large), Utilization (On-Demand only) (100%utilized/Month), Pricing strategy (OnDemand), Additional backup storage (100 TB), costo mensual 2.486,50 US\$.

Característica	Azure Data Lake Analytics	Amazon Redshift
<b>Tipo de Servicio</b>	Análisis de Big Data basado en el lago de datos	Data Warehouse en la nube
<b>Propósito Principal</b>	Procesamiento y análisis de grandes volúmenes de datos sin estructura	Almacenamiento y análisis de datos estructurados
<b>Modelo de Ejecución</b>	Servidorless; paga por trabajo realizado	Basado en clúster; paga por nodo/hora
<b>Lenguaje de Consulta</b>	U-SQL (combinación de SQL y C#)	SQL estándar con extensiones para análisis de datos
<b>Integración de Datos</b>	Directamente con Azure Data Lake Storage, puede acceder a otros datos en Azure Blob Storage	Integración con S3, AWS Glue para ETL, Redshift Spectrum para consulta de datos en S3
<b>Escalabilidad</b>	Escalabilidad automática según la carga de trabajo	Escalabilidad manual mediante adición de nodos
<b>Precio</b>	Pago por uso; basado en la cantidad de datos procesados y el tiempo de consulta	Pago por nodo/hora; diferentes tipos de nodos disponibles para optimización de costos
<b>Seguridad</b>	Integración con Azure Active Directory, cifrado de datos en reposo y en tránsito	Integración con AWS IAM, cifrado de datos en reposo y en tránsito, opciones de red privada virtual
<b>Facilidad de Uso</b>	Basado en scripts, ideal para desarrolladores acostumbrados a la programación	Basado en SQL, fácil de usar para analistas de datos y usuarios familiarizados con SQL
<b>Disponibilidad y Durabilidad</b>	Alta disponibilidad garantizada por la infraestructura de Azure	Alta disponibilidad con replicación en clúster y opción de replicación entre regiones
<b>Casos de Uso</b>	Procesamiento de grandes volúmenes de datos no estructurados, ETL, análisis de big data	Almacenamiento y análisis de datos estructurados, BI, reporting
<b>Herramientas de Integración</b>	Azure Synapse Analytics, Azure Databricks, Power BI, herramientas de desarrollo de Azure	Amazon QuickSight, AWS Glue, integración con herramientas de BI y ETL de terceros
<b>Rendimiento</b>	Optimizado para grandes volúmenes de datos y procesamiento distribuido	Alto rendimiento para consultas SQL complejas, optimizaciones automáticas de consultas

**Tabla 1.** Árbol del problema *cuadro comparativo tecnologías de análisis de datos.*



**Figura 8.** Árbol del problema *Estructura integral de gestión y análisis de datos*



**Figura 9.** Árbol del problema *Estructura propuesta para la solución AWS*

De acuerdo con las propuestas analizadas, la que más se ajusta a nuestra solución es la brindada por Amazon Redshift, esto debido a que la información a ingresar es netamente estructurada y la configuración es más accesible e interactiva.

#### 4. MODELO DE NEGOCIO

La expansión de consorcios empresariales como Sutex, ha permitido una diversificación valiosa y --- Con lo cual, como se ha mencionado en la descripción del problema (Numeral 1) el manejo y análisis de información requiere un tratamiento de datos complejo y dispendioso. El modelo de negocio propuesto, hace referencia a una solución de alta gama, enfocada en los pilares del manejo de la información:

- Confidencialidad
- Disponibilidad
- Integridad

Es claro que el manejo de grandes volúmenes de información, tiene que estar totalmente conectado a las evidencias demostradas por la integración de

herramientas sobre plataformas en la nube. Esto formaliza totalmente la gestión y análisis, teniendo presente que resulta inútil un gran almacenamiento, sin un análisis eficiente de dicha información. El objetivo siempre será la toma de decisiones basada en la minería de datos correspondiente.

Al desarrollar este modelo de negocio, es clave implementar una estructura comercial que ya cuenta con resultados comprobados.

#### 4.1 PROPUESTA DE MODELO DE NEGOCIO

##### **Selección de clientes:**

Es clave tener presente a quién creamos valor, ya que sus características forman parte del aporte para lograr exponer las necesidades de forma mucho más explícitas y detalladas. Al examinar el mercado y los objetivos particulares de una organización como Sutex, logramos afinar las herramientas que puedan proporcionar al final del proceso, una visualización mucho más detallada para la toma acertada de decisiones.

##### **PROPUESTA DE VALOR:**

Con el avance de nuevas tecnologías, es indispensable la aplicación de herramientas como Data Lake, Amazon S3, AWS; la interpretación de estos grandes volúmenes de datos mediante Machine Learning, requiere igualmente mecanismos de integración que están incluidas en AWS para la evaluación de datos, logrando minimizar el riesgo de consideraciones sin bases estadísticas (error humano)

##### **CANALES DE DISTRIBUCIÓN Y VENTA**

Es determinante establecer caminos claros que conduzcan a una comunicación confiable y directa con los clientes, sus organizaciones, y cada estructura comprometida con el proceso y el resultado final. Contar con tecnologías que brinden información con integridad y disponibilidad; herramientas de distribución asequibles y verificables para el usuario final, y elementos de fácil manejo, pero con respuestas objetivas y confiables, logran que el nivel de satisfacción mantenga un alto grado de calificación por parte del usuario final. Es por eso que incluir la información necesaria en los Sistemas de Información de Soporte, páginas Web,

Redes Sociales, y demás elementos de Marketing Digital, forma parte de la propuesta de valor que de manera permanente apoya los procesos de venta, entrega y posventa.

### **RELACIÓN CON LOS CLIENTES:**

Cada cliente tiene sus propias características; por lo tanto, es importante identificar el nicho al que pertenece, sus objetivos y desde luego sus necesidades. Esto requiere un lenguaje determinado conforme a cada situación, un pulso medido para cuidar y prolongar estas relaciones, integrando la tecnología de Web Services, plataformas de chat, Soporte Técnico básico y avanzado, los soportes en línea, pero igualmente la complejidad de la comunicación humana que hace parte de un servicio integral para cultivar las grandes relaciones comerciales que puedan generar no solo satisfacción al cliente final, sino un “voz a voz” que impulse este tipo de soluciones en otros gremios y organizaciones.

Para esto, utilizar sistemas de información que registren cada proceso, combinando la interacción humana con la agilidad de los llamados “chatbots” de respuesta inmediata, hace que haya un seguimiento asertivo y organizado fundamentado en canales de comunicación confiables y redundantes, lo cual previene la fuga de información y la retoma de datos, algo común en muchos de los sistemas de Soporte en Línea que terminan por fracturar las relaciones con los clientes.

### **INGRESOS:**

El ingreso inicial tiene como origen la implementación de un paquete de soluciones conforme a la integración y análisis de datos, pero al integrar plataformas mediante aplicaciones de terceros, se incluyen tanto costos como ingresos adicionales que van de la mano de requerimientos fuera de las propuestas; estos pueden estar relacionados con personalización de servicios, microanálisis sobre ambientes puntuales, datos estadísticos específicos y demás integraciones que el cliente defina o que comercialmente sean viables para una mejor toma de decisiones.

Permitir estas alianzas, da paso a abrir puertas de valor agregado que visualizan el valor de nuestro producto en el mercado mucho más fortalecido, y da paso a nuevas fuentes de ingreso con cada necesidad explorada. Para el caso Sutex, es válido determinar una propuesta que permite un ingreso considerable, entendiendo que la inversión en tecnología requiere una inversión importante en tecnología, que a largo plazo soluciona seguros retrasos en la gestión y toma de decisiones, lo cual termina por minimizar costos.

### **RECURSOS Y CAPACIDADES CLAVES**

Ya que el nombre del juego es “competencia”, es importante entender que esta es una propuesta que muestra un alto nivel de tecnología; lo cual requiere el compromiso de una respuesta eficiente, un alto grado de capacitación, y desde luego contar con una infraestructura que sobrepase la normatividad, y las exigencias de comunidades que avanzan a pasos agigantados en el mundo de las actualizaciones, teniendo presente la escalabilidad sin llegar a la sobredimensionabilidad. La constante autoevaluación, seguimiento, socios de negocio, planes de desarrollo y mantenimientos, son pilares de una plataforma eficiente y productiva.

### **ACTIVIDADES CLAVE:**

El monitoreo constante de los procesos y de los recursos, es un elemento prioritario en el factor de esta propuesta. Al realizar una implementación, hay que contar con un desgaste natural de los recursos, con la posibilidad de las fallas humanas, y crear programas de mantenimiento, e incluso planes de choque, junto con indicadores de gestión (KPIs), métricas y demás herramientas adaptadas a soluciones rápidas y de bajo impacto ante situaciones de fallas críticas o de alto riesgo.

### **ALIANZAS CLAVE:**

Esto significa igualmente contar con alianzas, convenios, (partnerships), y demás elementos de tercerización que apoyen de manera transparente el

cumplimiento de metas y demás objetivos trazados. Al final, uno de los logros de este punto, es evitar sobrecostos por fallas en la planeación estratégica, y en la anticipación de situaciones de riesgo. El análisis de proveedores forma parte crítica de los procesos, ya que, al no tener influencia directa, el riesgo de falla puede incrementarse. Es por esto que se hace preciso evaluar el background correspondiente, temas contractuales y demás elementos que permitan un seguimiento y cumplimiento dentro de las áreas a subcontratar. En esta propuesta, contar con marcas tecnológicas reconocidas hace que el nivel de confianza en los aplicativos, proveedores, etc, sea determinante.

### GASTOS / COSTOS:

Para determinar un margen de ganancia, es necesaria una evaluación consciente de gastos y costos, los cuales se diferencian al entender que los últimos forman parte de la producción (activos), mientras que los gastos, aunque forman parte de los desembolsos de la empresa, no están directamente relacionados con producción. Al aclarar lo anterior, sabemos que un presupuesto requiere el análisis de costos y gastos, que para esta propuesta tienen conexión con los asuntos administrativos como pago de nóminas, contrataciones, y demás recursos laborales, proveedores, infraestructura (de ser necesario), licenciamientos y demás elementos que requieran inversión o que, aunque no tengan un retorno directo forman parte del proceso correspondiente al producto o servicio.

### Modelo Canvas

Relaciones Clave	Actividades Clave	Propuesta de Valor	Relaciones con los clientes	Clientes
<p>Proveedores de servicios en la nube: AWS, Microsoft Azure, Google Cloud o cualquier otro proveedor que ofrezca la infraestructura de Data Lake y procesamiento de datos.</p> <p>Proveedores de herramientas de visualización: Tableau, Power BI, o cualquier plataforma de visualización que integre en la solución.</p>	<p>Configuración, mantenimiento y soporte de las plataformas de terceros en nombre de tus clientes</p> <p>Supervisión constante de los costos, rendimiento y seguridad en las plataformas de terceros utilizadas por el cliente.</p>	<p>Ofrecer a las empresas una solución "llave en mano" donde integras y gestionas servicios de terceros como AWS, Azure, o Google Cloud, eliminando la complejidad de configurar estas plataformas ellos mismos.</p> <p>Ayudar a las empresas a gestionar sus costos al optimizar el uso de los recursos en la nube de terceros, evitando sobrecargos innecesarios.</p> <p>Escalabilidad, seguridad y accesibilidad desde cualquier lugar.</p>	<p>Soporte técnico 24/7 a través de chat, email y teléfono.</p> <p>Consultoría personalizada durante la implementación y configuración del Data Lake.</p> <p>Programas de fidelización para ofrecer actualizaciones y mejoras periódicas a la plataforma.</p> <p>Capacitación y webinars para ayudar a los equipos de clientes a utilizar los dashboards y la analítica avanzada.</p>	<p>Empresas medianas y grandes que manejan grandes volúmenes de datos.</p> <p>Sectores con necesidades de analítica avanzada, como finanzas, retail, salud, telecomunicaciones, y manufactura.</p> <p>Equipos de analistas de datos, departamentos de TI, y ejecutivos que requieren insights visuales para la toma de decisiones.</p>
	<p><b>Recursos Clave</b></p> <p>Personal con experiencia en la configuración y manejo de plataformas de terceros como Data Lakes, soluciones de visualización, y servicios de nube.</p> <p>Tener acuerdos o partnerships con proveedores de nube y plataformas de visualización de datos.</p>		<p><b>Canales de Distribución</b></p> <p>Venta directa a través de un equipo comercial especializado en B2B.</p> <p>Presencia en línea mediante una página web y marketing digital (SEO, SEM).</p>	

### **Figura 10.** Modelo canvas

#### 4.2 VALIDACIÓN DEL MODELO DE NEGOCIO

Para la validación del modelo de negocio se plantean las siguientes hipótesis.

##### **Hipótesis por parte del cliente:**

Las empresas medianas y grandes que operan en sectores como finanzas, retail, salud, telecomunicaciones y manufactura experimentan dificultades para gestionar de manera eficiente grandes volúmenes de datos y están dispuestas a invertir en soluciones 'llave en mano' que integren servicios de nube y visualización de datos para facilitar su análisis y toma de decisiones.

##### **Cómo Validar esta Hipótesis:**

1. Entrevistas Cualitativas: Se realizarán entrevistas con responsables de TI, analistas de datos, y ejecutivos de empresas de estos sectores para confirmar si este es un problema real y cómo lo enfrentan actualmente.
2. Encuestas Cuantitativas: Se realizarán encuestas a muestra representativa de empresas en estos sectores para medir el interés en una solución "llave en mano" y la disposición a pagar por ella.
3. Pruebas de Concepto: Ofrecer una versión simplificada de la solución a un grupo de empresas para observar si realmente están dispuestas a probar y adoptar la solución.
4. Campañas de Marketing: Lanzar una campaña de marketing enfocada en estos segmentos para medir el interés y la tasa de respuesta de clientes potenciales.

##### **Hipótesis del problema.**

Las empresas medianas y grandes tienen dificultades para analizar grandes volúmenes de datos dispersos en silos de información, lo que ralentiza el análisis y consiguientemente la toma de decisiones y limita la capacidad de obtener insights integrados.

### **Cómo Validar esta Hipótesis**

Con el fin de validar si este es un problema real y crítico en nuestro mercado objetivo, se propone realizar las siguientes actividades:

1. Entrevistas en profundidad: Preguntar a líderes de TI y analistas de datos sobre sus mayores desafíos al manejar datos distribuidos en silos. Indagando cómo esta dispersión afecta la capacidad de generar informes integrales y tomar decisiones rápidas.
2. Estudios de casos de empresas: Se propone identificar casos donde la dispersión de datos en silos haya causado problemas de ineficiencia, incoherencia en los datos o retrasos en la toma de decisiones.
3. Encuestas dirigidas: Otro proceso es realizar encuestas en línea para entender cuántas empresas perciben la dispersión de datos en silos como un problema crítico, y si ya han intentado abordar esta situación.
4. Prueba piloto con un MVP: Se propone crear un demo o versión inicial de la solución que integre diferentes silos de datos en una plataforma centralizada, y medir la mejora en la calidad de los insights y la velocidad de la toma de decisiones.

### **Hipótesis del producto.**

Una plataforma centralizada que permita a las empresas integrar y analizar grandes volúmenes de datos de múltiples silos reducirá el tiempo necesario para generar insights estratégicos y mejorará la calidad de las decisiones de negocio.

### **Cómo Validar esta Hipótesis**

Con el fin de definir si este producto resuelve efectivamente el problema de tus clientes, podemos realizar las siguientes pruebas:

1. Desarrollar un MVP: Crear un prototipo básico que permita a los usuarios integrar datos de al menos dos o tres fuentes distintas y analizar los

resultados en una sola plataforma. Se debe medir el tiempo que les toma a los usuarios pasar de la carga de datos a la generación de insights.

2. Pruebas de usuario con clientes potenciales: Invitar a equipos de TI y analistas de datos de empresas medianas y grandes a probar la plataforma y proporcionar retroalimentación sobre la facilidad de uso, velocidad de integración y calidad de los insights obtenidos.
3. Métricas de éxito: Establecer métricas clave, como el tiempo de adopción, la reducción en el tiempo de análisis, y el nivel de satisfacción del usuario con la precisión y utilidad de los insights generados.
4. Pruebas de A/B: Si es posible, realizar pruebas de A/B para comparar la plataforma con las soluciones actuales que utilizan los usuarios, enfocándose en la velocidad de integración y la calidad del análisis.

La validación de este modelo de negocio se basa en tres hipótesis clave: que las empresas medianas y grandes necesitan integrar datos de silos para mejorar la toma de decisiones; que la dispersión de datos dificulta la generación de insights estratégicos; y que una plataforma centralizada puede resolver este problema de manera efectiva, reduciendo tiempos y mejorando la precisión del análisis. Confirmar estas hipótesis permitirá ajustar la solución a las necesidades del mercado, asegurando su viabilidad y potencial de crecimiento.

## **5. PROPUESTA DE LA SOLUCIÓN TECNOLÓGICA**

La propuesta de solución tecnológica consiste en implementar un Data Lake en la nube de AWS y GCP para centralizar y unificar la información de la empresa Sutex SAS, integrando datos de fuentes y sistemas internos distintas unas a las otras en un único repositorio escalable y eficiente. El servicio prestado incluirá la ingesta, almacenamiento, procesamiento y análisis de datos de la empresa que pueden estar enfocados a diferentes etapas que pueden ser asociados a módulos de la empresa, permitiendo la generación de informes estratégicos y la toma de decisiones informadas sobre esos datos y el enfoque que el usuario final le quiera dar. El Data Lake estará alojado en un entorno cloud para aprovechar la

infraestructura y su seguridad que nos ofrecen los servicios modernos que manejan este tipo de soluciones para empresas de todos los tamaños y enfocada a medianas como lo es Sutex.

La selección de tecnologías para esta solución se tuvo en cuenta los servicios de cloud computing, herramientas de procesamiento de datos y métodos de integración que trabajan mejor estas soluciones tecnológicas y que hoy en día están muy bien enfocadas para abordar los problemas de la empresa Sutex y nuestro foco. A continuación, se presentan las tecnologías para cada componente del Data Lake:

### **Almacenamiento en la Nube**

- **Herramienta Seleccionada:** Amazon S3 (Simple Storage Service)

Se eligió la herramienta dado que Amazon S3 es una solución escalable para el almacenamiento de datos en la nube. Su modelo de precios basado en el uso, la capacidad de manejar datos estructurados y no estructurados, y su integración con otros servicios de AWS lo convierten en una opción ideal para construir un Data Lake

### **Ingesta de Datos**

- **Herramienta Seleccionada:** AWS Glue

Se eligió AWS Glue ya que es un servicio de extracción, transformación y carga (ETL) serverless que facilita la integración de datos desde diversas fuentes y hoy en día eso es lo que tenemos en Sutex. Permitiendo la ingesta de datos en formato crudo y su organización en diferentes capas (bruta, estandarizada y depurada).

### **Procesamiento de Datos**

- **Herramienta Seleccionada:** Amazon Redshift

Se eligió y se ve en la arquitectura porque Amazon Redshift es un almacén de datos escalable que facilita la ejecución de consultas analíticas rápidas sobre volúmenes de datos. Su capacidad de integración con S3 permite

realizar análisis de datos directamente en el Data Lake sin necesidad de mover datos y ya se eligió en la parte superior en la arquitectura.

## **Análisis y Machine Learning**

- **Herramienta Seleccionada:** Google Bigquery  
Se eligió dado que Google Bigquery es una opción robusta, que nos permite el análisis de grandes cantidades de datos e integrarlos con plataformas de dashboard como Looker studio, Tableau o Power Bi.

## ***Arquitectura de la Solución***

La arquitectura propuesta tiene una estructura en capas que optimiza la gestión de datos y la escalabilidad de la solución. A continuación, se detalla cada capa:

1. **Capa de Ingesta de Datos:** Los datos de diversas fuentes (ERP CRM, plataformas de diseño, etc.) Se ingieren en su formato original y se almacenan en la capa de datos brutos en S3.
2. **Capa de Procesamiento ETL:** Utilizando AWS Glue, los datos se limpian y transforman para su análisis. Este proceso incluye la eliminación de duplicados, la corrección de errores y la estandarización de formatos.
3. **Capa de Almacenamiento Estandarizado:** Los datos transformados se almacenan en un formato optimizado para mejorar el rendimiento en las consultas de análisis.
4. **Capa de Almacenamiento Analítico:** Amazon Redshift se utiliza para almacenar los datos preparados para el análisis y la generación de informes.
5. **Capa de Machine Learning:** Amazon SageMaker extrae datos de Redshift y S3 para entrenar modelos de machine learning y generar predicciones.
6. **Capa de Presentación y Visualización:** Los resultados se visualizan a través de herramientas como BigQuery, lo que permite la creación de dashboards interactivos.

## **Seguridad y Gobernanza de los Datos**

Para garantizar la seguridad y la conformidad con normativas, se implementarán las siguientes medidas:

- **Cifrado de Datos:** Tanto en tránsito como en reposo, utilizando KMS (Key Management Service) de AWS.
- **Control de Acceso Basado en Roles (RBAC):** Configuración de políticas detalladas en IAM (Identity and Access Management) para gestionar permisos.
- **Auditoría y Monitoreo:** Utilización cloudwatch para registrar todas las actividades e información necesaria.

### **Propuesta de Valor**

La implementación de esta solución tecnológica permitirá a Sutex SAS mejorar significativamente la eficiencia en la gestión de la información y optimizar la toma de decisiones. Al integrar los datos de diferentes sistemas en un Data Lake, la empresa podrá:

- Eliminar la fragmentación de la información y consolidar una fuente única de datos.
- Escalar sus operaciones tecnológicas sin preocuparse por la capacidad de almacenamiento.
- Mejorar la precisión de sus análisis y la capacidad de prever tendencias y comportamientos.
- Facilitar la adaptación a cambios en el mercado y la adopción de nuevas tecnologías en el futuro.

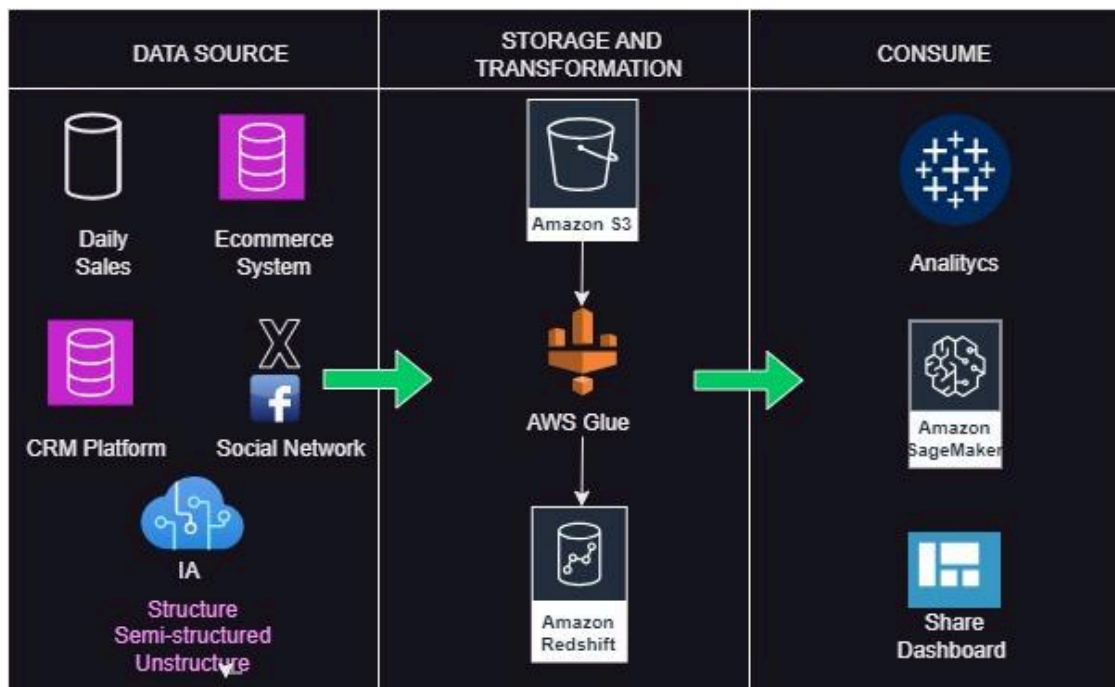
### **Características y Restricciones de la Solución**

- **Características:**
  - Escalabilidad: La solución se adapta al crecimiento de los datos y las necesidades de la empresa.
  - Acceso Controlado: Los datos están protegidos con políticas de acceso y encriptación.
  - Flexibilidad en el Procesamiento: Capacidad para manejar tanto análisis en tiempo real como en lotes.

- **Restricciones:**

- Dependencia de la Conectividad a Internet: La solución en la nube depende de una conexión estable.
- Costos de Almacenamiento y Procesamiento: El uso de servicios en la nube puede generar costos recurrentes que deben gestionarse adecuadamente.
- Complejidad en la Gestión Inicial: La configuración y despliegue del Data Lake puede requerir habilidades especializadas.

Esta propuesta detalla un plan integral para unificar la información de Sutex y aprovechar las tecnologías cloud e IoT más recientes para maximizar la eficiencia y la toma de decisiones estratégicas.



**Figura 8.** Árbol del problema. Estructura propuesta Solución Tecnológica.

Referencias

## 6. ANÁLISIS DEL PROCESO DE TRANSFORMACIÓN DIGITAL

Este proyecto de transformación digital no solo representa una solución técnica innovadora, sino que también evidencia la integración de competencias desarrolladas en la Especialización en Gestión de Servicios de Tecnologías de la Información. El enfoque multicloud propuesto no solo responde a las necesidades de Sutex, sino que también posiciona a la empresa como líder en la adopción de tecnologías avanzadas para la gestión y análisis de datos.

En este sentido, la transformación digital en Sutex implica más que la adopción de nuevas tecnologías. Se trata de un cambio en la manera en que la empresa opera y se conecta con sus empleados enfocados a la tecnología y su nivel de madurez en cuanto a ello, clientes y mercados. Este proceso no solo abarca la implementación de soluciones tecnológicas avanzadas, sino que también implica una renovación cultural que permita a Sutex adaptarse de manera ágil y efectiva a las demandas del entorno digital vistas en la clase de transformación digital.

- ***Cambio Cultural y Mentalidad Digital***

Para Sutex, esto significa pasar de una mentalidad centrada en procesos tradicionales y manuales a una mentalidad digitalmente integrada y basada en datos dado que es el foco de nuestra implementación. Se debe fomentar una cultura de innovación, colaboración y agilidad en todos los niveles de la empresa como hoy en día se puede ver en empresas como Google y algunos bancos en Latinoamérica. Los empleados deben ser capacitados para utilizar las herramientas digitales de manera eficiente, y se debe promover la adopción de nuevas formas de trabajo que favorezcan la cooperación y la toma de decisiones basada en datos.

- ***Automatización e Integración de Sistemas***

Para Sutex, es la integración de sus sistemas dispersos mediante la implementación de una arquitectura multicloud. Esta infraestructura permitirá consolidar datos de diversas fuentes y mejorar la accesibilidad y el análisis de la información en tiempo real por lo cual nuestra arquitectura soñada debe implementar la automatización de procesos, como la recopilación y procesamiento de datos, lo que se traducirá en una toma de decisiones más precisa y rápida.

- ***Uso de Análisis de Datos y Business Intelligence***

Para Sutex, el análisis de datos será otro pilar fundamental en la transformación digital. Ya que la implementación de herramientas de Business Intelligence y análisis avanzados que se dejaron en la arquitectura, permitirá a la empresa obtener insights valiosos sobre sus operaciones y el comportamiento del mercado. Proporcionarán una base sólida para la toma de decisiones estratégicas, ayudando a la empresa a adaptarse de manera más rápida.

Este proceso de transformación digital no solo es implementar nuevas tecnologías, sino de un cambio cultural que permita a Sutex ser más ágil, innovadora y eficiente en todos sus ámbitos.

## **7. ASPECTOS LEGALES Y CONTRATACIÓN**

El proceso de implementación de un Data Lake tiene una serie de consideraciones legales y contractuales importantes para garantizar el éxito y cumplimiento normativo, a continuación se nombran los aspectos claves:

- Asegurar el cumplimiento de la ley de protección de datos 1581 de 2012 esta norma protege el derecho de las personas a conocer, actualizar y rectificar la información que se tenga sobre ellas en bases de datos o archivos.
- En caso de almacenar información personal obtener el consentimiento explícito de los usuarios para el tratamiento de sus datos.
- Implementar medidas de seguridad técnicas y organizativas para garantizar la seguridad de los datos y evitar accesos no autorizados.
- Asegurar el cumplimiento de la Ley 23 de 1982 para regular los derechos de autor en Colombia. Esta ley establece las normas para los derechos de autor y los derechos conexos, y define el alcance de la protección para las obras intelectuales.
- Asegurar que las licencias de uso de software y hardware sean compatibles con la implementación del Data Lake.

A nivel de contratación tenemos dos puntos importantes que se deben tener en cuenta para tener feliz término en la implementación:

- Establecer y asegurar contratos claros con los proveedores de servicios en la nube, software y hardware, incluyendo cláusulas sobre confidencialidad, responsabilidad y resolución de disputas.
- Actualizar los contratos de los empleados para incluir cláusulas sobre el manejo de datos confidenciales y la protección de la información de la empresa.

## REFERENCIAS

1. ¿Qué es un data lake? | Google Cloud. (s. f.). Google Cloud.  
<https://cloud.google.com/learn/what-is-a-data-lake?hl=es-419>.
2. Hernández-Leal, E. J., Duque-Méndez, N. D., & Moreno-Cadavid, J. (2017). Big Data: una exploración de investigaciones, tecnologías y casos de aplicación. Tomado de:  
<https://www.redalyc.org/journal/3442/344251476001/html/#:~:text=Los%20m>

[%C3%A9todos%20de%20an%C3%A1lisis%20de.la%20totalidad%20de%20la%20poblaci%C3%B3n.](#)

3. *¿Qué es un lago de datos? - Introducción a los lagos de datos y los análisis - AWS.* (s. f.). Amazon Web Services, Inc.  
<https://aws.amazon.com/es/what-is/data-lake/>
4. KYOCERA Document Solutions. (2020, 24 marzo). 5 tipos de tecnología que pueden mejorar los resultados de tu empresa | Kyocera. *KYOCERA Document Solutions.*  
<https://www.kyoceradocumentsolutions.es/es/smarter-workspaces/business-challenges/paperless/5-tipos-de-tecnologia-que-pueden-mejorar-los-resultados-de-tu-empresa.html>
5. Tozzi, C. (2019, 25 noviembre). *Opciones de IaaS vs. PaaS en AWS, Azure y Google Cloud Platform.* *ComputerWeekly.es.*  
<https://www.computerweekly.com/es/consejo/Opciones-de-IaaS-vs-PaaS-en-AWS-Azure-y-Google-Cloud-Platform>
6. *Blockchain, bitcoin y criptomonedas.* (s. f.). Google Books.  
<https://books.google.es/books?hl=es&lr&id=f7SIDwAAQBAJ&oi=fnd&pg=PA11&dq=que+tecnologia+deberia+usarse+en+una+empresa+documentos&ots=L6sQQ1nfZf&sig=O4f5mh1eDgTnemRPybo6bI0GEB4#v=onepage&q&f=false>
7. Nambiar, Athira, and Divyansh Mundra. 2022. "An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management." *Big Data and Cognitive Computing* 5, no. 2 (Marzo): 230.
8. Aplicaciones de Data Lake y Data Warehouse en Big Data. (Marzo de 2023). *Revista de Tecnología y Gestión*, 5(2), 123-130. Recuperado el 3 de junio de 2024.
9. G Colaborativo. (s.f.).gruposim.eu. (2024). [www/gruposim.eu](http://www.gruposim.eu).iberdrola. (2023). [www.iberdrola.com](http://www.iberdrola.com).
10. PowerData. (2024). *PowerData*. Obtenido de Data Lake: definición, conceptos clave y mejores prácticas: <https://www.powerdata.es/data-lake>  
SAP. (2023). [www.sap.com](http://www.sap.com).

11. Estrella, R. (2013). Gerencia de tecnologías emergentes. Boletín Científico de las Ciencias Económico Administrativas del ICEA, 1(2).
12. ¿Qué es un data lake? | SAP. (s. f.-b). SAP.  
<https://www.sap.com/latinamerica/products/technology-platform/hana/what-is-a-data-lake.html>
13. Ramos-Rivadeneira, D. X., & Jiménez-Toledo, J. A. (2024). La innovación desde las tecnologías emergentes para la competitividad empresarial. *Gestión y Desarrollo Libre*, 9(17).
14. *Integración en la nube - Blog Incentro*. (2023, 24 julio). Integración En la Nube - Blog Incentro. <https://www.incentro.com/es-ES/blog/integracion-en-la-nube>
15. Slotnisky, D. (2016). *Transformación digital: cómo las empresas y los profesionales deben adaptarse a esta revolución*. Digital House. Coding School..
16. Alonso Guerra, I. (2017). La transformación digital de la empresa.
17. Townsend Valencia, J., & Figueroa Filián, J. (2022). Los modelos de transformación digital en la gestión de las empresas comerciales. *Cooperativismo y Desarrollo*, 10(2), 407-429.
18. Rojas, F. R. F. (2019). Análisis empírico de la transformación digital en las organizaciones. *International Journal of Information Systems and Software Engineering for Big Companies*, 6(1), 35-52.
19. Tagarro Martí, J. M. (2019). Desarrollo de una solución business intelligence mediante un paradigma de data lake.
20. Balón De La Cruz, M. L. (2022). Propuesta de una guía de adopción en un Data Lake enmarcando los aspectos importantes de Cloud Computing. Caso de estudio: Federación de Organizaciones Populares Peninsulares FOPOPE (Bachelor's thesis, La Libertad: Universidad Estatal Península de Santa Elena. 2022).
21. Beltrán García, V. E. (2022). Implementación de un Data Lake de los datos de uso del LMS Canvas de la Universidad Internacional SEK.
22. DeSL. (2023, 12 enero). Digital Transformation Fashion Industry - Download the White Paper.  
[https://www.desl.net/fashion-retail-apparel-footwear-digital-transformation/?utm\\_medium=adwords&utm\\_campaign=dynamic&utm\\_source=ad\\_group\\_1&utm\\_content=dynamic&gad\\_source=1&gclid=Cj0KCQiAr7C6BhDRARIsAOUKifj](https://www.desl.net/fashion-retail-apparel-footwear-digital-transformation/?utm_medium=adwords&utm_campaign=dynamic&utm_source=ad_group_1&utm_content=dynamic&gad_source=1&gclid=Cj0KCQiAr7C6BhDRARIsAOUKifj)

[V0ao0TRC0BOZQuFLVKxg4KpxxTqcMcl2Vrm3Uu71qNFqETzKpcucaAqp7EALw\\_wcB](https://www.mastercard.com/colombia/transformacion-digital)

23. Mastercard. (2024, 17 octubre). Digital labs. Mastercard Data & Services.  
[https://www.mastercard.com/en/capabilities/digital-labs?campaign\\_id=701UH000009uJvdYAE&channel=sep&cmp=2024.q2.bau-sem-digital-labs.digital%20transformation%20strategy&keyword=digital%20transformation%20strategy&gad\\_source=1&gclid=Cj0KCQiAr7C6BhDRARIsAOUKifhnDwxcFYgvdhjHi7TQpARZ3WebE7Np87x6aFSS1THS-mY2jhXcQT4aAo7IEALw\\_wcB](https://www.mastercard.com/en/capabilities/digital-labs?campaign_id=701UH000009uJvdYAE&channel=sep&cmp=2024.q2.bau-sem-digital-labs.digital%20transformation%20strategy&keyword=digital%20transformation%20strategy&gad_source=1&gclid=Cj0KCQiAr7C6BhDRARIsAOUKifhnDwxcFYgvdhjHi7TQpARZ3WebE7Np87x6aFSS1THS-mY2jhXcQT4aAo7IEALw_wcB)
24. Zúñiga, T., & Francini, K. (2021). Diseñar un data lake en aws lake formation para mejorar el almacenamiento y el acceso de los datos (Doctoral dissertation, Universidad Cenfotec).
25. Landa Felices, R. A., & Lujan Canicela, O. J. Integración de servicios de AWS enfocado al uso de DevSecOps para mejorar el rendimiento y la seguridad en proyectos de Data & Analytics.
26. El Manual de Estrategias para la Omnicanalidad. (s. f.). Commercetools.  
[https://commercetools.com/es/resources/whitepaper/el-manual-de-estrategias-para-la-omnicanalidad?utm\\_source=Paid-Search&utm\\_medium=Google-Search&utm\\_content=Spanish-omnichannel&utm\\_campaign=composable%20AMER-LATAM&utm\\_term=comercio%20electr%C3%B3nico&utm\\_campaign=composable%202024%20AMER-LATAM&utm\\_source=adwords&utm\\_medium=ppc&hsa\\_acc=9131799125&hsa\\_cam=21883528254&hsa\\_grp=170881592835&hsa\\_ad=720583443266&hsa\\_src=g&hsa\\_tgt=kwd-16650341&hsa\\_kw=comercio%20electr%C3%B3nico&hsa\\_mt=b&hsa\\_net=adwords&hsa\\_ver=3&gad\\_source=1&gbraid=0AAAAADRrS8cKQ23Sk2rIfSJhwsORa8e3e&gclid=Cj0KCQiAr7C6BhDRARIsAOUKifj1GRv-GP7GyKiX2VsB7oc0J5hxnLU9KU9JMojqwFlzCrdii1ZD8EaAtfBEALw\\_wcB](https://commercetools.com/es/resources/whitepaper/el-manual-de-estrategias-para-la-omnicanalidad?utm_source=Paid-Search&utm_medium=Google-Search&utm_content=Spanish-omnichannel&utm_campaign=composable%20AMER-LATAM&utm_term=comercio%20electr%C3%B3nico&utm_campaign=composable%202024%20AMER-LATAM&utm_source=adwords&utm_medium=ppc&hsa_acc=9131799125&hsa_cam=21883528254&hsa_grp=170881592835&hsa_ad=720583443266&hsa_src=g&hsa_tgt=kwd-16650341&hsa_kw=comercio%20electr%C3%B3nico&hsa_mt=b&hsa_net=adwords&hsa_ver=3&gad_source=1&gbraid=0AAAAADRrS8cKQ23Sk2rIfSJhwsORa8e3e&gclid=Cj0KCQiAr7C6BhDRARIsAOUKifj1GRv-GP7GyKiX2VsB7oc0J5hxnLU9KU9JMojqwFlzCrdii1ZD8EaAtfBEALw_wcB)
27. Hurtado, J. S. (2024, 19 octubre). ¿Cuáles son los modelos de negocios digitales más utilizados? Thinking for Innovation.  
<https://www.iebschool.com/blog/modelos-negocios-digitales-mas-utilizados-digital-business/>
28. Ley 1581 de 2012 - Gestor Normativo. (s. f.). Función Pública.  
<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=49981>

29. Los modelos de negocios digitales más utilizados. (2024, 28 junio). EAE Barcelona Business School.  
<https://www.eaebarcelona.com/es/blog/modelo-de-negocios>
30. Los modelos de negocios digitales más utilizados. (2024, 28 junio). EAE Barcelona Business School.  
<https://www.eaebarcelona.com/es/blog/modelo-de-negocios>
31. Política de Protección de Datos Personales -. (2024, 30 septiembre).  
<https://www.minambiente.gov.co/politica-de-proteccion-de-datos-personales/>
32. Ghoshal, A. (2023). Por qué Microsoft está combinando todos sus productos de análisis de datos en Fabric. Computerworld Spain, NA-NA.