UNIVERSIDAD DE EL SALVADOR

FACULTAD DE INGENIERÍA Y ARQUITECTURA

ESCUELA DE INGENIERÍA DE SISTEMAS INFORMÁTICOS

CURSO DE ESPECIALIZACIÓN INGENIERÍA DE DATOS

ANÁLISIS DE INFORMACIÓN SOBRE MANEJO DE INVENTARIO DE LA EMPRESA *IMPORTACIONES EL ÁNGEL* GESTIONADA A TRAVÉS DEL SISTEMA MAGENTO COMMERCE

PRESENTADO POR:

GUSTAVO ERNESTO HENRÍQUEZ CORVERA

CRISTIAN VIDAL GARCÍA MEJÍA

LEONEL ENRIQUE RAMOS RAMOS

PARA OPTAR AL TITULO DE:

INGENIERO DE SISTEMAS INFORMÁTICOS

CIUDAD UNIVERSITARIA, DICIEMBRE 2023

UNIVERSIDAD DE EL SALVADOR

RECTOR:

MSC. JUAN ROSA QUINTANILLA

SECRETARIO GENERAL:

LIC. PEDRO ROSARIO ESCOBAR CASTANEDA

FACULTAD DE INGENIERIA Y ARQUITECTURA

DECANO:

ING. LUIS SALVADOR BARRERA MANCÍA

ESCUELA DE INGENIERIA DE SISTEMAS INFORMÁTICOS

DIRECTOR:

ING. CESAR AUGUSTO GONZÁLEZ

UNIVERSIDAD DE EL SALVADOR

FACULTAD DE INGENIERIA Y ARQUITECTURA

ESCUELA DE INGENIERIA DE SISTEMAS INFORMÁTICOS

Curso de Especialización previo a la opción al Grado de:

INGENIERO DE SISTEMAS INFORMÁTICOS

Titulo:

CURSO DE ESPECIALIZACIÓN INGENIERÍA DE DATOS

ANÁLISIS DE INFORMACIÓN SOBRE MANEJO DE INVENTARIO DE LA EMPRESA IMPORTACIONES EL ÁNGEL GESTIONADA A TRAVÉS DEL SISTEMA MAGENTO COMMERCE

Presentado por:

GUSTAVO ERNESTO HENRÍQUEZ CORVERA

CRISTIAN VIDAL GARCÍA MEJÍA

LEONEL ENRIQUE RAMOS RAMOS

Curso de Especialización aprobado por:

Docente Asesor:

ING. RENE FABRICIO QUINTANILLA GOMEZ

SAN SALVADOR, DICIEMBRE 2023

SAN SALVADOR, DICIEMBRE 2023

Trabajo de Graduación aprobado por:

Docente Asesor:

ING. RENE FABRICIO QUINTANILLA GOMEZ

INDICE

Introducción 6

Capítulo I: Especificación del proyecto 7

a. Situación actual 7

I. Antecedentes 7
II. Descripción del problema 7
III. Planteamiento del problema 7
i. Objetivos 8
Objetivos generales 8
Objetivos específicos 8
ii. Alcances 8
iii. Justificación 9
iv. Marco teórico 10
v. Cronograma de actividades 16
vi. Presupuesto 16

Capitulo II: Análisis y diseño de la propuesta de la solución 17

a. Metodología de trabajo 17

Introducción a la lógica del negocio 17
Descripción del Dataset 18
Diccionario de datos del Dataset 29
Resultado de data Profiling 33
Modelo dimensional propuesto 38
Mapping por tabla 39

b. Descripción de la propuesta de solución 43

c. Especificaciones de necesidades analíticas 44

d. Descripción de la tecnología a utilizar 44

e. Diagrama arquitectónico de la solución 47

f. Descripción de cada componente de la solución 47

Capitulo III: Estrategia de implementación de propuesta de solución 49

a. Estrategia de implementación 49

b. Presupuesto de implementación 57

c. Análisis de resultados 59

Conclusiones 62

Recomendaciones 62

Bibliografía 63

Glosario 63

Anexos 65

Introducción

En el ámbito empresarial, saber gestionar el inventario de manera eficiente es clave para que las organizaciones prosperen a largo plazo. En este contexto, el presente proyecto se embarca en un profundo análisis del manejo de inventario de Importaciones El Ángel, una empresa dedicada a la importación y comercialización de productos decorativos para el hogar, cuya operación se centra en una tienda en línea con presencia exclusiva en El Salvador.

El proyecto aborda la creación de un modelo multidimensional sólido que sirva de base para el análisis detallado del proceso de negocio seleccionado (estratégico, operativo, control). Esta fase implica la implementación del modelo dimensional en el entorno robusto y escalable de Amazon Web Services (AWS), con un enfoque en los servicios Redshift y S3.

Como herramienta de elección para la visualización y generación de informes, Power BI se posiciona como una plataforma potente que permitirá la creación de Dashboards intuitivos y análisis detallados para informar decisiones estratégicas. La implementación de Fact Table, como parte esencial del modelo dimensional, proporcionará una estructura que permitirá evaluar el rendimiento del inventario y tomar decisiones basadas en datos.

En resumen, este proyecto fusiona tecnología innovadora, metodologías analíticas y estrategias de negocio para ofrecer a Importaciones El Ángel una perspectiva profunda y valiosa sobre su gestión de inventario. A lo largo de dos fases, desde el diseño conceptual hasta la implementación práctica.

Capítulo I: Especificación del proyecto

a. Situación actual

I. Antecedentes

Con base a estudios realizados de empresas del mismo rubro, se centró en las complejidades experimentadas ante la gestión de un inventario diversificado en constante evolución. Enfocándose en la expansión de su oferta de productos, se determinó la necesidad de una solución analítica avanzada para optimizar la gestión de inventario.

II. Descripción del problema

La empresa se enfrenta a desafíos significativos en su capacidad para anticipar y adaptarse de manera efectiva a las fluctuaciones del mercado. La falta de herramientas analíticas robustas dificulta la comprensión profunda de las dinámicas del mercado y la rápida respuesta a las cambiantes demandas de los clientes. Esta carencia ha resultado en dificultades para tomar decisiones estratégicas informadas y ha limitado la agilidad de la empresa para ajustar sus operaciones de manera eficiente.

La ausencia de un sistema analítico sofisticado ha llevado a una respuesta reactiva en lugar de proactiva ante los cambios del mercado, lo que ha afectado la capacidad de la empresa para mantenerse competitiva. Este proyecto surge como una respuesta estratégica para superar estas limitaciones mediante la implementación de una herramienta analítica avanzada. La meta es proporcionar a la empresa una visión más clara y profunda de sus operaciones, permitiéndole tomar decisiones fundamentadas y ágiles para mantenerse adaptable y competitiva en un entorno empresarial dinámico.

III. Planteamiento del problema

En el entorno empresarial actual, Importaciones El Ángel se enfrenta a desafíos sustanciales relacionados con la gestión efectiva de su inventario y la capacidad de adaptación a las dinámicas cambiantes del mercado. La falta de herramientas analíticas avanzadas ha generado limitaciones en la capacidad de la empresa para comprender a fondo las complejidades de su inventario, anticipar las variaciones en la demanda y tomar decisiones estratégicas.

La ausencia de un sistema integral de análisis ha llevado a una respuesta reactiva ante las fluctuaciones del mercado, lo que impacta directamente en la eficiencia operativa y la capacidad competitiva de Importaciones El Ángel. La falta de visibilidad en tiempo real sobre el estado del inventario, combinada con la dificultad para analizar tendencias y patrones de compra, ha resultado en un manejo óptimo de recursos y en la pérdida de oportunidades de optimización.

Este escenario problemático crea la necesidad de realizar una solución que no solo apunte a superar las limitaciones actuales de gestión de inventario, sino también proporcionar una actitud proactiva a las condiciones del mercado, una mejora continua de sus operaciones y el fortalecimiento de su posición competitiva en el sector.

i. Objetivos

Objetivos generales

El propósito fundamental de este proyecto es diseñar e implementar un modelo dimensional específico, con la finalidad de proporcionar a Importaciones El Ángel una visión más detallada y precisa de su inventario. Este modelo aspira a ser una herramienta fundamental para la toma de decisiones estratégicas y operativas.

Objetivos específicos

- Analizar fuentes de datos proporcionada por la empresa.
- Realizar perfilado de datos.
- Diseñar modelo dimensional del proceso de negocio.
- Desarrollar procesos ETL personalizados del proceso de negocio.
- Utilizar herramientas para integrar el modelo dimensional.
- Diseñar un panel interactivo para el análisis de datos.

ii. Alcances

Para el proyecto de desarrollo e implementación de un modelo dimensional en el ámbito de la gestión de inventario, se proponen los siguientes alcances:

- <u>Análisis de datos:</u>

Recolectar datos proporcionados por la empresa, para el exhaustivo análisis.

- <u>Diseño de Esquema Estrella:</u>

Desarrollar un esquema estrella del modelo dimensional que responda a las necesidades analíticas actuales del proceso de gestión de inventario, proporcionando una estructura que pueda adaptarse a futuros requerimientos analíticos.

- Implementación de Data Lake para Inventarios:

Establecer una solución de Data Lake en AWS S3 para almacenar y gestionar datos generados en el proceso de inventario, brindando soporte al proceso de toma de decisiones basadas en datos internos, se destacará la integración de herramientas como Redshift.
El proyecto se centrará en el diseño e implementación de un modelo dimensional, aprovechando servicios de AWS con un enfoque específico en Redshift para el almacenamiento de datos.

- Desarrollo de Dashboards Analíticos:

Crear Dashboards que aborden las necesidades analíticas del proyecto de inventario, permitiendo la visualización y análisis de datos para orientar estrategias a corto, mediano y largo plazo en la gestión de inventarios, (agregar POWER BI)

- Repositorio en GitHub con Pasos Técnicos:

Objetivo Específico: Mantener un repositorio en GitHub que contenga los pasos técnicos seguidos en las cuatro etapas del proyecto de inventario: perfilado de datos, diseño dimensional, desarrollo e implementación, y resultados obtenidos. Facilitará la revisión y replicación del proceso.

iii. Justificación

La pandemia global ha traído consigo cambios significativos en el comportamiento de los consumidores y en la operación de los negocios, especialmente en el sector del comercio electrónico. "Importaciones El Ángel", al operar exclusivamente en línea en El Salvador, enfrenta desafíos únicos en la gestión de su inventario, acentuados por las fluctuaciones en la demanda y los cambios en las preferencias de los consumidores. La eficiencia y la precisión en la gestión del inventario no son solo deseables, sino cruciales para mantener la competitividad y satisfacer las expectativas de los clientes.

Este proyecto busca abordar estos desafíos utilizando un modelo de análisis de inventario avanzado, implementado mediante la integración de tecnologías y herramientas de vanguardia. La utilización de Amazon Web Services, específicamente Amazon Redshift y AWS S3, proporcionará una infraestructura de almacenamiento y análisis de datos robusta y escalable, capaz de manejar el volumen y la complejidad de los datos de inventario generados por la plataforma Magento Commerce.

La inclusión de Power BI en nuestro enfoque permitirá la visualización y el análisis detallado de estos datos, proporcionando perspectivas claras y accionables para la toma de decisiones estratégicas. Además, la utilización de herramientas como Talend para el proceso de ETL (Extract, Transform, Load) facilitará la integración y transformación eficiente de los datos, asegurando que la información analizada sea precisa y relevante.

Este análisis detallado y en tiempo real del inventario permitirá a "Importaciones El Ángel" responder rápidamente a las cambiantes demandas del mercado, optimizar los niveles de stock, reducir los costos de almacenamiento y mejorar la satisfacción del cliente. Además, el proyecto

tiene el potencial de revelar patrones y tendencias en el comportamiento del consumidor, lo que podría influir en futuras decisiones de compra y estrategias de marketing.

En resumen, la implementación de esta solución de big data no solo es una respuesta a las necesidades inmediatas de "Importaciones El Ángel", sino que también establece una base para una gestión de inventario más inteligente y basada en datos en el futuro. Esta solución proporcionará a la empresa las herramientas necesarias para adaptarse y prosperar en el dinámico entorno del comercio electrónico moderno.

iv. Marco teórico

1. Comercio Electrónico y gestión de inventario

Importancia de la Gestión de Inventario en el Comercio Electrónico:

- El comercio electrónico requiere una gestión de inventarios efectiva para asegurar el éxito del negocio. Esto implica no solo mantener un stock óptimo sino también entender qué productos tienen mayor o menor demanda. Una buena gestión de inventario ayuda a realizar pedidos de manera eficiente y evita la sobrecarga de inventarios no deseados.
- Una gestión eficiente de inventarios asegura que los productos estén disponibles para los clientes, lo que se traduce en mayores ventas y satisfacción del cliente. Por otro lado, una mala gestión puede tener consecuencias negativas, como pérdida de ventas y insatisfacción de los clientes.

Consecuencias de una Gestión de Inventarios Deficiente:

- Una gestión inadecuada de inventarios puede generar costos adicionales, como almacenamiento de exceso de producto, pérdida de productos por obsolescencia, y necesidad de contar y reordenar inventario nuevo. Además, puede llevar a errores en la gestión de pedidos, afectando la entrega de productos y la satisfacción del cliente.

Integración de la Gestión de Inventarios con la Gestión de Pedidos:

- La integración de la gestión de inventarios con la gestión de pedidos es clave para el éxito en el comercio electrónico. Esto implica que la información del inventario disponible coincida con los pedidos realizados en tiempo real. El uso de software especializado permite una comunicación constante entre ambas áreas, reduciendo errores en la gestión de pedidos y optimizando el espacio de almacenamiento, lo que lleva a una mayor eficiencia en la gestión de la cadena de suministro.

2. Magento Commerce como plataforma de E-Commerce

Definición y Función de un Data Warehouse:

- Un Data Warehouse es un sistema de almacenamiento de datos estructurados que sirve como la única fuente de verdad en una organización y es esencial para consolidar información de diversas áreas, como ventas, operaciones y atención al cliente.
- Los Data Warehouses mantienen datos históricos y de calidad, los cuales son fundamentales para el análisis retrospectivo y la toma de decisiones basada en hechos pasados. Estos datos suelen estar agregados o segmentados para facilitar su análisis o reporte.

Data Warehouse en la Nube:

- La tendencia actual mueve los Data Warehouses hacia infraestructuras en la nube para aprovechar la reducción de costos operativos y de mantenimiento.
- Las tecnologías en la nube ofrecen elasticidad y escalabilidad, permitiendo a las organizaciones ajustar recursos de forma rápida y eficiente según la demanda, y aprovechar el procesamiento paralelo masivo (MPP) para acelerar las consultas y la transformación de datos.

Ejemplos de Data Warehouse en la Nube:

- Ejemplos de soluciones de Data Warehouse en la nube incluyen Amazon Redshift, Azure Synapse Analytics (anteriormente Azure SQL Data Warehouse) y Google BigQuery. Estas plataformas ofrecen capacidades específicas, como almacenamiento columnar, MPP, y lenguaje de consulta SQL, y están diseñadas para integrarse con otras herramientas de procesamiento de datos como Apache Spark.
- Estas soluciones son críticas en el comercio electrónico para manejar grandes volúmenes de transacciones y datos de clientes, y para proporcionar la infraestructura necesaria para análisis avanzados y la generación de insights accionables para la empresa.

3. Data lakes y big data

Fundamentos y Rol del Big Data en el Comercio Electrónico:

- Big Data se refiere a grandes volúmenes de datos almacenados y analizados para obtener información útil y relevante para las organizaciones. En el comercio electrónico, estos datos se utilizan para desarrollar estrategias de marketing y ventas, mejorar la atención al cliente y optimizar la logística de los productos.

Impacto del Big Data en las Estrategias Comerciales:

- El uso del Big Data en el comercio electrónico permite a las empresas estructurar bases de datos completas que incluyen desde información demográfica básica hasta detalles más específicos, como el comportamiento del cliente, sus preferencias y patrones de compra.
- Este enfoque ayuda a las empresas de comercio electrónico a obtener ventajas competitivas, al proporcionarles conocimientos detallados sobre sus clientes y el mercado, y permitiéndoles optimizar sus estrategias de ventas y personalización.

Rastreo y Análisis de Datos en el Ambiente Digital:

- En un mercado digital cada vez más competitivo, la capacidad de rastrear y analizar datos es crucial para guiar las decisiones comerciales, reducir riesgos y tomar acciones estratégicas.
- El seguimiento preciso del comportamiento del consumidor es fundamental para entender sus necesidades y preferencias, lo cual se traduce en insights valiosos para el negocio.

Identificación de Puntos de Mejora y Estrategias de Personalización:

- El Big Data ayuda a identificar qué necesita mejorar en los procesos de comercio electrónico y cómo hacerlo, utilizando datos concretos en lugar de suposiciones.
- También favorece la personalización de la experiencia del usuario, crucial en la era actual donde los consumidores valoran no solo el producto, sino toda la experiencia de compra.

Predicción de Tendencias y Demanda:

- Con herramientas de predicción potentes, el Big Data permite anticipar tendencias y patrones de consumo, lo cual es esencial para preparar estrategias de marketing y ventas eficaces para productos con demanda creciente.

Posicionamiento de la Marca y Optimización de la Distribución:

- Las empresas pueden utilizar el Big Data para desarrollar acciones de marketing digital más eficientes y segmentadas, contribuyendo a un posicionamiento de marca más fuerte en internet.
- Además, el Big Data juega un papel clave en la optimización de la distribución de productos y el almacenaje, permitiendo a los comercios electrónicos identificar períodos de alta demanda y planificar una distribución más eficiente.

4. Modelado Dimensional

Definición de Objetivos y Requerimientos de Información:

- En esta fase inicial, se definen los objetivos específicos del Data Warehouse y se identifican los requerimientos de información. Es crucial comprender las necesidades del negocio y cómo el Data Warehouse las abordará.

Diseño y Modelización:

- Se identifican las fuentes de datos y se realizan las transformaciones necesarias para construir el modelo lógico del Data Warehouse. Este modelo incluye entidades y relaciones que reflejan las necesidades del negocio. Luego, se traduce este modelo lógico en un modelo físico que define la arquitectura de almacenamiento del Data Warehouse.

Implementación:

- La implementación del Data Warehouse incluye la extracción y transformación de datos del sistema operacional, la carga de datos en el Data Warehouse. Los metadatos juegan un papel esencial en esta fase, almacenando información crítica sobre los datos y cómo se usan
.

Revisión y Mejora Continua:

- Después de la implementación, se realiza una revisión del Data Warehouse para identificar áreas de mejora y potenciación. Esta fase es iterativa y se basa en el aprendizaje continuo de las experiencias anteriores.

Formación y Capacitación de Usuarios:

- Se diseñan cursos de formación personalizados para los usuarios del Data Warehouse. Estos cursos buscan proporcionar la formación necesaria para el mejor aprovechamiento de las funcionalidades del sistema y asegurar que los conceptos clave sean bien entendidos y aplicados por los usuarios.

5. Procesos ETL y Talend Open Studio

Fundamentos de ETL (Extract, Transform, Load):

- ETL es un proceso clave en la gestión de datos que implica extraer datos de varias fuentes, transformarlos para cumplir con los requisitos del negocio, y cargarlos en un sistema de destino como un Data Warehouse.
- Este proceso es esencial para asegurar que los datos sean consistentes, limpios y útiles para análisis.

Talend Open Studio para la Integración de Datos:

- Talend Open Studio es una herramienta de integración de datos de código abierto que facilita la creación y gestión de flujos de trabajo ETL.
- Ofrece una interfaz gráfica de usuario para diseñar procesos de ETL sin necesidad de escribir código extenso, lo que agiliza y simplifica la integración de datos.

Aplicaciones de ETL en Comercio Electrónico:

- En el comercio electrónico, los procesos ETL son fundamentales para consolidar datos de ventas, inventario y comportamiento del cliente de múltiples plataformas en un único repositorio.
- Esto permite realizar análisis detallados y generar insights para mejorar la experiencia del cliente, optimizar la gestión del inventario y aumentar las ventas.

Talend Open Studio en la Práctica:

- Descripción de cómo Talend se utiliza para conectar con diversas fuentes de datos, transformar los datos según las reglas de negocio y cargarlos en sistemas como Amazon Redshift o bases de datos locales.
- Importancia de Talend en la automatización de procesos ETL, asegurando eficiencia y precisión en la gestión de datos.

6. AWS para análisis de datos

Amazon S3: Almacenamiento en la Nube:

- Amazon S3 es un servicio de almacenamiento en la nube creado hace más de 12 años, diseñado para almacenar datos y permitir su recuperación desde cualquier ubicación.
- Se define como un servicio multipropósito, que facilita el almacenamiento y recuperación de datos de cualquier volumen, siendo ideal para respaldar sitios web estáticos, imágenes, y otros archivos.

Beneficios de Amazon S3 para Empresas:

- Amazon S3 ofrece beneficios significativos, especialmente para pequeñas y medianas empresas, al reducir la necesidad de inversiones en infraestructura informática y proporcionar opciones de transferencia de datos fiables y sencillas.
- Destaca por su alta seguridad, garantizando la integridad y confidencialidad de los datos almacenados, y proporcionando una disponibilidad del 99,99%.

Costos y Escalabilidad de Amazon S3:

- Uno de los atractivos de Amazon S3 es su estructura de costos, que elimina la necesidad de inversiones iniciales en infraestructura de TI y reduce los costos fijos. Además, ofrece una escalabilidad que permite ajustar rápidamente el volumen de datos almacenados según las necesidades.

Integración y Distribución de Contenido con Amazon S3:

- Amazon S3 facilita la integración de datos almacenados con aplicaciones de terceros, mejorando el rendimiento de aplicaciones móviles y web. Además, puede ser utilizado para crear redes de distribución de contenido (CDN), mejorando el rendimiento de las aplicaciones al reducir los tiempos de carga.

7. Visualización de datos y Power BI

Power BI: Herramienta de Visualización de Datos e Inteligencia Empresarial:

- Power BI es una solución de análisis empresarial que permite visualizar datos y compartir conocimientos en toda la organización o integrarlos en aplicaciones o sitios web. Conectándose a cientos de fuentes de datos, Power BI facilita la creación de paneles e informes en vivo.

Beneficios de Power BI para Empresas:

- Power BI es una plataforma de análisis de negocios basada en la nube que brinda a los usuarios la capacidad de crear informes y paneles de control. Los beneficios clave incluyen un costo inicial mínimo, con la versión de escritorio (Power BI Desktop) gratuita para todos los usuarios, y la posibilidad de acceder a datos en cualquier lugar y en cualquier momento mediante su aplicación móvil.

Obtención de Datos de Múltiples Fuentes:

- Power BI permite conectar más de 70 fuentes de datos, como SQL Server, Azure DB, SharePoint, Excel y servicios en línea, consolidando los datos en un conjunto único para la creación de informes y paneles interactivos.

Actualización de Datos y Funcionalidades Avanzadas:

- La versión Pro de Power BI permite programar actualizaciones de datos diarias o cada hora, garantizando que las organizaciones tengan acceso a la información más reciente para tomar decisiones comerciales rentables.

Integración de Power BI y Visualización Interactiva:

- La incrustación de Power BI integra elementos visuales completamente interactivos en aplicaciones personalizadas, utilizando Power BI REST y la API de JavaScript. Esto permite a las empresas personalizar cómo se presentan y analizan los datos en sus sistemas internos.

Q&A y Análisis de Datos en Lenguaje Natural:

- Power BI incluye una herramienta de consulta en lenguaje natural (Q&A) que permite a los usuarios hacer preguntas sobre sus datos y obtener resultados en forma de gráficos y tablas, facilitando el acceso y comprensión de la información clave.

8. Modelo de entidad-valor-atributo (EAV) en el comercio electrónico

Definición y Características del EAV:

- El EAV es un modelo de datos que se utiliza para diseñar esquemas de base de datos flexibles y capaces de almacenar información que varía ampliamente en cantidad de atributos.
- Es especialmente útil en escenarios donde la estructura de los datos no es fija o donde los atributos de las entidades pueden cambiar con el tiempo.

EAV en Magento y Comercio Electrónico:

- Magento, una plataforma popular de comercio electrónico, utiliza el modelo EAV para gestionar los datos de productos, clientes y pedidos. Esto permite una gran flexibilidad en términos de personalización de productos y categorías.
- En tu proyecto, el análisis de datos extraídos de Magento implica comprender y trabajar con este modelo de datos, lo cual puede presentar desafíos y oportunidades únicos en términos de modelado y análisis.

Importancia del EAV en la Gestión de Inventario:

- En el contexto de la gestión de inventario, el modelo EAV permite almacenar detalles variados y extensos de productos, lo cual es crucial para un análisis de inventario detallado y personalizado.
- Comprender cómo se estructuran y almacenan estos datos en Magento es esencial para diseñar un proceso de ETL eficaz y un modelo dimensional que refleje con precisión las necesidades de análisis de inventario.

Desafíos y Estrategias de Trabajo con EAV:

- Trabajar con un modelo EAV puede presentar desafíos en términos de rendimiento y complejidad de consulta. En tu marco teórico, sería útil abordar cómo estos desafíos se mitigarán o manejarán en el proyecto.
- Puedes incluir estrategias para transformar datos EAV en un formato más adecuado para el análisis, como la normalización o la conversión a un esquema más tradicional para su uso en herramientas de BI como Power BI.

v. *Cronograma de actividades*

El cronograma de actividades estará en el GitHub puesto en los anexos por las dimensiones de dicho recuadro.

vi. *Presupuesto*

PRESUPUESTO HIPOTÉTICO PARA EL PROYECTO DE "IMPORTACIONES EL ÁNGEL"

1.	***Servicios de Nube (AWS)***
	• **Amazon Redshift:**
	▪ Costo estimado: $0.25 por hora para una instancia dc2.large.
	▪ Uso estimado: 4 horas al día durante 6 meses (180 días).
	▪ Costo total: 4 horas/día * $0.25/hora * 180 días = $180.
	• **AWS S3**:
	▪ Costo estimado: $0.023 por GB al mes.
	▪ Almacenamiento estimado: 50 GB al mes durante 6 meses.
	▪ Costo total: 50 GB * $0.023/GB * 6 meses = $6.90.
2.	***Software***
	• **Power BI Pro:**
	▪ Costo de suscripción: $9.99 por usuario al mes.
	▪ Número de usuarios: 3.
	▪ Duración: 6 meses.
	▪ Costo total: $9.99/usuario/mes * 3 usuarios * 6 meses = $179.82.
3.	***Recursos Humanos***
	• **Desarrolladores y Analistas de Datos**:
	▪ Costo estimado: $25 por hora por persona.
	▪ Número de personas: 3.
	▪ Horas estimadas de trabajo: 10 horas a la semana por persona durante 6 meses (aproximadamente 240 horas por persona).
	▪ Costo total: $25/hora * 3 personas * 240 horas = $18,000.

COSTO TOTAL ESTIMADO DEL PROYECTO:

- Servicios de Nube (AWS): $186.90
- Software (Power BI Pro): $179.82
- Recursos Humanos: $18,000
- **Total General**: $18,366.72

Capitulo II: Análisis y diseño de la propuesta de la solución

a. Metodología de trabajo

Introducción a la lógica del negocio

El sistema Magento Commerce se utiliza para gestionar las operaciones de la tienda en línea de Importaciones El Ángel. El inventario, los clientes y otros aspectos clave del negocio. A través del análisis de estos datos, se busca obtener información valiosa para mejorar la gestión del inventario y la toma de decisiones estratégicas.

Se aplicará el modelado dimensional, conformado por el proceso de negocio, identificación de dimensiones y determinación de métricas que medirán los eventos que suceden en el proceso de negocio. En la primera etapa se cubrirá todo el proceso de profiling sobre el dataset de Inventario, el diccionario de datos para poder conocer tabla por tabla y su estructura.

Además, se evidenciará el resultado del modelado dimensional por medio del esquema estrella y finalmente se cubrirá el proceso de mappings por cada tabla que se utilizará para construir el modelo dimensional como tal.

El objetivo principal de este proyecto de análisis de inventario es utilizar el sistema Magento y las tablas de datos proporcionadas para desarrollar una solución analítica que permita una visión más precisa y completa de la gestión del inventario. Mediante el uso de técnicas de análisis y generación de informes, los niveles de inventario, la rotación de productos y la rentabilidad.

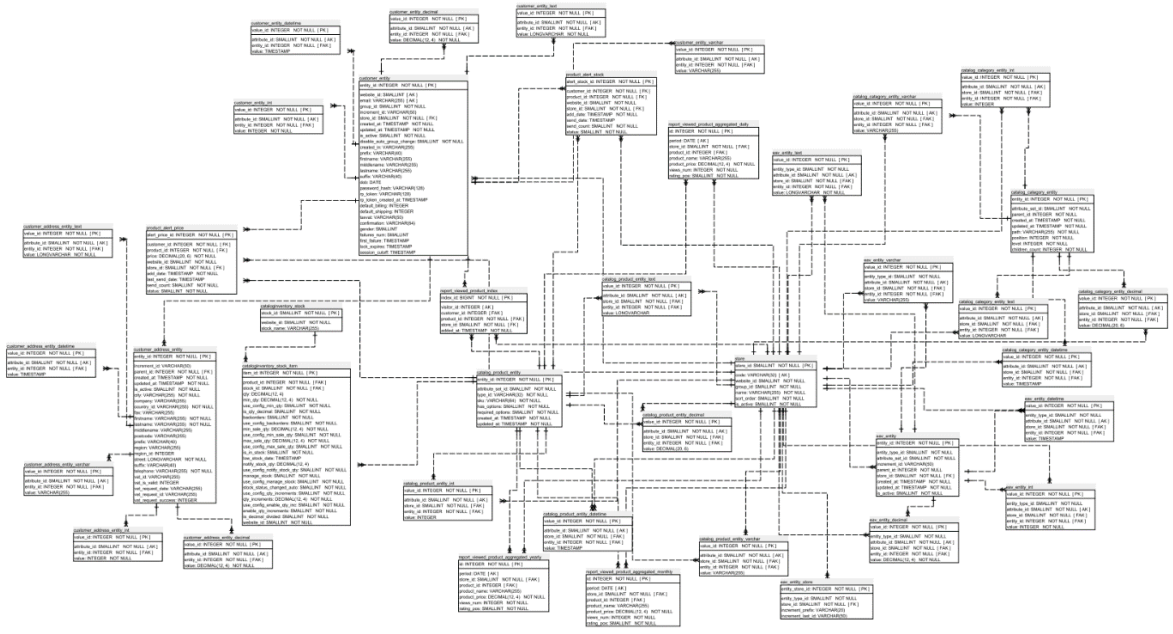
La solución propuesta permitirá a Importaciones El Ángel tomar decisiones más informadas y estratégicas en relación con su inventario. Con un análisis exhaustivo, la empresa podrá identificar patrones de demanda, gestionar adecuadamente los niveles de stock y mejorar la rentabilidad general del negocio.

Descripción del Dataset

A continuación, se detalla de manera concreta como los datos que conforman el dataset de Magento están organizados:

Origen de los datos:	Los datos provienen de un modelo de base de datos transaccional o relacional, almacenados en la plataforma de Magento
Formato:	Los datos están organizados en 348 tablas con formato csv, con encabezados en sus tablas, uso de separadores tales como [,]
Estructura:	Cada una de las tablas con las que cuenta el modelo relacional poseen sus correspondientes llaves primarias y foráneas por las cuales se relacionan. Teniendo consigo metada correspondiente al esquema que cada una posee, es decir sus respectivos tipos de datos para cada columna que la conforma
Periodo de tiempo del dataset:	

A continuación, presentaremos el correspondiente modelo relacional que constituye nuestra principal fuente de datos para la construcción del modelo dimensional:



Nota: Para poder visualizar el modelo relacional, puede hacerlo en el repositorio de Github: ***Magento.pdf***, ya que su diseño es muy grande y no podrá cumplir su función de una visualización clara y entendible.

A continuación, se listará las tablas que conforman Magento y una breve descripción de la selección de ellas para implementar nuestro Inventario y aplicación del modelo dimensional.

Nombre de tabla	Columnas	Utilizadas para MD
admin_analytics_usage_version_log	2	
admin_passwords	5	
admin_system_messages	3	
admin_user	19	
admin_user_expiration	2	
admin_user_session	7	
adminnotification_inbox	8	
authorization_role	8	
authorization_rule	5	
cache	5	
cache_tag	2	
captcha_log	4	
catalog_category_entity	9	Si
catalog_category_entity_datetime	5	Si
catalog_category_entity_decimal	5	Si
catalog_category_entity_int	5	Si
catalog_category_entity_text	5	Si
catalog_category_entity_varchar	5	Si
catalog_category_product	4	Si
catalog_category_product_index	6	
catalog_category_product_index_replica	6	
catalog_category_product_index_store1	6	
catalog_category_product_index_store1_replica	6	
catalog_category_product_index_tmp	6	
catalog_compare_item	6	
catalog_compare_list	3	
catalog_eav_attribute	24	
catalog_product_bundle_option	5	
catalog_product_bundle_option_value	5	
catalog_product_bundle_price_index	5	
catalog_product_bundle_selection	10	
catalog_product_bundle_selection_price	5	

catalog_product_bundle_stock_index	5	
catalog_product_entity	8	Si
catalog_product_entity_datetime	5	Si
catalog_product_entity_decimal	5	Si
catalog_product_entity_gallery	6	
catalog_product_entity_int	5	Si
catalog_product_entity_media_gallery	5	
catalog_product_entity_media_gallery_value	7	
catalog_product_entity_media_gallery_value_to_entity	2	
catalog_product_entity_media_gallery_value_video	7	
catalog_product_entity_text	5	Si
catalog_product_entity_tier_price	8	
catalog_product_entity_varchar	5	Si
catalog_product_frontend_action	6	
catalog_product_index_eav	5	
catalog_product_index_eav_decimal	5	
catalog_product_index_eav_decimal_idx	5	
catalog_product_index_eav_decimal_replica	5	
catalog_product_index_eav_decimal_tmp	5	
catalog_product_index_eav_idx	5	
catalog_product_index_eav_replica	5	
catalog_product_index_eav_tmp	5	
catalog_product_index_price	9	
catalog_product_index_price_bundle_idx	13	
catalog_product_index_price_bundle_opt_idx	9	
catalog_product_index_price_bundle_opt_tmp	9	
catalog_product_index_price_bundle_sel_idx	9	
catalog_product_index_price_bundle_sel_tmp	9	
catalog_product_index_price_bundle_tmp	13	
catalog_product_index_price_cfg_opt_agr_idx	6	
catalog_product_index_price_cfg_opt_agr_tmp	6	
catalog_product_index_price_cfg_opt_idx	6	
catalog_product_index_price_cfg_opt_tmp	6	
catalog_product_index_price_downlod_idx	5	
catalog_product_index_price_downlod_tmp	5	
catalog_product_index_price_final_idx	10	
catalog_product_index_price_final_tmp	10	
catalog_product_index_price_idx	9	

catalog_product_index_price_opt_agr_idx	7	
catalog_product_index_price_opt_agr_tmp	7	
catalog_product_index_price_opt_idx	6	
catalog_product_index_price_opt_tmp	6	
catalog_product_index_price_replica	9	
catalog_product_index_price_tmp	10	
catalog_product_index_tier_price	4	
catalog_product_index_website	4	
catalog_product_link	4	
catalog_product_link_attribute	4	
catalog_product_link_attribute_decimal	4	
catalog_product_link_attribute_int	4	
catalog_product_link_attribute_varchar	4	
catalog_product_link_type	2	
catalog_product_option	10	
catalog_product_option_price	5	
catalog_product_option_title	4	
catalog_product_option_type_price	5	
catalog_product_option_type_title	4	
catalog_product_option_type_value	4	
catalog_product_relation	2	
catalog_product_super_attribute	4	
catalog_product_super_attribute_label	5	
catalog_product_super_link	3	
catalog_product_website	2	
catalog_url_rewrite_product_category	3	
cataloginventory_stock	3	Si
cataloginventory_stock_item	26	Si
cataloginventory_stock_status	5	
cataloginventory_stock_status_idx	5	
cataloginventory_stock_status_replica	5	
cataloginventory_stock_status_tmp	5	
catalogrule	12	
catalogrule_customer_group	2	
catalogrule_group_website	3	
catalogrule_group_website_replica	3	
catalogrule_product	11	
catalogrule_product_price	8	

catalogrule_product_price_replica	8	
catalogrule_product_replica	11	
catalogrule_website	2	
catalogsearch_recommendations	3	
checkout_agreement	8	
checkout_agreement_store	2	
cms_block	7	
cms_block_store	2	
cms_page	20	
cms_page_store	2	
core_config_data	6	
cron_schedule	8	
customer_address_entity	25	Si
customer_address_entity_datetime	4	Si
customer_address_entity_decimal	4	Si
customer_address_entity_int	4	Si
customer_address_entity_text	4	Si
customer_address_entity_varchar	4	Si
customer_eav_attribute	13	
customer_eav_attribute_website	6	
customer_entity	29	Si
customer_entity_datetime	4	Si
customer_entity_decimal	4	Si
customer_entity_int	4	Si
customer_entity_text	4	Si
customer_entity_varchar	4	Si
customer_form_attribute	2	
customer_grid_flat	26	
customer_group	3	
customer_group_excluded_website	3	
customer_log	4	
customer_visitor	5	
design_change	5	
design_config_grid_flat	5	
directory_country	3	
directory_country_format	4	
directory_country_region	4	
directory_country_region_name	3	

directory_currency_rate	3	
downloadable_link	11	
downloadable_link_price	4	
downloadable_link_purchased	10	
downloadable_link_purchased_item	16	
downloadable_link_title	4	
downloadable_sample	6	
downloadable_sample_title	4	
eav_attribute	17	
eav_attribute_group	7	
eav_attribute_label	4	
eav_attribute_option	3	
eav_attribute_option_swatch	5	
eav_attribute_option_value	4	
eav_attribute_set	4	
eav_entity	9	Si
eav_entity_attribute	6	
eav_entity_datetime	6	Si
eav_entity_decimal	6	Si
eav_entity_int	6	Si
eav_entity_store	5	Si
eav_entity_text	6	Si
eav_entity_type	16	
eav_entity_varchar	6	Si
eav_form_element	5	
eav_form_fieldset	4	
eav_form_fieldset_label	3	
eav_form_type	6	
eav_form_type_entity	2	
email_template	13	
flag	5	
gift_message	5	
googleoptimizer_code	5	
import_history	7	
importexport_importdata	6	
indexer_state	5	
integration	10	
jwt_auth_revoked	3	

layout_link	5	
layout_update	5	
login_as_customer	4	
login_as_customer_assistance_allowed	1	
magento_acknowledged_bulk	2	
magento_bulk	7	
magento_login_as_customer_log	6	
magento_operation	10	
media_content_asset	4	
media_gallery_asset	12	
media_gallery_asset_keyword	2	
media_gallery_keyword	2	
mview_state	6	
newsletter_problem	5	
newsletter_queue	11	
newsletter_queue_link	4	
newsletter_queue_store_link	2	
newsletter_subscriber	7	
newsletter_template	12	
oauth_consumer	8	
oauth_nonce	3	
oauth_token	13	
oauth_token_request_log	5	
password_reset_request_event	5	
patch_list	2	
paypal_billing_agreement	9	
paypal_billing_agreement_order	2	
paypal_cert	4	
paypal_payment_transaction	4	
paypal_settlement_report	5	
paypal_settlement_report_row	19	
persistent_session	6	
product_alert_price	10	Si
product_alert_stock	9	Si
queue	2	
queue_lock	3	
queue_message	3	
queue_message_status	6	

queue_poison_pill	1	
quote	51	
quote_address	64	
quote_address_item	49	
quote_id_mask	3	
quote_item	53	
quote_item_option	5	
quote_payment	22	
quote_shipping_rate	12	
rating	5	
rating_entity	2	
rating_option	5	
rating_option_vote	10	
rating_option_vote_aggregated	8	
rating_store	2	
rating_title	3	
release_notification_viewer_log	3	
report_compared_product_index	6	
report_event	7	
report_event_types	3	
report_viewed_product_aggregated_daily	8	Si
report_viewed_product_aggregated_monthly	8	Si
report_viewed_product_aggregated_yearly	8	Si
report_viewed_product_index	6	Si
reporting_counts	4	
reporting_module_status	6	
reporting_orders	6	
reporting_system_updates	4	
reporting_users	4	
review	5	
review_detail	7	
review_entity	2	
review_entity_summary	6	
review_status	2	
review_store	2	
sales_bestsellers_aggregated_daily	8	
sales_bestsellers_aggregated_monthly	8	
sales_bestsellers_aggregated_yearly	8	

sales_creditmemo	51	
sales_creditmemo_comment	6	
sales_creditmemo_grid	24	
sales_creditmemo_item	34	
sales_invoice	45	
sales_invoice_comment	6	
sales_invoice_grid	26	
sales_invoice_item	34	
sales_invoiced_aggregated	9	
sales_invoiced_aggregated_order	9	
sales_order	139	
sales_order_address	26	
sales_order_aggregated_created	20	
sales_order_aggregated_updated	20	
sales_order_grid	26	
sales_order_item	80	
sales_order_payment	54	
sales_order_status	2	
sales_order_status_history	8	
sales_order_status_label	3	
sales_order_status_state	4	
sales_order_tax	11	
sales_order_tax_item	10	
sales_payment_transaction	10	
sales_refunded_aggregated	8	
sales_refunded_aggregated_order	8	
sales_sequence_meta	4	
sales_sequence_profile	9	
sales_shipment	18	
sales_shipment_comment	6	
sales_shipment_grid	20	
sales_shipment_item	12	
sales_shipment_track	11	
sales_shipping_aggregated	8	
sales_shipping_aggregated_order	8	
salesrule	24	
salesrule_coupon	10	
salesrule_coupon_aggregated	13	

salesrule_coupon_aggregated_order	10	
salesrule_coupon_aggregated_updated	13	
salesrule_coupon_usage	3	
salesrule_customer	4	
salesrule_customer_group	2	
salesrule_label	4	
salesrule_product_attribute	4	
salesrule_website	2	
search_query	10	
search_synonyms	4	
sendfriend_log	4	
sequence_creditmemo_0	1	
sequence_creditmemo_1	1	
sequence_invoice_0	1	
sequence_invoice_1	1	
sequence_order_0	1	
sequence_order_1	1	
sequence_shipment_0	1	
sequence_shipment_1	1	
session	3	
setup_module	3	
shipping_tablerate	9	
sitemap	6	
store	7	Si
store_group	6	
store_website	6	
tax_calculation	5	
tax_calculation_rate	9	
tax_calculation_rate_title	4	
tax_calculation_rule	5	
tax_class	3	
tax_order_aggregated_created	8	
tax_order_aggregated_updated	8	
theme	9	
theme_file	7	
translation	6	
ui_bookmark	9	
url_rewrite	10	

variable	3	
variable_value	5	
vault_payment_token	11	
vault_payment_token_order_payment_link	2	
weee_tax	7	
widget	4	
widget_instance	7	
widget_instance_page	8	
widget_instance_page_layout	2	
wishlist	5	
wishlist_item	7	
wishlist_item_option	5	

Diccionario de datos del Dataset

catalog_category_entity

Physical Column Name	Type	Is Mandatory	Is Primary	Is Foreing	Nullable
entity_id	INTEGER	YES	YES	NO	NOT NULL
attribute_set_id	SMALLINT	YES	NO	NO	NOT NULL
parent_id	INTEGER	YES	NO	NO	NOT NULL
created_at	TIMESTAMP	YES	NO	NO	NOT NULL
updated_at	TIMESTAMP	YES	NO	NO	NOT NULL
path	VARCHAR(255)	YES	NO	NO	NOT NULL
position	INTEGER	YES	NO	NO	NOT NULL
level	INTEGER	YES	NO	NO	NOT NULL
children_count	INTEGER	YES	NO	NO	NOT NULL

catalog_category_entity_datetime

Physical Column Name	Type	Is Mandatory	Is Primary	Is Foreing	Nullable
value_id	INTEGER	YES	YES	NO	NOT NULL
attribute_id	SMALLINT	YES	NO	NO	NOT NULL
store_id	SMALLINT	YES	NO	YES	NOT NULL
entity_id	INTEGER	YES	NO	YES	NOT NULL
value	TIMESTAMP	YES	NO	NO	

catalog_category_entity_decimal

Physical Column Name	Type	Is Mandatory	Is Primary	Is Foreing	Nullable
value_id	INTEGER	YES	YES	NO	NOT NULL
attribute_id	SMALLINT	YES	NO	NO	NOT NULL
store_id	SMALLINT	YES	NO	YES	NOT NULL
entity_id	INTEGER	YES	NO	YES	NOT NULL
value	DECIMAL(20,6)	YES	NO	NO	

Catalog_category_entit_int

Physical Column Name	Type	Is Mandatory	Is Primary	Is Foreing	Nullable
value_id	INTEGER	YES	YES	NO	NOT NULL
attribute_id	SMALLINT	YES	NO	NO	NOT NULL
store_id	SMALLINT	YES	NO	YES	NOT NULL
entity_id	INTEGER	YES	NO	YES	NOT NULL
value	INTEGER	YES	NO	NO	

catalog_category_entity_text

Physical Column Name	Type	Is Mandatory	Is Primary	Is Foreing	Nullable
value_id	INTEGER	YES	YES	NO	NOT NULL
attribute_id	SMALLINT	YES	NO	NO	NOT NULL
store_id	SMALLINT	YES	NO	YES	NOT NULL
entity_id	INTEGER	YES	NO	YES	NOT NULL
value	LONGVARCHAR	YES	NO	NO	

catalog_category_entity_varchar

Physical Column Name	Type	Is Mandatory	Is Primary	Is Foreing	Nullable
value_id	INTEGER	YES	YES	NO	NOT NULL
attribute_id	SMALLINT	YES	NO	NO	NOT NULL
store_id	SMALLINT	YES	NO	YES	NOT NULL
entity_id	INTEGER	YES	NO	YES	NOT NULL
value	VARCHAR(255)	YES	NO	NO	

catalog_product_entity

Physical Column Name	Type	Is Mandatory	Is Primary	Is Foreing	Nullable
entity_id	INTEGER	YES	YES	NO	NOT NULL
attribute_set_id	SMALLINT	YES	NO	NO	NOT NULL
type_id	VARCHAR(32)	YES	NO	NO	NOT NULL
sku	VARCHAR(64)	YES	NO	NO	NOT NULL
has_options	SMALLINT	YES	NO	NO	NOT NULL
required_options	SMALLINT	YES	NO	NO	NOT NULL
created_at	TIMESTAMP	YES	NO	NO	NOT NULL
updated_at	TIMESTAMP	YES	NO	NO	NOT NULL

catalog_product_entity_datetime

Physical Column Name	Type	Is Mandatory	Is Primary	Is Foreing	Nullable
value_id	INTEGER	YES	YES	NO	NOT NULL
attribute_id	SMALLINT	YES	NO	NO	NOT NULL
store_id	SMALLINT	YES	NO	YES	NOT NULL
entity_id	INTEGER	YES	NO	YES	NOT NULL
value	TIMESTAMP	YES	NO	NO	

catalog_product_entity_decimal

Physical Column Name	Type	Is Mandatory	Is Primary	Is Foreing	Nullable
value_id	INTEGER	YES	YES	NO	NOT NULL
attribute_id	SMALLINT	YES	NO	NO	NOT NULL
store_id	SMALLINT	YES	NO	YES	NOT NULL
entity_id	INTEGER	YES	NO	YES	NOT NULL
value	DECIMAL(20,6)	YES	NO	NO	

catalog_product_entity_text

Physical Column Name	Type	Is Mandatory	Is Primary	Is Foreing	Nullable
value_id	INTEGER	YES	YES	NO	NOT NULL
attribute_id	SMALLINT	YES	NO	NO	NOT NULL
store_id	SMALLINT	YES	NO	YES	NOT NULL
entity_id	INTEGER	YES	NO	YES	NOT NULL
value	LONGVARCHAR	YES	NO	NO	

catalog_product_entity_varchar

Physical Column Name	Type	Is Mandatory	Is Primary	Is Foreing	Nullable
value_id	INTEGER	YES	YES	NO	NOT NULL
attribute_id	SMALLINT	YES	NO	NO	NOT NULL
store_id	SMALLINT	YES	NO	YES	NOT NULL
entity_id	INTEGER	YES	NO	YES	NOT NULL
value	VARCHAR(255)	YES	NO	NO	

cataloginventory_stock

Physical Column Name	Type	Is Mandatory	Is Primary	Is Foreing	Nullable
stock_id	SMALLINT	YES	YES	NO	NOT NULL
website_id	SMALLINT	YES	NO	NO	NOT NULL
stock_name	VARCHAR(255)	YES	NO	NO	

cataloginventory_stock_item

Physical Column Name	Type	Is Mandatory	Is Primary	Is Foreing	Nullable
item_id	INTEGER	YES	YES	NO	NOT NULL
product_id	INTEGER	YES	NO	YES	NOT NULL
stock_id	SMALLINT	YES	NO	YES	NOT NULL
qty	DECIMAL(12,4)	YES	NO	NO	
min_qty	DECIMAL(12,4)	YES	NO	NO	NOT NULL
use_config_min_qty	SMALLINT	YES	NO	NO	NOT NULL
is_qty_decimal	SMALLINT	YES	NO	NO	NOT NULL
backorders	SMALLINT	YES	NO	NO	NOT NULL
use_config_backorders	SMALLINT	YES	NO	NO	NOT NULL
min_sale_qty	DECIMAL(12,4)	YES	NO	NO	NOT NULL
use_config_min_sale_qty	SMALLINT	YES	NO	NO	NOT NULL
max_sale_qty	DECIMAL(12,4)	YES	NO	NO	NOT NULL
use_config_max_sale_qty	SMALLINT	YES	NO	NO	NOT NULL
is_in_stock	SMALLINT	YES	NO	NO	NOT NULL
low_stock_date	TIMESTAMP	YES	NO	NO	
notify_stock_qty	DECIMAL(12,4)	YES	NO	NO	
use_config_notify_stock_qty	SMALLINT	YES	NO	NO	NOT NULL
manage_stock	SMALLINT	YES	NO	NO	NOT NULL
use_config_manage_stock	SMALLINT	YES	NO	NO	NOT NULL
stock_status_changed_auto	SMALLINT	YES	NO	NO	NOT NULL
use_config_qty_increments	SMALLINT	YES	NO	NO	NOT NULL
qty_increments	DECIMAL(12,4)	YES	NO	NO	NOT NULL
use_config_enable_qty_inc	SMALLINT	YES	NO	NO	NOT NULL
enable_qty_increments	SMALLINT	YES	NO	NO	NOT NULL
is_decimal_divided	SMALLINT	YES	NO	NO	NOT NULL
website_id	SMALLINT	YES	NO	NO	NOT NULL

store

Physical Column Name	Type	Is Mandatory	Is Primary	Is Foreing	Nullable
store_id	SMALLINT	YES	YES	NO	NOT NULL
code	VARCHAR(32)	YES	NO	NO	
website_id	SMALLINT	YES	NO	NO	NOT NULL
group_id	SMALLINT	YES	NO	NO	NOT NULL
name	VARCHAR(255)	YES	NO	NO	NOT NULL
sort_order	SMALLINT	YES	NO	NO	NOT NULL
is_active	SMALLINT	YES	NO	NO	NOT NULL

Resultado de data Profiling

A continuación, se mostrarán los correspondientes resultados del perfilado de datos, describiéndolo por cada una de las tablas como tal:

catalog_category_entity

Attribute Name	Type	Description	suitable for DM
entity_id	Integer	Identificador único de la entidad de categoría	Sí
attribute_set_id	Smallint	Identificador del conjunto de atributos al que pertenece	Sí
parent_id	Integer	Identificador del padre de la categoría	Sí
created_at	Timestamp	Fecha y hora de creación de la entidad de categoría	No
updated_at	Timestamp	Fecha y hora de actualización de la entidad de categoría	No
path	Varchar(225)	Ruta o ubicación de la categoría en la estructura jerárquica	No
position	Integer	Posición relativa de la categoría en relación con otras	No
level	Integer	Nivel jerárquico de la categoría	No
children_count	Integer	Cantidad de categorías secundarias o hijos de la categoría	No

catalog_category_entity_datetime

Attribute Name	Type	Description	suitable for DM
value_id	Integer	Identificador único del valor de fecha y hora	No
attribute_id	Smallint	Identificador del atributo de fecha y hora	Sí
store_id	Smallint	Identificador de la tienda relacionada	No
entity_id	Integer	Identificador de la entidad de categoría	Sí
value	Timestamp	Fecha y hora del atributo específico de la entidad	Sí

catalog_category_entity_decimal

Attribute Name	Type	Description	suitable for DM
value_id	Integer	Identificador único del valor decimal	No
attribute_id	Smallint	Identificador del atributo relacionado	No
store_id	Smallint	Identificador de la tienda relacionada	No
entity_id	Integer	Identificador de la entidad relacionada	No
value	Decimal(20,6)	Valor decimal	Sí

catalog_category_entity_int

Attribute Name	Type	Description	suitable for DM
value_id	Integer	Identificador único del valor entero	No
attribute_id	Smallint	Identificador del atributo relacionado	No
store_id	Smallint	Identificador de la tienda relacionada	No
entity_id	Integer	Identificador de la entidad relacionada	No
value	Integer	Valor entero	Sí

catalog_category_entity_text

Attribute Name	Type	Description	suitable for DM
value_id	Integer	Identificador único del valor de texto	No
attribute_id	Smallint	Identificador del atributo relacionado	No
store_id	Smallint	Identificador de la tienda relacionada	No
entity_id	Integer	Identificador de la entidad relacionada	No
value	Longvarchar	Valor de texto	Sí

catalog_category_entity_varchar

Attribute Name	Type	Description	suitable for DM
value_id	Integer	Identificador único del valor varchar	No
attribute_id	Smallint	Identificador del atributo relacionado	No
store_id	Smallint	Identificador de la tienda relacionada	No
entity_id	Integer	Identificador de la entidad relacionada	No

catalog_product_entity

Attribute Name	Type	Description	suitable for DM
entity_id	Integer	Identificador único de la entidad del producto	Sí
attribute_set_id	Smallint	Identificador del conjunto de atributos al que pertenece	Sí
type_id	Varchar(32)	Identificador del tipo de producto	No
Sku	Varchar(64)	Código SKU del producto	Sí
has_options	Smallint	Indicador de si el producto tiene opciones	Sí
required_options	Smallint	Indicador de si el producto requiere opciones	Sí
created_at	Timestamp	Fecha y hora de creación del producto	No
updated_at	Timestamp	Fecha y hora de actualización del producto	No

catalog_producto_entity_datetime

Attribute Name	Type	Description	suitable for DM
value_id	Integer	Identificador único del valor de fecha y hora	No
attribute_id	Smallint	Identificador del atributo de fecha y hora	Sí
store_id	Smallint	Identificador de la tienda relacionada	No
entity_id	Integer	Identificador de la entidad del producto	Sí
value	Timestamp	Fecha y hora del atributo específico de la entidad	Sí

catalog_product_entity_decimal

Attribute Name	Type	Description	suitable for DM
value_id	Integer	Identificador único del valor decimal	No
attribute_id	Smallint	Identificador del atributo relacionado	No
store_id	Smallint	Identificador de la tienda relacionada	No
entity_id	Integer	Identificador de la entidad relacionada	No
value	Decimal(20,6)	Valor decimal	Sí

catalog_product_entity_text

Attribute Name	Type	Description	suitable for DM
value_id	Integer	Identificador único del valor de texto	No
attribute_id	Smallint	Identificador del atributo relacionado	No
store_id	Smallint	Identificador de la tienda relacionada	No
entity_id	Integer	Identificador de la entidad relacionada	No
value	Longvarchar	Valor de texto	Sí

catalog_product_entity_varchar

Attribute Name	Type	Description	suitable for DM
value_id	Integer	Identificador único del valor varchar	No
attribute_id	Smallint	Identificador del atributo relacionado	No
store_id	Smallint	Identificador de la tienda relacionada	No
entity_id	Integer	Identificador de la entidad relacionada	No
value	Varchar(255)	Valor varchar	Sí

cataloginventory_stock

Attribute Name	Type	Description	suitable for DM
stock_id	Smallint	Identificador único del stock	Sí
website_id	Smallint	Identificador del sitio web relacionado	Sí
stock_name	Varchar(225)	Nombre del stock	Sí

cataloginventory_stock_item

Attribute Name	Type	Description	suitable for DM
item_id	Integer	Identificador único del elemento de stock	Sí
product_id	Integer	Identificador del producto relacionado	Sí
stock_id	Smallint	Identificador del stock relacionado	Sí
qty	Decimal(12,4)	Cantidad en stock del producto	Sí
min_qty	Decimal(12,4)	Cantidad mínima requerida en stock	Sí
use_config_min_qty	Smallint	Indicador de si se usa la configuración de cantidad mínima	Sí
is_qty_decimal	Smallint	Indicador de si la cantidad es decimal	Sí
backorders	Smallint	Indicador de si se permiten backorders	Sí
use_config_backorders	Smallint	Indicador de si se usa la configuración de backorders	Sí
min_sale_qty	Decimal(12,4)	Cantidad mínima requerida para una venta	Sí
use_config_min_sale_qty	Smallint	Indicador de si se usa la configuración de cantidad mínima de venta	Sí
max_sale_qty	Decimal(12,4)	Cantidad máxima permitida para una venta	Sí

use_config_max_sale_qty	Smallint	Indicador de si se usa la configuración de cantidad máxima de venta	Sí
is_in_stock	Smallint	Indicador de si el producto está en stock	Sí
low_stock_date	Timestamp	Fecha y hora de bajo stock	No
notify_stock_qty	Decimal(12,4)	Cantidad para notificar el bajo stock	Sí
use_config_notify_stock_qty	Smallint	Indicador de si se usa la configuración de cantidad para notificar el bajo stock	Sí
manage_stock	Smallint	Indicador de si se gestiona el stock	Sí
use_config_manage_stock	Smallint	Indicador de si se usa la configuración para gestionar el stock	Sí
stock_status_changed_auto	Smallint	Indicador de si el estado de stock se cambia automáticamente	Sí
use_config_qty_increments	Smallint	Indicador de si se usa la configuración de incrementos de cantidad	Sí
qty_increments	Decimal(12,4)	Incrementos de cantidad	Sí
use_config_enable_qty_inc	Smallint	Indicador de si se usa la configuración para habilitar incrementos de cantidad	Sí
enable_qty_increments	Smallint	Indicador de si se habilitan incrementos de cantidad	Sí
is_decimal_divided	Smallint	Indicador de si la cantidad se divide en decimal	Sí
website_id	Smallint	Identificador del sitio web relacionado	Sí

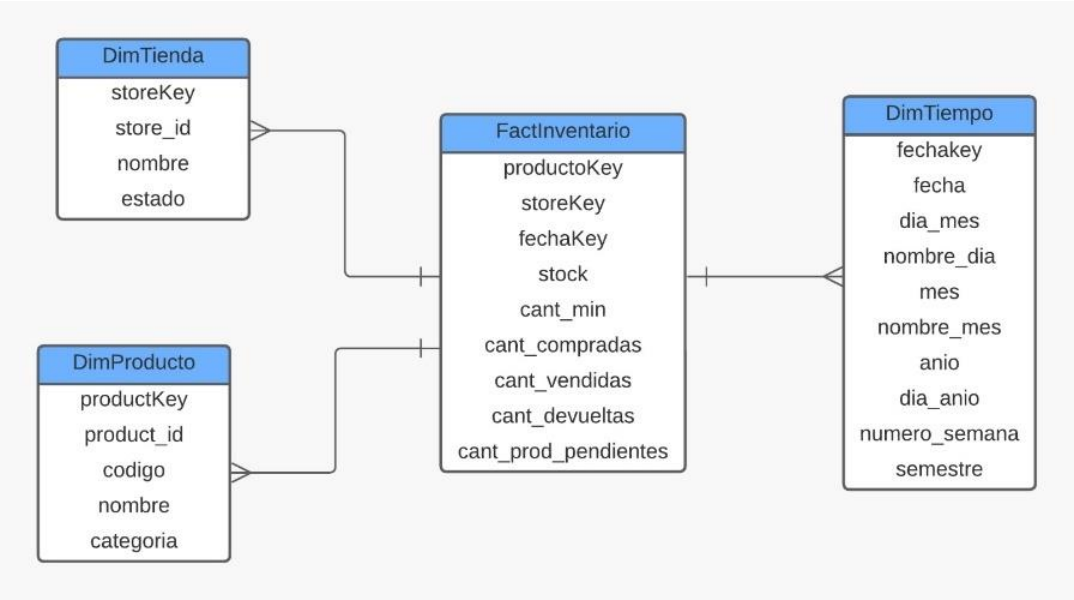
store

Attribute Name	Type	Description	suitable for DM
store_id	Smallint	Identificador de la tienda	Sí
código	Varchar(32)	Código de la tienda	Sí
website_id	Smallint	Identificador del sitio web relacionado	No
group_id	Smallint	Identificador del grupo de tiendas relacionado	No
nombre	Varchar(225)	Nombre de la tienda	Sí
sort_order	Smallint	Orden de clasificación de la tienda	No
is_active	Smallint	Estado de la tienda (activo/inactivo)	Sí

Modelo dimensional propuesto

Luego de la aplicación del modelado dimensional ante los requerimientos iniciales presentamos a continuación el resultado:

Esquema Estrella	
Proceso de negocio:	productos agregados diariamente
Granularidad:	1 fila corresponde a un producto agregado
Dimensiones:	dimTienda, dimProducto, dimFecha
Métricas:	stock, cant_min, cant_vendidas, cant_devueltas, cant_compradas



Mapping por tabla

NOMBRE DE LA TABLA:	**DIMPRODUCTO**
TIPO DE TABLA:	Tabla de dimensión
DESCRIPCIÓN:	Dimensión de producto, que incluye todos los productos de la base transaccional
LLAVE PRIMARIA:	productKey
LLAVE FORÁNEA:	tiendaId
POLÍTICA DE NULIDAD:	Todos los campos son requeridos
POLÍTICA DE SCD	SCD tipo 0: código SCD tipo 1: tiendaId SCD tipo 2: nombre, costo

Nombre de la columna	Descripción	Grupo de atributos	Tipo de dato	Sistema origen	Tabla origen	Columna origen
Código	Llave primaria subrogada	Identificador	Int(10)	Autoincremento	Catalog_product_entity	Sku
product Key	Llave primaria subrogada	Identificador	Int(10)	elAngelBD	dimProducto	productKey
tiendaId	Llave primaria subrogada	Identificador	Int(10)	elAngelBD	dimTienda	Storekey
Nombre	Nombre del producto	Detalle del nombre de producto	Varchar(255)	elAngelBD	Catalog_product_entit_varchar	Name
Costo	Precio del producto	Detalle del precio del producto	Float(2,4)	elAngelBD	Catalag_product_entity_decimal	Value

NOMBRE DE LA TABLA:	**DIMTIENDA**
TIPO DE TABLA:	Tabla de dimensión
DESCRIPCIÓN:	Dimensión de producto, que incluye todas las tiendas de la base transaccional
LLAVE PRIMARIA:	StoreKey
LLAVE FORÁNEA:	n/a
POLÍTICA DE NULIDAD:	Todos los campos son requeridos
POLÍTICA DE SCD	SCD tipo 0: código SCD tipo 1: nombre

Nombre de la columna	Descripción	Grupo de atributos	Tipo de dato	Sistema origen	Tabla origen	Columna origen
Código	Llave primaria subrogada	Identificador	Int(10)	Autoincremento	Store	Code
storeKey	Llave primaria subrogada	identificador	Int(10)	elAngelBD	dimTienda	storeKey
Nombre	Nombre de la tienda	Detalle del nombre de la tienda	Varchar(255)	elAngelBD	Store	Name
Estado	Estado de la tienda	Detalle del estado de la tienda	Int(10)	elAngelBD	store	Is_active

NOMBRE DE LA TABLA:	**DIMFECHA**
TIPO DE TABLA:	Tabla de dimensión
DESCRIPCIÓN:	Dimensión de tiempo, que incluye la fecha y su detalle diario para un período de tiempo
LLAVE PRIMARIA:	fechaKey
LLAVE FORÁNEA:	n/a
POLÍTICA DE NULIDAD:	Todos los campos son requeridos
POLÍTICA DE SCD	SCD tipo 0: SCD tipo 1: SCD tipo 2:

Nombre de la columna	Descripción	Grupo de atributos	Tipo de dato	Sistema origen	Tabla origen	Columna origen
fechaKey	Llave primaria, formato ISO de la fecha	Identificador	Int(10)	Script de fechas	-	-
Fecha	Fecha	Detalles de la fecha	Smalldatetime	Script de fechas	-	-
Dia_mes	Número del día del mes	Detalle de la fecha	Int(10)	Script de fechas	-	-
Nombre_dia	Nombre del día	Detalle de la fecha	Varchar(15)	Script de fechas	-	-
Mes	Número del mes	Detalle de la fecha	Int(10)	Script de fechas	-	-
Nombre_mes	Nombre del mes	Detalle de la fecha	Varchar(15)	Script de fechas	-	-
Año	Número del año	Detalle de la fecha	Int(10)	Script de fechas	-	-
Dia_año	Día del año	Detalle de la fecha	Int(10)	Script de fechas	-	-
Numero_semana	Numero de la semana	Detalle de la fecha	Int(5)	Script de fechas	-	-
semestre	Número del semestre	Detalle de la fecha	Int(5)	Script de fechas	-	-

Nombre de la tabla:	**FactVentas**
Tipo de tabla:	Tabla de hechos
Descripción:	Contiene los hechos para el proceso de ventas
Llave primaria:	Llave compuesta por: productKey, storeKey, fechaKey
Llave foránea:	ProductKey hace referencia a la DimProducto StoreKey hace referencia a la DimTienda FechaKey hace referencia a la DimFecha
Política de nulidad:	Todos los campos son requeridos

Nombre de la columna	Descripción	Grupo de atributos	Tipo de dato	Sistema origen	Tabla origen	Column a origen
productKey	Llave primaria a la dimesión producto	identificador	Int(10)	DW	productKey	-
Storekey	Llave primaria a la dimesión tienda	identificador	Int(10)	DW	storeKey	-
Fechakey	Llave primaria a la dimesión fecha	identificador	Int(10)	DW	fechaKey	-
Stock	Cantidad de producto en inventario	Detalle de producto	Int(10)	ElAngelBD	Cataloginventory_stock_item	Qty
Cant_min	Cantidad mínima del producto	Detalle de cantidad de producto	Int(10)	ElAngelBD	Cataloginventory_stock_item	Min_qty
Cant_comprada	Cantidad comprada del producto	Detlla de compras de producto	Int(10)	ElAngelBD	Inventory_sourve_item	Quantity
Cant_vendida	Cantidad vendiad del producto	Detalle de cantidad vendida de produto	Int (10)	ElAngelBD	Sales_order	Invoiced
Cant_devuelta	Cantidad devuelta del producto	Detalle de productos devueltos	Int(10)	ElAngelBD	Sales_order	Refuded
Cant_prod_pendiente	Cantidad pendiente del producto	Detalle de productos pendientes	Int(10)	ElAngelBD	Cataloginventory_stock_item	backorders

b. Descripción de la propuesta de solución

La solución propuesta para "Importaciones El Ángel" representa un enfoque integral y avanzado para la gestión y análisis de inventario, aprovechando la potencia y flexibilidad de las tecnologías de big data y BI. El proceso para llegar a la implementación efectiva en Power BI comprende varias etapas clave, cada una construida sobre la base de la anterior, asegurando un flujo de trabajo cohesivo y eficiente.

Análisis de la Base de Datos de Magento: El proyecto comenzó con un análisis en profundidad de la base de datos de Magento, donde se extrajo información crítica relacionada con el inventario. Esta fase fue crucial para comprender la estructura de datos existente y para identificar los datos relevantes para el análisis de inventario.

Modelado Dimensional: Con las perspectivas obtenidas del análisis de datos, se diseñó un modelo dimensional que refleja las necesidades específicas de "Importaciones El Ángel". Este modelo, estructurado alrededor de una tabla de hechos central y varias dimensiones relacionadas, permite un análisis multifacético y en profundidad del inventario.

Conexiones y Transformaciones con Talend: Utilizando Talend Open Studio, se establecieron conexiones entre la base de datos de pgAdmin y el modelo dimensional. Se desarrollaron procesos de ETL que transforman y limpian los datos, preparándolos para su carga en el data warehouse en AWS. Este paso fue esencial para asegurar que los datos sean precisos y estén listos para el análisis.

Implementación en AWS: Los datos procesados se cargaron luego en Amazon Redshift, proporcionando una plataforma de data warehousing en la nube que es tanto escalable como segura. Además, se utilizó AWS S3 como un data lake para almacenar grandes volúmenes de datos en su forma más bruta, ofreciendo así flexibilidad para análisis futuros.

Visualización en Power BI: Finalmente, los datos almacenados en AWS se conectaron a Power BI, donde se crearon dashboards y reportes interactivos. Estas visualizaciones permiten a los stakeholders de "Importaciones El Ángel" obtener una visión clara y actualizada del estado de su inventario, identificar tendencias y tomar decisiones informadas rápidamente.

Durante todo el proceso, se puso especial énfasis en la precisión, la seguridad de los datos y la escalabilidad del sistema, asegurando que la solución no solo satisfaga las necesidades actuales, sino que también pueda adaptarse a los cambios futuros y al crecimiento de la empresa.

Este enfoque holístico y la implementación detallada de la solución propuesta demuestran la eficacia de combinar herramientas de vanguardia y técnicas avanzadas de análisis de datos para mejorar la gestión de inventario en el ámbito del comercio electrónico.

c. Especificaciones de necesidades analíticas

1) ¿Cuánto producto se descargó del inventario de una tienda en un determinado periodo de tiempo?
2) ¿Cuánto producto ingreso al inventario de una tienda en un determinado periodo de tiempo?
3) ¿Cuál fue el inventario de los productos al final del día?
4) ¿Cuántos productos están agotados por tienda al final del día?
5) ¿Cantidad mínima de productos que deben mantenerse en stock para evitar agotamientos?
6) ¿Cantidad de pedidos pendientes por tienda y producto?

d. Descripción de la tecnología a utilizar

Talend Open Studio

- Descripción General: Talend Open Studio es una herramienta de integración de datos de código abierto y gratuita ofrecida por Talend.
- Características Clave: Ofrece capacidades para diseñar, testear y ejecutar tareas de integración de datos. Incluye una interfaz gráfica de usuario para facilitar la creación de flujos de trabajo de datos y soporta la integración de una amplia variedad de fuentes y destinos de datos.
- Uso en Proyectos: Ampliamente utilizado para proyectos de integración de datos de menor escala, transformación de datos, migración de bases de datos y procesamiento por lotes.

AWS S3 (Amazon Simple Storage Service)

- Descripción General: AWS S3 es un servicio de almacenamiento de objetos en la nube que ofrece escalabilidad, seguridad de datos, y alta disponibilidad.
- Características Clave: Almacenamiento de cualquier cantidad de datos, con acceso desde cualquier lugar, robusta seguridad y cifrado de datos, y una amplia gama de opciones de gestión de datos.
- Uso en Proyectos: Ampliamente utilizado para el almacenamiento de backup, sitios web estáticos, aplicaciones móviles, y como almacenamiento para aplicaciones de big data y análisis.

Power BI

- Descripción General: Power BI es una herramienta de visualización de datos y business intelligence que permite transformar datos en información accionable.
- Características Clave: Incluye capacidades de modelado de datos, visualizaciones interactivas, y la capacidad de integrar con diversas fuentes de datos. También permite la creación de reportes y dashboards personalizables.
- Uso en Proyectos: Se usa para el análisis de datos empresariales, la creación de reportes y paneles para la toma de decisiones basada en datos, y la distribución de insights en una organización.

phpMyAdmin

- Descripción General: phpMyAdmin es una herramienta de administración de bases de datos MySQL y MariaDB basada en web.
- Características Clave: Permite la gestión de bases de datos, tablas, columnas, relaciones, índices, usuarios, y permisos. Facilita la ejecución de sentencias SQL, y la exportación e importación de datos.
- Uso en Proyectos: Esencial para la administración de bases de datos en proyectos web, especialmente útil en entornos de hosting compartido o para usuarios que prefieren una interfaz gráfica.

Magento

- Descripción General: Magento es una plataforma de comercio electrónico de código abierto que ofrece una gran flexibilidad y control sobre la apariencia, contenido y funcionalidad de una tienda en línea.
- Características Clave: Ofrece un potente motor de personalización, una amplia variedad de plugins y temas, capacidades SEO robustas, y soporte para múltiples tiendas y lenguajes.
- **Uso en Proyectos**: Utilizado para desarrollar tiendas en línea escalables y personalizadas, desde pequeñas empresas hasta grandes corporaciones, con la capacidad de adaptarse a diversas necesidades de negocio.

pgAdmin

- Descripción General: pgAdmin es una herramienta de administración y desarrollo de código abierto para PostgreSQL, la base de datos relacional de objetos.
- Características Clave: Incluye herramientas para gestionar bases de datos, diseñar esquemas, ejecutar consultas y scripts SQL, y visualizar y editar datos.
- Uso en Proyectos: Utilizado por administradores de bases de datos, desarrolladores y analistas para gestionar bases de datos PostgreSQL, tanto en entornos de desarrollo como de producción.

Amazon Redshift

- Descripción General: Amazon Redshift es un servicio de data warehouse en la nube rápido, escalable y completamente gestionado, diseñado para el análisis de grandes volúmenes de datos y la integración con herramientas de BI.
- Características Clave: Almacenamiento en columnas, compresión de datos, ejecución paralela de consultas, escalabilidad masiva, gestión simplificada, alta seguridad y cifrado, y una estructura de precios predecible.
- Uso en Proyectos: Ideal para análisis de big data, almacenamiento y análisis de datos históricos para business intelligence, y análisis en tiempo real cuando se combina con otras herramientas y servicios de AWS.

e. Diagrama arquitectónico de la solución

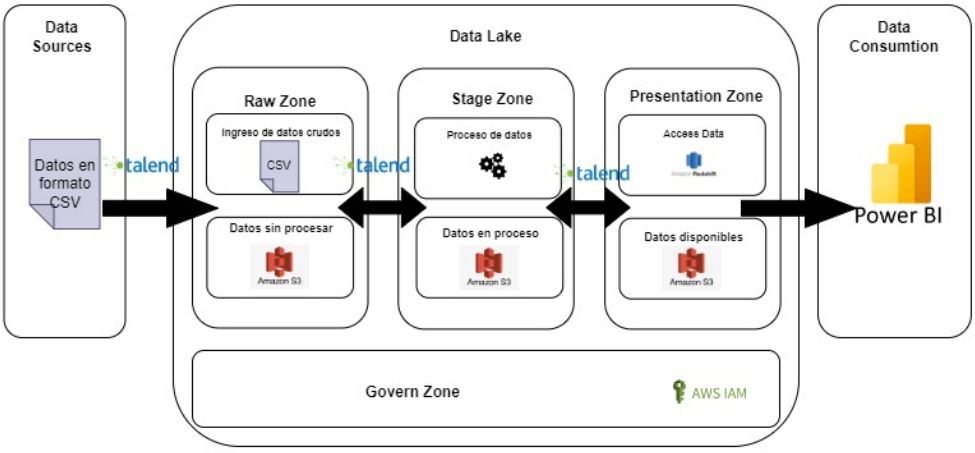


Diagrama arquitectónico de la solución basado en la arquitectura de Ralp Kimball

f. Descripción de cada componente de la solución

En este apartado se describirán y demostrará algunos de los componentes descritos en la arquitectura anterior junto con un ejemplo de los procesos ETL que se llevaron a cabo para el desarrollo de este proyecto.

Data sources: Conexión a la base de datos origen.

La fuente origen de este proyecto se encuentra alojada en el gestor de base de datos PostgreSQL con el nombre de "almacén". Por lo cual para poder utilizar esos datos se establece una conexión con esa base de datos origen a través de Talend. Esta herramienta tiene la opción para guardar diferentes tipos de componentes como: conexiones a bases de datos, esquemas de bases de datos, ficheros de todo tipo, servicios web, conexiones FTP, LDAP, etc., esa opción se llama "Metadatos" la cual se ocupará para guardar la conexión a la base de datos MySQL, la conexión a Redshift y esquemas de los archivos .CSV a utilizar a lo largo del proyecto.

Procesos ETL (Extracción, Transformación y Carga): Se refiere a los procesos con los cuales se realiza el movimiento de datos desde las múltiples fuentes, así como también los procesos para reformatearlos, limpiarlos y cargarlos en otras bases de datos, para el proyecto se hace uso de la herramienta Talend Open Studio para extraer la información desde la fuente origen, así como hacer los procesos correspondientes entre las diferentes zonas que componen el Data Lake.

Data Lake: Es un repositorio centralizado diseñado para almacenar, procesar y proteger grandes cantidades de datos estructurados, semiestructurados o no estructurados. Para nuestro proyecto se utilizó Amazon S3 como repositorio para almacenar todos los archivos .CSV que vaya generando el proceso de Data Warehouse.

Dentro del Data Lake nos encontramos 4 zonas:

- **Raw Zone:** Área donde se almacenan los datos crudos extraídos de la base de datos origen. Estos datos son almacenados sin ningún tipo de transformación y conservando su formato original, tal cual están almacenados en las fuentes de datos. Para el proyecto se utilizó Amazon S3 dentro de la carpeta denominada RAW, donde se guardan los datos crudos en formato .CSV.
- **Stage Zone:** Área donde se almacenan los datos que están siendo procesados o han sufrido alguna transformación final (es decir, han sido limpiados, filtrados, se les ha convertido el tipo de dato o formato, etc, entre otro tipo de transformaciones), estos datos serán utilizados para la construcción de las respectivas tablas de hechos y dimensiones. Para el proyecto se utilizó Amazon S3 dentro de la carpeta denominada STAGE, donde se guardan los datos en proceso en formato .CSV.
- **Presentation Zone:** Área donde se almacenan los datos listos para su respectivo consumo por aplicaciones de BI. Estos datos ya están limpios, filtrados, transformados y organizados en las respectivas tablas de hechos y dimensiones. Para el proyecto se utilizó Amazon S3 dentro de la carpeta denominada PRESENTATION, donde se guardan los datos listos para ser consumidos en formato .CSV.
- **Access Data:** Dentro del área de Presentation Zone se hace uso de la herramienta Amazon Redshift la cual se encarga guardar la estructura del modelo dimensional y sus datos procesados y listos. A través de esta herramienta se da acceso a los datos para que diferentes aplicaciones BI puedan consumirlos.
- **Govern Zone:** Área donde se administran las reglas, políticas y usuarios que tendrán acceso a tanto a los datos que se utilizarán para la construcción del Data Warehouse, como a la administración de las diferentes herramientas utilizadas dentro del proceso. Es la capa de seguridad que se aplica en cada área que compone el Data Lake. Para el proyecto se hace uso de Amazon IAM.
- **Data consumption:** Área donde se encuentran las aplicaciones BI las cuales son las que consumen los datos procesados por el Data Warehouse para dar respuesta a los requerimientos analíticos del negocio. Para el proyecto se hace uso de la herramienta Power BI en la cual se generan los respectivos Dashboards para cada proceso de negocio donde se analizan y diseñan los gráficos y reportes a presentar.

Capitulo III: Estrategia de implementación de propuesta de solución

a. Estrategia de implementación

Estrategia de Implementación para el Análisis de Inventario de Importaciones El Ángel

La solución diseñada está enfocada en empresas como Importaciones El Ángel, una empresa que opera en el sector de comercio electrónico B2C, especializada en la venta de productos decorativos para el hogar. Este modelo de negocio implica la venta directa a los consumidores, tanto en plataformas digitales como físicas, lo que conlleva desafíos únicos en la gestión del inventario.

Adaptabilidad a Variaciones de Datos: Dado que la cantidad de datos manejados por Importaciones El Ángel puede variar significativamente dependiendo de la temporada y las tendencias de mercado, la solución utiliza Amazon Web Services (AWS). Esta elección permite una escalabilidad y flexibilidad en el manejo de los datos, adecuando el uso y coste de los recursos según las necesidades de cada periodo.

Acceso y Manejo de Datos: Para minimizar la carga de trabajo de Importaciones El Ángel, la solución propuesta no requiere esfuerzos adicionales por parte de la empresa para proporcionar datos de transacciones. Mediante herramientas específicas, se accederá directamente a los datos almacenados en la plataforma Magento Commerce para realizar el análisis. Esto facilita la extracción y procesamiento de datos, permitiendo un análisis profundo y la generación de insights valiosos.

Proceso ETL y Herramientas Utilizadas: Para el proceso ETL (Extract, Transform, Load), se ha elegido Talend Open Studio debido a su curva de aprendizaje relativamente baja y su eficacia en la gestión de datos. Talend permitirá un manejo eficiente de los datos de inventario, desde la extracción de las fuentes originales en Magento hasta la transformación y carga de estos en un formato adecuado para el análisis.

Análisis y Visualización de Datos: Una vez procesados los datos, se utilizará Power BI para la visualización. Esto proporcionará a los usuarios de negocios de Importaciones El Ángel herramientas de visualización de datos intuitivas y poderosas, facilitando la toma de decisiones basada en datos sobre la gestión del inventario, identificación de tendencias, y optimización de la cadena de suministro.

En resumen, la estrategia de implementación para el proyecto de análisis de inventario de Importaciones El Ángel está diseñada para ser eficiente, escalable y de bajo impacto operativo para la empresa, mientras se proporciona un análisis detallado y visualizaciones que impulsan la toma de decisiones estratégicas.

Estrategias para Implementar el Proyecto de Análisis de Inventario

Para asegurar una implementación exitosa en un entorno de producción, es crucial instalar y configurar adecuadamente todas las herramientas necesarias. A continuación, se detallan los pasos y consideraciones clave:

1. **Instalación y Configuración de Talend Open Studio**:
    - **Talend** es fundamental para el desarrollo de los procesos ETL requeridos para construir un modelo dimensional eficaz para el análisis de inventario.
    - **Instalación de Java**: Antes de instalar Talend, es necesario contar con el Java Development Kit (JDK). Talend utiliza Java para compilar archivos fuente, por lo que es imprescindible tenerlo instalado.
    - **Versión del JDK**: Conforme a la documentación oficial de Talend, las versiones recomendadas del JDK para ejecutar Talend Open Studio 8.0 son JDK 8 o JDK 11, siendo este último el más aconsejable.
2. **Configuración de Conexión a Magento Commerce**:
    - Para acceder a los datos de inventario almacenados en Magento, es necesario establecer una conexión segura y eficiente. Esto implica configurar las credenciales de acceso y garantizar una integración adecuada entre Talend y Magento.
3. **Preparación del Entorno de AWS**:
    - Dado que se utilizará AWS para el almacenamiento y procesamiento de datos, es importante configurar adecuadamente los servicios de AWS, como Amazon Redshift, para manejar eficientemente el volumen de datos y las necesidades de escalabilidad.
4. **Configuración de Power BI para la Visualización de Datos**:
    - Tras el procesamiento y modelado de los datos, se utilizará Power BI para la visualización. Esto implica preparar dashboards y reportes que permitan a los usuarios de negocio de Importaciones El Ángel tomar decisiones informadas basadas en los datos analizados.
5. **Pruebas y Validación**:
    - Antes de pasar a producción, es crucial realizar pruebas exhaustivas para asegurar que todos los componentes del sistema funcionan como se espera y que los datos procesados son precisos y confiables.
6. **Documentación y Capacitación**:
    - Preparar documentación detallada sobre el uso y mantenimiento del sistema y ofrecer capacitación a los usuarios finales para asegurar una transición fluida y eficiente al nuevo sistema de análisis de inventario.

Esta estrategia de implementación asegura que todos los aspectos técnicos y operativos necesarios para el proyecto de análisis de inventario de Importaciones El Ángel sean considerados y ejecutados adecuadamente, proporcionando una base sólida para el análisis eficiente y efectivo del inventario.

Instalación y Configuración de Talend Open Studio

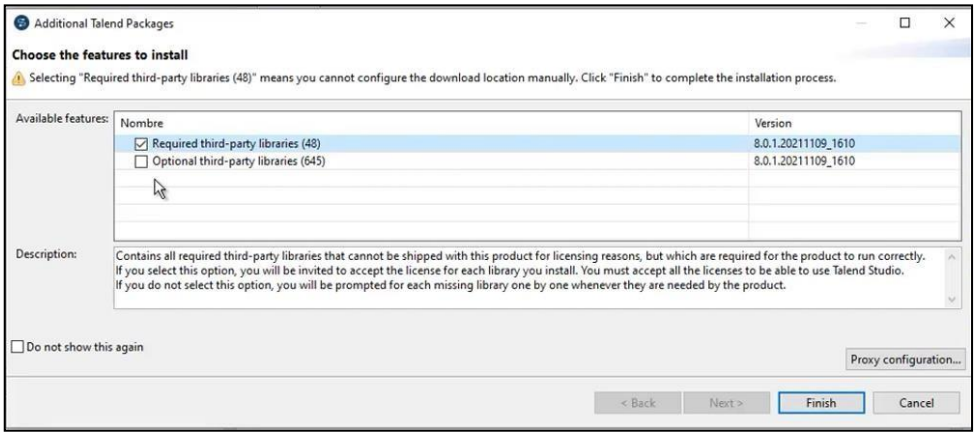
Para implementar efectivamente los procesos ETL en el proyecto de análisis de inventario de Importaciones El Ángel, la instalación y configuración adecuada de Talend Open Studio es un paso crucial. A continuación, se detallan los pasos para realizar esta instalación:

1. **Instalación del JDK 11**:
    - Antes de instalar Talend Open Studio, asegúrese de que el Java Development Kit (JDK) versión 11 esté instalado en su sistema, ya que es necesario para la compilación y ejecución de Talend.
2. **Descarga de Talend Open Studio**:
    - Visite la página oficial de Talend en https://www.talend.com/products/talend-open-studio/.
    - Para descargar el software, es necesario registrarse proporcionando su correo electrónico y otra información requerida.
    - Una vez registrado, recibirá un correo electrónico con enlaces de descarga para los diferentes sistemas operativos.
3. **Proceso de Instalación**:
    - Elija el enlace de descarga adecuado para su sistema operativo y proceda con la descarga.
    - Siga las instrucciones para instalar Talend Open Studio en su computadora.
4. **Configuración Inicial de Talend**:
    - Al abrir Talend por primera vez, el software solicitará la descarga de ciertas librerías adicionales requeridas para su funcionamiento.
    - Acepte la descarga, y Talend instalará automáticamente estas librerías.
    - Es importante asegurarse de que todas las librerías y complementos necesarios estén instalados correctamente para evitar errores durante el desarrollo de los procesos ETL.

5. **Verificación y Pruebas**:

    - Una vez completada la instalación y configuración, es recomendable realizar una verificación y pruebas básicas para asegurarse de que Talend funciona correctamente y está listo para su uso en el proyecto.

Al seguir estos pasos, se garantiza una instalación y configuración exitosa de Talend Open Studio, preparando el terreno para el desarrollo eficiente de los procesos ETL necesarios para el análisis de inventario de Importaciones El Ángel.



Configuración Adicional de Talend para PgAdmin y Componentes de Amazon

Tras la instalación inicial de Talend Open Studio para el proyecto de Importaciones El Ángel, es necesario realizar pasos adicionales para asegurar la integración con PgAdmin (PostgreSQL) y los servicios de Amazon Web Services, como S3 y Redshift.

1. **Instalación de .jar para PostgreSQL**:

    - Para la conexión y manipulación de bases de datos gestionadas a través de PgAdmin, Talend requiere archivos .jar específicos para PostgreSQL.

    - Al utilizar componentes relacionados con PostgreSQL en Talend, se te solicitará la descarga de los archivos .jar necesarios. Estos archivos permiten que Talend se conecte y realice operaciones en las bases de datos PostgreSQL.

    - Acepta la descarga cuando Talend la solicite, y el software gestionará automáticamente la instalación de estos archivos.

2. **Configuración para Componentes de Amazon (S3 y Redshift)**:

    - Para la integración con servicios de AWS como Amazon S3 y Redshift, Talend necesita archivos .jar adicionales.

- Al acceder a los componentes de Talend relacionados con estos servicios, se te pedirá descargar los correspondientes archivos .jar. Acepta estas solicitudes para permitir que Talend los instale automáticamente.

3. **Verificación de la Instalación**:

    - Una vez completadas estas instalaciones, es crucial verificar que todos los componentes funcionan adecuadamente. Esto incluye probar la conexión a PostgreSQL a través de PgAdmin y realizar operaciones básicas en AWS, como la manipulación de datos en S3 o la realización de consultas en Redshift.

4. **Documentación y Respaldo**:

    - Se recomienda documentar detalladamente los pasos de instalación y mantener copias de seguridad de los archivos .jar, en caso de que sean necesarios para futuras referencias, reinstalaciones o solución de problemas.

Al finalizar estos pasos, Talend Open Studio estará completamente configurado y preparado para integrarse efectivamente con bases de datos PostgreSQL gestionadas a través de PGAdmin y con servicios de AWS, facilitando así un análisis de datos de inventario eficiente y robusto para Importaciones El Ángel.

Instalación y Configuración de Power BI Desktop para Análisis de Inventario

Power BI Desktop es una herramienta esencial en este proyecto para la creación de dashboards y reportes que faciliten la visualización de datos y la presentación de resultados. A continuación, se detallan los pasos para su instalación y configuración:

1. **Descarga de Power BI Desktop**:

    - Visita la página oficial de Power BI Desktop en https://powerbi.microsoft.com/es-es/desktop/.

    - La descarga de Power BI Desktop es gratuita y se puede iniciar directamente desde el sitio web.

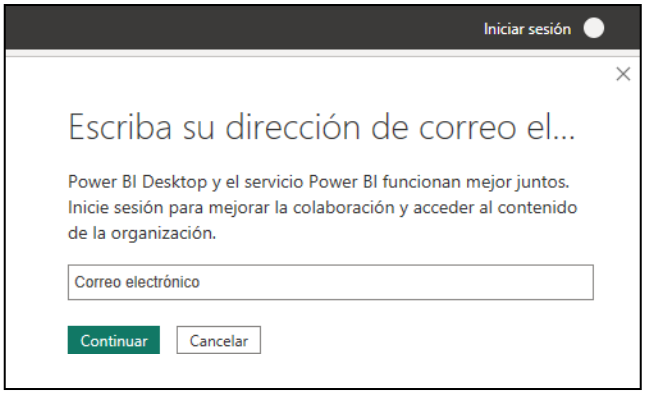
2. **Proceso de Instalación**:

    - Una vez descargado el instalador, sigue las instrucciones para instalar Power BI Desktop en tu computadora.

    - La instalación suele ser sencilla y directa, con opciones predeterminadas adecuadas para la mayoría de los usuarios.

3. **Inicio de Sesión y Configuración Inicial**:
    - Se recomienda iniciar sesión en Power BI Desktop utilizando una cuenta de correo electrónico asociada a Power BI Web, si se tiene acceso a una.
    - Iniciar sesión con una cuenta de Power BI ofrece ventajas en términos de colaboración, organización y seguridad para los dashboards creados.
    - La integración con Power BI Web también facilita la compartición de dashboards y reportes con los usuarios de negocio y otros stakeholders del proyecto.
4. **Configuración de Seguridad y Colaboración**:
    - Asegúrate de configurar las opciones de seguridad y colaboración adecuadas en Power BI Desktop, especialmente si los dashboards contienen información sensible o confidencial.
    - Explora las opciones de publicación y compartición para determinar la mejor manera de presentar los resultados a los usuarios finales.
5. **Creación y Diseño de Dashboards**:
    - Comienza a explorar las funcionalidades de Power BI Desktop para el diseño de dashboards y reportes.
    - Considera utilizar plantillas, si están disponibles, para acelerar el proceso de creación de informes y asegurar una presentación consistente y profesional.

Al seguir estos pasos, Power BI Desktop estará listo para ser utilizado en el análisis de datos de inventario de Importaciones El Ángel, permitiendo una visualización efectiva y una presentación impactante de los resultados del análisis.

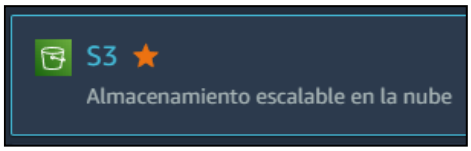


Configuración de los Componentes de AWS para el Proyecto

La utilización efectiva de los servicios de Amazon Web Services es un paso esencial en nuestro proyecto, y requiere la creación de una cuenta en esta plataforma.

Amazon S3: Almacenamiento de Datos

- Para empezar, accede a la consola de AWS y localiza Amazon S3, como se muestra en la siguiente figura

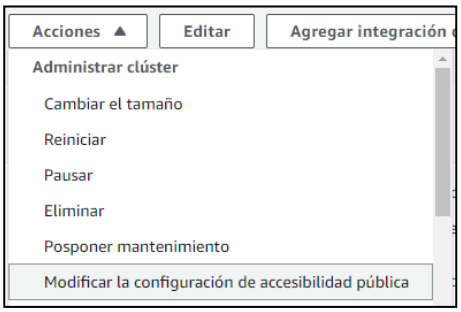


- Crea un nuevo bucket seleccionando la opción correspondiente y asigna un nombre siguiendo las reglas de AWS, optando por la región EE. UU. Este (Norte de Virginia) us-east-1.
- Deshabilita las propiedades ACL y bloquea el acceso público para garantizar la seguridad de los datos, manteniendo desactivada la opción de control de versiones.
- Selecciona la opción de cifrado SSE-S3 para los objetos almacenados y procede a la creación de carpetas dentro del bucket, aplicando el mismo cifrado.

Amazon Redshift: Gestión y Almacenamiento de Datos

- Ingresa a la consola de AWS y busca Amazon Redshift, cuyo icono se puede ver en la siguiente figura.

- Crea un clúster seleccionando la opción adecuada, asigna un nombre y elige la configuración según tus necesidades, incluyendo el tipo y número de nodos.

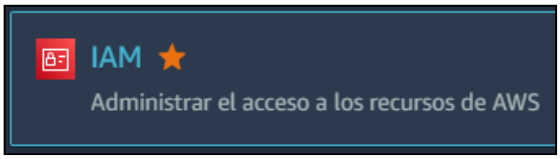


- Establece el usuario administrador y su contraseña. Una vez creado el clúster, configura la accesibilidad pública, lo cual se ilustra en la siguiente figura.

Usuarios IAM: Gestión de Identidad y Acceso

- Inicia sesión en AWS, busca IAM, y procede a la creación de una nueva política de usuario, como se muestra en la siguiente figura.

- Busca S3 en los servicios y elige el nivel de acceso. En la sección de recursos, agrega el ARN del bucket al que se dará acceso.



- Revisa y asigna un nombre a la política. Luego, crea un nuevo usuario y asigna los permisos necesarios, seleccionando la política creada anteriormente.

- Tras la creación del usuario, AWS proporcionará una URL única que debe guardarse para el acceso futuro.

Estos pasos garantizan una implementación segura y eficiente de los servicios de AWS, crucial para el manejo y análisis de datos en nuestro proyecto de inventario.

b. Presupuesto de implementación

Considerando los componentes clave que conforman la solución propuesta y con el objetivo de llevarla a un entorno de producción, se detalla a continuación una estimación presupuestaria general.

Recursos humanos

Cargo	Funciones	Salario mensual	Cantidad	Costo
Ingeniero de datos	• Desarrollo del esquema dimensional • Elaboración y ejecución de procesos ETL • Administración de los servicios de AWS	$4000.00	1	$4,000.00
Analista de datos	• Elaboración de informes utilizando Power BI • Análisis deresultados.	$2000.00	1	$2,000.00
			Total mensual	**$6,000.00**

Hardware

Equipo	Cantidad	Costo	Subtotal
PC procesador I7 12th, 16 GB RAM, 2TBSSD	2	$1,950.00	$3,900.00
		Total	**$3,900.00**

Este equipo es con las configuraciones mínimas para poder tener un rendimiento optimo en todas las operaciones que se realicen.

Licencias

Licencia	Cantidad	Costo	Subtotal
Licencia premium para Power BI	20	$20.15	$403.00
		Total mensual	**$403.00**

Servicios AWS

Según calculadora de servicios AWS

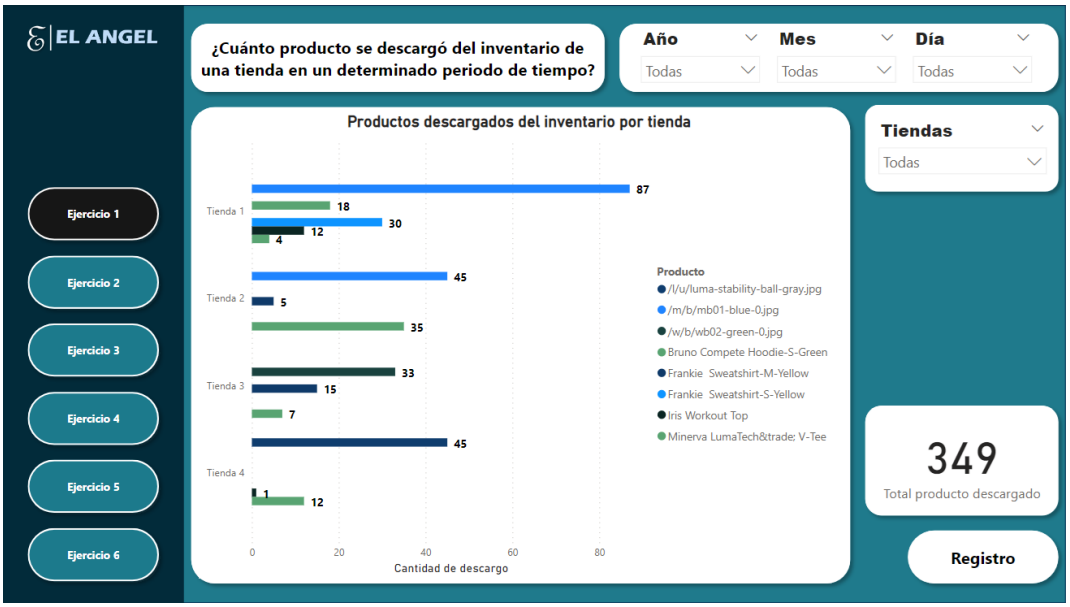
Servicio	Descripción	Costo	Subtotal
S3 Standard	• Almacenamiento y datos devueltos/escaneados para 10TB. • 20000 solicitudes PUT, COPY, POST y LIST a S3 y SolicitudesGET, SELECT desde S3.	$360.04	$360.00
Instancia de Redshift	96 GB de almacenamiento, disponible el 100% del tiempo.	$2,379.80	$2,379.80
		Total mensual	**$2,739.80**

La demanda de servicios de AWS puede variar en función de los requerimientos crecientes de almacenamiento de datos o de una mayor necesidad de acceso a los mismos, como se refleja en los componentes mencionados anteriormente

Elemento	Costo
Recursos humanos	$6,000.00
Hardware	$3,900.00
Licencias	$403.00
Servicios AWS	$2,739.84
Total	**$14,799.88**

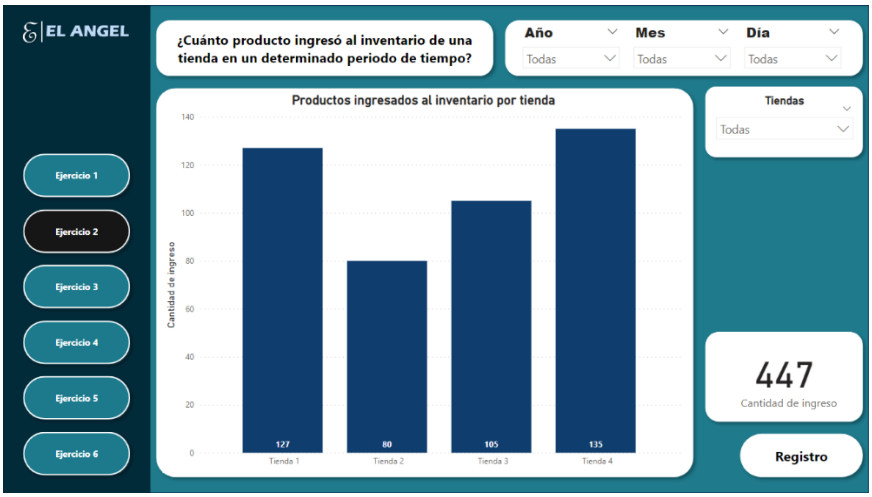
c. Análisis de resultados

Problema 1:



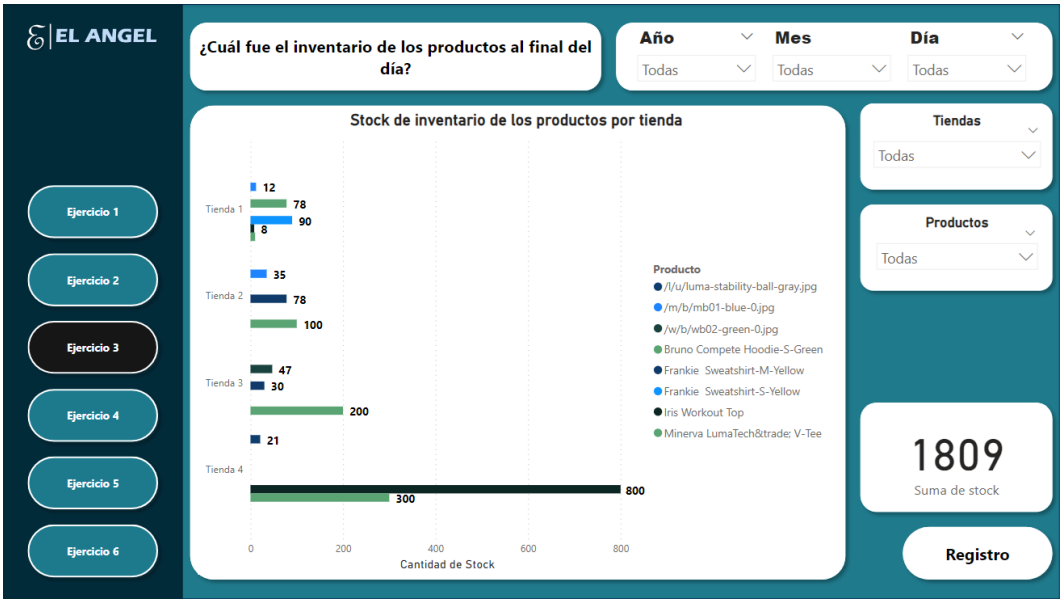
Descripción: Este dashboard nos proporciona movimientos de descarga de productos durante un periodo de tiempo brindando un análisis comparativo de cuáles son los productos que tienen un mayor movimiento por cada una de las tiendas. Se puede interactuar tanto con el día, mese, año y tienda.

Problema 2:



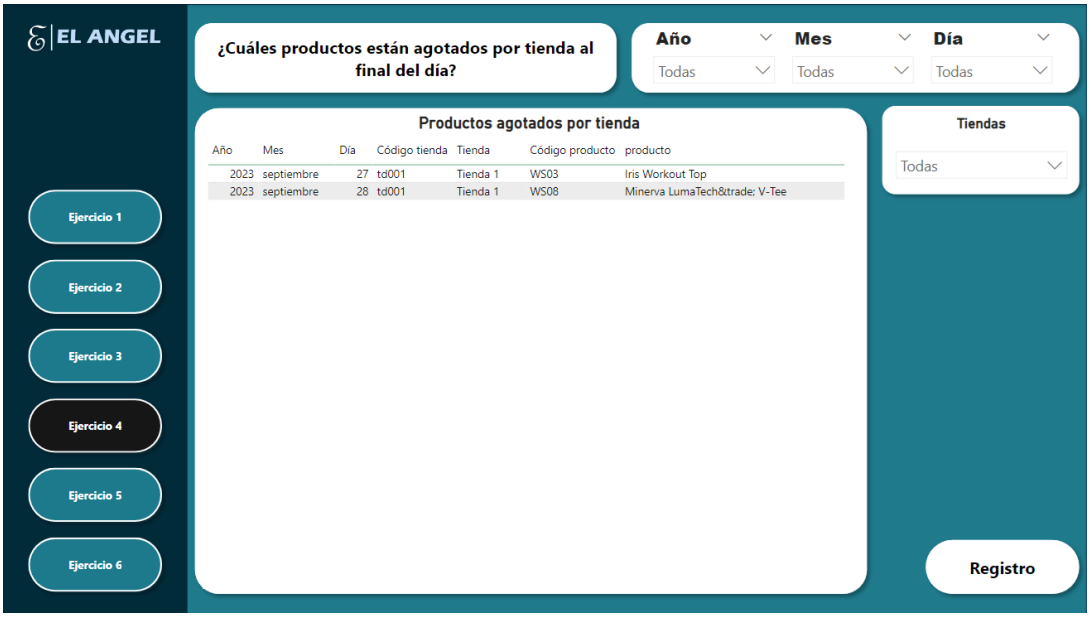
Descripción: El dashboard presentado al igual que el problema anterior muestra los movimientos de carga que se han tenido en cada tienda, al igual que se puede brindar un análisis más específico mostrando los productos de una tienda especifica. Se puede interactuar tanto con el día, mes, año y tienda.

Problema 3:



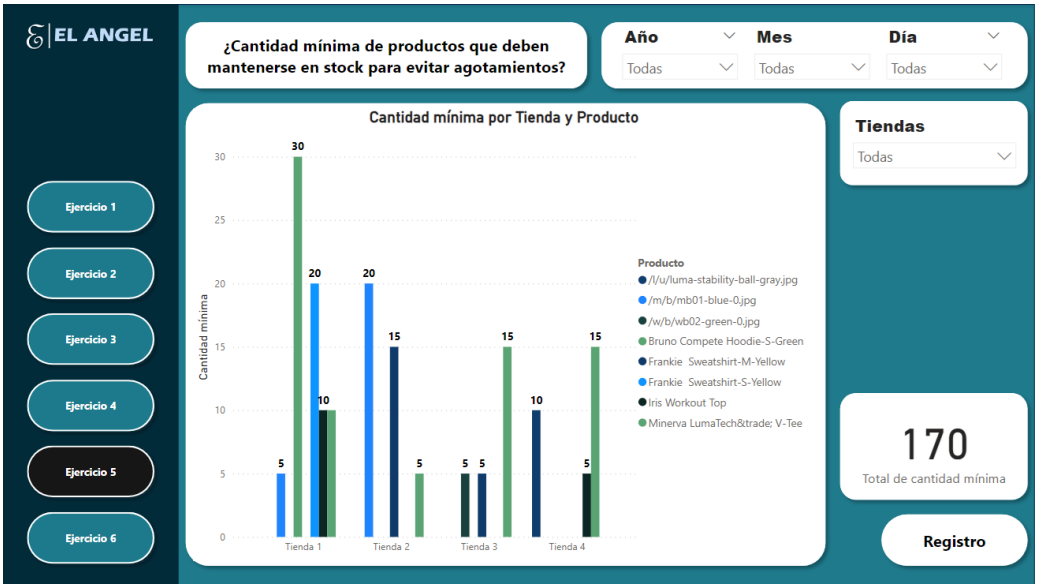
Descripción: Este dashbord nos ayudara a tener un indicador del estado final del inventario después de las cargas y descargas que este ha sufrido a lo largo del día. Se puede interactuar tanto con el día, mes, año, tienda y producto.

Problema 4:



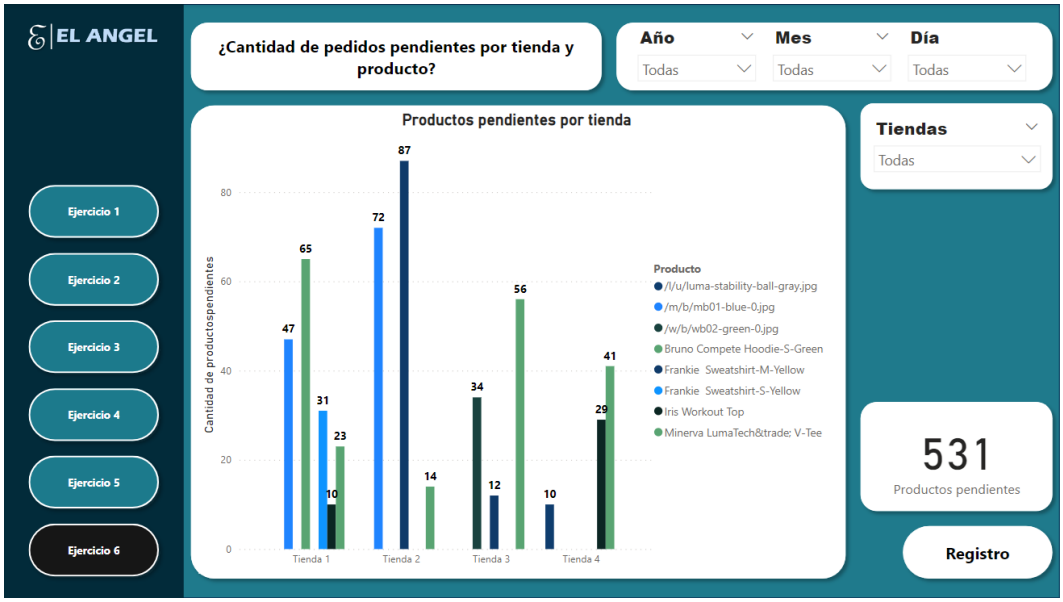
Descripción: Este dashboard nos proporcionara uno de los reportes más importantes para una empresa, ya que nos indica que productos están agotados por tienda, ayudando a tener una descripción clara y completa de que productos deben tener prontamente un ingreso. Se puede interactuar tanto con el día, mes, año y tienda.

Problema 5:



Descripción: En el dashboard presentado nos indicara la cantidad mínima que un producto posee, esta cantidad ayuda a saber si un producto está en proceso de estar en estado de agotamiento. Se puede interactuar tanto con el día, mes, año y tienda.

Problema 6:



Descripción: El dashboard mostrado indica que existen algunos productos que generaran una descarga para una fecha posterior y se debe verificar si se posee en inventario o se debe generar un ingreso para cumplir con dichos pendientes. Se puede interactuar tanto con el día, mes, año y tienda.

Conclusiones

El análisis de inventario realizado en el marco de este proyecto permitió obtener información valiosa sobre el manejo de inventario de Importaciones El Ángel. Mediante el diseño de un modelo multidimensional y la explotación de los datos almacenados en el sistema Magento Commerce, se logró analizar aspectos clave del inventario, como los niveles de stock, las categorías de productos más relevantes, la demanda por producto, entre otros.

El modelo propuesto y los informes generados proporcionan una visión detallada y actualizada del inventario, lo que facilita la toma de decisiones estratégicas, como la planificación de compras, la identificación de productos de baja rotación o la optimización de la gestión de stocks. Además, se brinda una base sólida para la generación de reportes, gráficos y análisis que permiten evaluar el desempeño del inventario y detectar oportunidades de mejora.

Recomendaciones

Recomendaciones para Futuras Implementaciones y Mejoras

A partir de los resultados obtenidos y las conclusiones del proyecto, se recomienda lo siguiente para Importaciones El Ángel:

Integración Continua de Datos: Para mantener la relevancia y precisión del análisis, es crucial asegurar una integración continua y automatizada de nuevos datos de inventario en el modelo multidimensional. Esto permitirá a Importaciones El Ángel mantenerse al día con las tendencias actuales del mercado y los cambios en la demanda de los consumidores.

Expansión del Análisis de Datos: Considerar la ampliación del análisis a otras áreas clave del negocio, como la logística de envío, la satisfacción del cliente y el análisis de ventas. Esto proporcionará una visión más holística del negocio y ayudará a identificar nuevas oportunidades de optimización y crecimiento.

Capacitación y Desarrollo del Personal: Invertir en formación continua para el equipo de Importaciones El Ángel en el uso de herramientas como Power BI y el entendimiento del modelo multidimensional. Esto potenciará la capacidad interna para realizar análisis independientes y tomar decisiones basadas en datos de manera más eficiente.

Evaluación Periódica del Modelo: Realizar evaluaciones regulares del modelo multidimensional y los procesos ETL para asegurar que sigan siendo relevantes y efectivos. Esto incluye la actualización de los modelos en función de los cambios en el negocio y el entorno del mercado.

Implementación de Alertas y Monitoreo en Tiempo Real: Desarrollar un sistema de alertas y monitoreo en tiempo real para detectar rápidamente desviaciones significativas en el inventario, como escasez de stock o cambios inesperados en la demanda.

Estas recomendaciones buscan no solo mantener la eficacia del sistema actual sino también potenciar la capacidad de Importaciones El Ángel para adaptarse a cambios futuros y aprovechar al máximo los datos disponibles para la toma de decisiones estratégicas.

Bibliografía

https://www.conectapyme.com/blog/gestion-de-inventarios-la-clave-del-exito-del-comercio-electronico/

https://rockcontent.com/es/blog/big-data-en-el-comercio-electronico/

https://www.dataprix.com/es/fases-implantacion-un-data-warehouse

https://apser.es/que-es-amazon-s3-y-cuales-son-sus-principales-beneficios/

https://powerbi.microsoft.com/es-es/pricing/

https://aws.amazon.com/es/s3/pricing/?p=pm&c=s3&z=4

Glosario

1. **Data Warehouse**: Un sistema utilizado para almacenar y gestionar grandes cantidades de datos históricos, organizados de manera que facilitan el análisis y la generación de informes. En el contexto de tu proyecto, se refiere al almacenamiento de datos de inventario y transacciones de Importaciones El Ángel.
2. **Estratégicas**: Decisiones o acciones enfocadas en objetivos a largo plazo y el logro de metas generales de la empresa. En tu proyecto, se refiere a decisiones sobre la gestión general del inventario basadas en análisis de datos.
3. **Operativas**: Decisiones o actividades diarias relacionadas con el funcionamiento habitual de la empresa. En tu proyecto, incluye tareas como la gestión diaria del stock y la respuesta a cambios en la demanda de productos.
4. **ETL (Extract, Transform, Load)**: Un proceso en la gestión de bases de datos donde los datos se extraen de varias fuentes, se transforman según sea necesario, y luego se cargan en otro sistema como un data warehouse. En tu proyecto, se usa para preparar datos para análisis desde Magento Commerce.
5. **Data Lake**: Un repositorio que almacena una gran cantidad de datos en bruto en su formato nativo hasta que se necesitan. Diferente de un data warehouse, que almacena datos estructurados y procesados. En tu proyecto, podría ser utilizado para almacenar datos de inventario en su forma más cruda.

6. **Magento Commerce**: Una plataforma de comercio electrónico que en tu proyecto se utiliza para gestionar las operaciones de venta en línea, incluyendo datos de inventario, ventas y clientes.

7. **Data Profiling**: El proceso de examinar, analizar y resumir datos de una base de datos para entender su estructura, consistencia y calidad. En tu proyecto, se utiliza para comprender mejor los datos de inventario de Importaciones El Ángel.

8. **Dataset**: Una colección de datos, generalmente presentados en formato tabular. En tu proyecto, se refiere a los conjuntos de datos de inventario, ventas y otros datos relacionados con la operación de Importaciones El Ángel.

9. **Mapping**: El proceso de definir cómo se transfieren y transforman los datos de una fuente a otra. En tu proyecto, se refiere a cómo los datos de Magento se transforman y se cargan en el sistema de análisis.

10. **Esquema Estrella**: Un modelo de diseño de base de datos utilizado en data warehouses donde una tabla de hechos central se conecta a varias tablas de dimensiones. En tu proyecto, se utiliza para organizar datos de inventario para análisis y reportes.

11. **Dashboard**: Un panel que muestra visualizaciones de datos clave como gráficos y métricas. En tu proyecto, se refiere a los paneles en Power BI que muestran información sobre el inventario.

12. **Stock**: Se refiere a la cantidad total de bienes almacenados en un punto dado. En tu proyecto, es una métrica clave relacionada con la gestión de inventario.

13. **Big Data**: Grandes conjuntos de datos que requieren sistemas avanzados para su procesamiento y análisis. En tu proyecto, puede referirse a la gran cantidad de datos de transacciones e inventario manejados.

14. **Q&A**: Siglas en inglés de "Preguntas y Respuestas". En el contexto de tu proyecto, podría referirse a la sección donde se resuelven dudas sobre la interpretación de los datos o el uso de las herramientas de análisis.

15. **EAV (Entity-Attribute-Value)**: Un modelo de base de datos utilizado para describir entidades donde el número de atributos (valores) es extenso y varía entre las entidades. En Magento Commerce, se usa para manejar la flexibilidad de los datos de productos.

Anexos

Link de enlace a repositorio GitHub

https://github.com/lordspider47/TesinaElAngel

Scripts de modelos dimensional

```
CREATE TABLE IF NOT EXISTS public.dimfecha
(
    fechakey smallint,
    fecha date,
    dia_mes smallint,
    nombre_dia character varying(255) COLLATE pg_catalog."default",
    mes smallint,
    nombre_mes character varying(255) COLLATE pg_catalog."default",
    anio smallint,
    dia_anio smallint,
    numero_semana smallint,
    semestre smallint
)

CREATE TABLE IF NOT EXISTS public.dimproducto
(
    productokey smallint NOT NULL,
    codigo character varying(255) COLLATE pg_catalog."default",
    nombre character varying(255) COLLATE pg_catalog."default",
    tiendaid smallint NOT NULL DEFAULT 1,
```

```
    costo double precision,
    CONSTRAINT dimproducto_pkey PRIMARY KEY (productokey)
)
CREATE TABLE IF NOT EXISTS public.dimtienda
(
    storekey integer NOT NULL,
    codigo character varying(32) COLLATE pg_catalog."default",
    nombre character varying(255) COLLATE pg_catalog."default" NOT NULL,
    estado integer NOT NULL
)
CREATE TABLE IF NOT EXISTS public.factinventario
(
    productokey smallint NOT NULL,
    storekey integer NOT NULL DEFAULT nextval('factinventario_storekey_seq'::regclass),
    fechakey smallint,
    stock numeric,
    cant_min numeric NOT NULL,
    cant_vendidas numeric,
    cant_devueltas numeric,
    cant_prod_pendientes smallint NOT NULL,
    CONSTRAINT factinventario_pkey PRIMARY KEY (productokey)
)
```