



# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

---

---

FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN

ELABORACIÓN DE UN DASHBOARD, APLICADO A  
LA MINERÍA DE DATOS PARA EL APOYO A LA  
TOMA DE DECISIONES EN EL DEPARTAMENTO DE  
VENTAS.

PRESENTA:

**ISMAEL MARTÍNEZ LUNA**

QUE PARA OBTENER EL TÍTULO DE:

INGENIERO INDUSTRIAL

ASESOR DE TESIS:

L.I. CARLA PAMELA PÉREZ SOLANO

CUAUTITLÁN IZCALLI, ESTADO DE MÉXICO 2024.



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## **DEDICATORIA**

A mis padres y hermana por haberme forjado como la persona que soy actualmente; muchos de mis logros se los debo a ustedes entre los que se incluye este, siempre los amaré y estaré agradecido eternamente.

A mi familia que siempre ha estado conmigo en todo momento y a pesar de todas las adversidades nunca me dejaron solo, gracias a ellos soy quien soy y me esfuerzo cada día para ser mejor.

A mi esposa por seguir motivándome día a día, su amor, sus consejos y apoyo incondicional que me da ánimos para seguir adelante, te amo y te amaré por siempre, gracias.

A mi asesora de tesis, la licenciada Carla Pamela Pérez Solano, por haberme guiado en este proyecto el cual no habría sido posible sin ella, con base a su experiencia y sabiduría que ha sabido direccionar mis conocimientos.



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

DR. DAVID QUINTANAR GUERRERO  
DIRECTOR DE LA FES CUAUTITLÁN  
PRESENTE

FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN  
SECRETARÍA GENERAL  
DEPARTAMENTO DE TITULACIÓN



ASENTO: VOTO APROBATORIO

UNAM

ATN: DRA. MARÍA DEL CARMEN VALDERRAMA BRAVO

Clasificación: Departamento de Titulación  
de la FES Cuautitlán.

Con base en el Reglamento General de Exámenes, y la Dirección de la Facultad, nos permitimos comunicar a usted que revisamos el trabajo de: **Tests**.

Elaboración de un dashboard, aplicado a la minería de datos para el apoyo a la toma de decisiones en el departamento de ventas.

Que presenta el pasante: **Ismael Martínez Luna**  
Con número de cuenta: **412058460** para obtener el Título de: **Ingeniero Industrial**.

Considerando que dicho trabajo reúne los requisitos necesarios para ser discutido en el **EXAMEN PROFESIONAL** correspondiente, otorgamos nuestro **VOTO APROBATORIO**.

**ATENTAMENTE**  
"POR MI RAZA HABLARÁ EL ESPÍRITU"  
Cuautitlán Izcalli, Méx. a 27 de septiembre de 2023.

PROFESORES QUE INTEGRAN EL JURADO

	NOMBRE	FIRMA
<b>PRESIDENTE</b>	Mtra. Maritza Nova Hernández	
<b>VOCAL</b>	Ing. Rocio Cruz Chávez	
<b>SECRETARIO</b>	Lic. Carla Pamela Pérez Solano	
<b>1er. SUPLENTE</b>	Dr. Raúl Dalí Cruz Morales	
<b>2do. SUPLENTE</b>	Mtro. Gonzalo Hedain López Mera	

NOTA: los sinodales suplentes están obligados a presentarse el día y hora del Examen Profesional.

MCVb/nsm\*

# ÍNDICE

RESUMEN.....	8
INTRODUCCIÓN .....	10
OBJETIVO GENERAL .....	12
OBJETIVOS PARTICULARES .....	12
PROBLEMÁTICA .....	13
HIPÓTESIS .....	13
CAPÍTULO I: MARCO TEÓRICO .....	14
1.1 Bases de datos. ....	14
1.2 Clasificación de las bases de datos. ....	15
1.2.1 Datawarehouse. ....	16
1.2.2 Bases de datos multidimensionales. ....	18
1.2.2.1 Estrella simple. ....	20
1.2.2.2. Copo de nieve. ....	21
1.2.2.3 Datamart. ....	22
1.2.2.3.1 Objetivo. ....	23
1.2.2.3.2 Ventajas. ....	24
1.2.2.3.3 Desventajas. ....	25
1.2.2.4 Constelación.....	26
1.3 Herramienta de desarrollo.....	27
A) Modelo conceptual e-r.....	27
B) Modelo lógico relacional.....	29
C) Modelo Físico.....	30

1.4 Sistema de base de datos.....	30
1.5 Sistema gestor de base de datos (SGBD).....	32
1.5.1 SGBDR – MariaDB. ....	32
CAPÍTULO 2: METODOLOGÍA. ....	34
2. Metodología KDD.....	34
2.1 Fases de la metodología.....	35
2.1.1 Fase de integración y recopilación:.....	36
2.1.2 Fase de selección, limpieza y transformación. ....	39
2.1.2.1 ETL.....	39
2.1.3 Fase de minería de datos. ....	40
2.1.3.1 Series de tiempo. ....	41
2.1.3.2 Promedios móviles simples.....	47
2.1.3.3 Coeficiente de Correlación. ....	47
2.1.4 Fase de Evaluación e interpretación. ....	48
2.1.5 Fase de difusión, uso y monitorización. ....	48
2.2 Dashboard.....	49
2.2.1 Pasos para la creación de dashboard:.....	49
2.2.2 Ventajas y desventajas. ....	50
2.2.3 Herramientas de Desarrollo. ....	51
a) Instalación del programa Día. ....	51
b) Instalación del programa XAMPP.....	54
d) Instalación del programa MySQL.....	57
e) Descarga del connector odbc. ....	60
f) Descarga el programa Power BI. ....	62

g) Descarga el programa Javajdk.....	64
h) Descarga el programa: Talend Open Studio.....	67
CAPÍTULO 3. CASO PRÁCTICO DISEÑO DE LA SOLUCIÓN. ....	70
3. Contexto del negocio. ....	70
3.1 Propuesta de solución.....	71
3.2 Arquitectura de solución.....	71
3.3 ETL.....	72
3.4 Diseño de la base de datos.....	92
B) Creación de las entidades. ....	92
C) Generación de los atributos para cada entidad. ....	92
D) Constraints. ....	93
E) Integridad de la entidad (Primary Key).....	93
F) Llaves primarias y foráneas. ....	94
G) Integridad referencial (Foreign Key).....	94
H) Llaves foráneas.....	95
3.5 E-R (Entidad-Relación).....	95
3.6 Relacional.....	97
3.7 Modelado Físico. ....	105
3.8 Dashboard.....	112
a) Vistas.....	112
3.9 Modelado de datos. ....	119
3.10 Construcción del dashboard. ....	120
3.11 Modelos de minería de datos. ....	134
RESULTADOS.....	136

ANÁLISIS DE RESULTADOS. ....	140
DISCUSIÓN DE RESULTADOS. ....	147
CONCLUSIONES.....	151
BIBLIOGRAFÍA: .....	152

## RESUMEN

Hoy en día muchas organizaciones aún trabajan de forma tradicional, en donde todavía no involucran nuevas e innovadoras tecnologías que ayuden a automatizar y optimizar procesos, y que a su vez maximicen la eficiencia, poniendo en riesgo su productividad y permanencia en el mercado.

Como consecuencia se manifiestan ciertas problemáticas especialmente en el manejo y gestión de los datos para la obtención de resultados precisos. Por lo dicho, este trabajo va a versar sobre la correcta aplicación de ciertas herramientas informáticas modernas que garanticen la construcción de un dashboard, utilizando una metodología llamada KDD (*Knowledge Discovery in Databases*) acompañada de un proceso ETL (*Extract Transform Load*), donde se aplicarán modelos estadísticos de la minería de datos a partir de una vista minable, recopilada y centralizada en una base de datos relacional multidimensional. Para lograr un mejor tratamiento a los datos dando como resultado la extracción de conocimiento, para la obtención de reportes confiables y consistentes, dirigidos a los altos mandos permitiéndoles una mejor toma de decisiones. Tratando de perfeccionar los procesos que se manejan actualmente en el área de ventas en una organización encargada del marketing y promotoría de productos de diferentes marcas con la finalidad de impulsar las ventas.

El contenido de esta tesis se distribuye en tres capítulos, el primero explica la definición de una base de datos, así como su clasificación, poniendo mayor énfasis en las bases de datos multidimensionales, mostrando su modelo y las distintas arquitecturas, así como la utilizada para la construcción e implementación de este proyecto (datamart), acompañado de las herramientas de desarrollo para su creación desde la modelación hasta la codificación con el uso de un sistema gestor de base de datos y se expone el concepto de sistema de base de datos y se expone el concepto de sistema de base de datos

El segundo capítulo detalla la metodología a utilizar llamada KDD en sus cinco fases, acompañada del proceso ETL. Se exponen la técnica estadística de minería de datos aplicada en la fase tres, una vez cargada la data en la base de datos multidimensional, adicional se muestran las herramientas de desarrollo, así como su instalación y se redacta una descripción, de lo que es un tablero de control.

El capítulo final muestra el caso práctico donde se detalla el contexto de la empresa, la propuesta y arquitectura de solución, además del uso de las herramientas informáticas usadas para el diseño y creación de la base de datos, la carga de datos a la misma mediante ETL, permitiendo tener la información disponible para analizar y aplicar técnicas de minería de datos, alimentando modelos estadísticos que

generaran resultados de interés para el negocio de forma resumida mediante gráficos.

## **INTRODUCCIÓN**

En la actualidad la tecnología a generado cambios en la sociedad y las empresas las cuales han tenido que adaptarse a esta evolución, pues muchas de sus actividades dependen de ella, teniendo como propósito cumplir con ciertos objetivos de negocio estratégico, entre ellos, la excelencia operacional, la creación de nuevos productos y servicios, la innovación de nuevos modelos de negocio, manejar una mejor intimidad con los clientes, conservar una ventaja competitiva, una mejora continua y mantener una correcta toma de decisiones.

Existen factores que obstaculizan esta transformación, como es la resistencia al cambio, la falta de infraestructura y recursos, la nula formación del personal, no contar con procesos eficientes que generen datos correctos, la carencia de visión por parte de los directivos y el personal, el no implementar nuevas tecnologías, etc. Sin embargo, las empresas se ven en la necesidad de incorporarse a esta transición para estar siempre a la vanguardia, afrontando los nuevos retos para sobrevivir dentro de la industria. En ocasiones implementando un conjunto de herramientas informáticas como el uso de software específico.

Este trabajo apoyará al departamento de ventas, establecido en una organización que se encarga del marketing y la promoción de productos de diferentes marcas activando el impulso de las ventas. Con la finalidad de implementar un grupo de herramientas digitales que ayuden a la construcción de un dashboard aplicado a la minería de datos siguiendo una metodología adecuada. Haciendo uso de una base de datos relacional que contendrá la vista minable.

Este tablero de control auxiliará en la automatización e innovación de procesos operativos claves, así como en la retroalimentación con otros sistemas, a partir de ciertos reportes e informes para analizar y hacer proyecciones para el negocio de acuerdo a los resultados obtenidos, ofrecer soluciones adaptadas a las necesidades de cada cliente, optimizar los tiempos para brindar un mejor servicio, checar fácilmente el estado y las etapas en que se encuentra cada venta y lo principal, apoyará a la toma de decisiones, tratando de evitar deficiencias y datos erróneos.

## **OBJETIVO GENERAL**

Implementar un tablero de control que incorpore algunas técnicas estadísticas de minería de datos, que ayuden a gestionar la información del área de ventas en una empresa de servicios, con el propósito de encontrar patrones, tendencias, reglas en el comportamiento de los datos, así como visualizar indicadores clave de desempeño y métricas para la toma de decisiones que llevarán a la organización a lograr sus objetivos.

## **OBJETIVOS PARTICULARES**

Generar un datawarehousing para el área de ventas con la finalidad de organizar, comprender y utilizar la información implementando herramientas tecnológicas para la automatización de las bases de datos con la finalidad de minimizar tiempos de entrega de reportes e indicadores de rendimiento.

## **PROBLEMÁTICA**

El presente trabajo versa en la automatización de informes, resultados y proyecciones generados en el departamento de ventas del año 2023 en una empresa de mercadotecnia de donde se sustenta esta tesis. Ya que se han reflejado ciertas problemáticas en la consulta, análisis y gestión de la información con la inexistencia de una base de datos. Al elaborar reportes ineficientes e incongruentes, generando conflictos con los clientes, por la inconsistencia de la información, lo que provoca una salida errónea y una baja calidad de los resultados, evitando llegar a una buena toma de decisiones al momento requerido.

## **HIPÓTESIS**

De acuerdo con la problemática observada en la organización, la implementación de un dashboard mediante la metodología KDD, el proceso ETL y modelos estadísticos de avanzado nivel (minería de datos) ayudará a optimizar los procesos de análisis y gestión de la información, logrando una mejora continua en el tratamiento y análisis de los datos para una mejor toma de decisiones de acuerdo con los resultados obtenidos, a través de ciertos indicadores claves, desde la recopilación de los datos hasta la entrega final de los reportes presentados en forma gráfica y resumida.

## **CAPÍTULO I: MARCO TEÓRICO**

### **1.1 Bases de datos.**

Una base de datos se define como:

“Un conjunto de datos homogéneos relacionados entre sí, que se encuentran estructurados de forma fácil y de rápido acceso, para el manejo y actualización de la información” (ELEMENTOS DE UNA BASE DE DATOS, 2020, pág. 1)

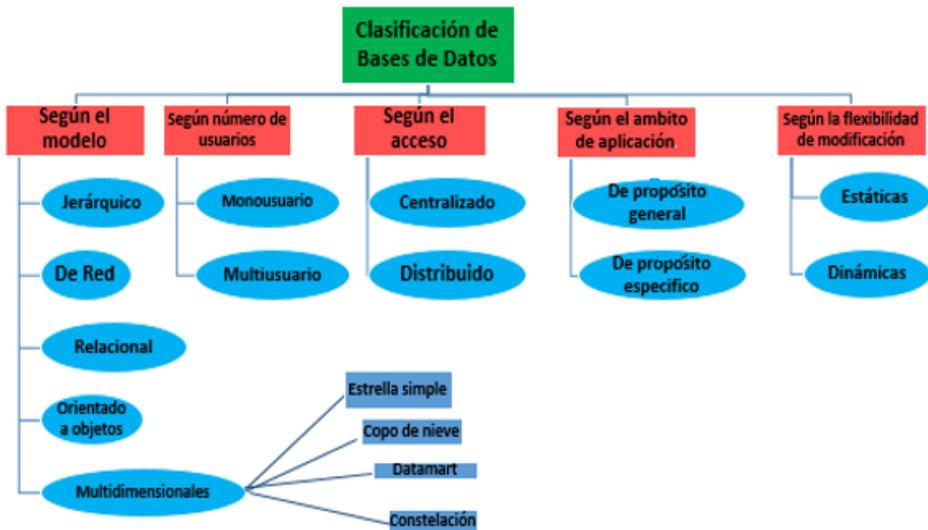
Por otra parte, los autores Miguel y Mario Pattini, precisan que es una:

“Colección o depósito de datos integrados, almacenados en soporte secundarios (no volátil) y con redundancia controlada. Los datos, que han de ser compartidos por diferentes usuarios y aplicaciones, deben mantenerse independientes de ellos y su definición (estructura del BD), única y almacenada junto con los datos, se ha de apoyar en un modelo de datos, el cual ha de permitir captar las interrelaciones y restricciones existentes en el mundo real. Los procedimientos de actualización y recuperación, comunes y bien determinados, facilitarán la seguridad del conjunto de los datos” (Mora Rioja Arturo. Base de datos Diseño y Gestión. Edit. Síntesis. 2014.Pdf. pág. 2)

A partir de las definiciones anteriores podemos deducir que una base de datos es un conjunto de datos interrelacionados entre sí con el objetivo de almacenar información y poder acceder a ella cuando sea necesario y pertinente.

## 1.2 Clasificación de las bases de datos.

Existe una clasificación de bases de datos en función del modelo de datos, el número de usuarios, su acceso, su ámbito de aplicación o su flexibilidad de modificación, etc. A continuación, podemos observar esta distribución (Figura 1):



**Figura 1.** *Clasificación de la base de datos según el modelo.*

Nota. El organizador grafico representa la clasificación de Bases de datos. Creación propia.

Posteriormente se pondrá mayor interés en las bases de datos multidimensionales ya que en el presente proyecto se diseña una base de datos multidimensional con arquitectura tipo datamart.

### 1.2.1 Datawarehouse.

Bill Inmon y R. D. Hackathorn, “El almacén de datos es una colección de datos orientados al tema, integrados, no volátiles e historiado, , organizados para ofrecer apoyo a procesos de ayuda de decisión” (Gamazo. A. G. Introducción a la Data Warehouse.2016. pág. 7)

Las características más relevantes son:

- **Orientado a temas.** - Los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.
- **Integrado.** - La base de datos contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes.
- **No volátil.** - La información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en

información de sólo lectura, y se mantiene para futuras consultas.

- **Variante en el tiempo.** - Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones. (¿Qué es el data warehouse?. Blog de CEUPE <https://www.ceupe.com/blog/que-es-el-datawarehouse.html>)

José Hernández Orallo menciona que “La tecnología de bases de datos ha respondido a este reto con una nueva arquitectura surgida recientemente: el almacén de datos (data warehouse). Se trata de un repositorio de fuentes heterogéneas de datos, integrados y organizados bajo un esquema unificado para facilitar su análisis y dar soporte a la toma de decisiones”. (Hernández Orallo J. (2004). Introducción a la minería de datos, 2004. Pearson educación libgenli compress.pdf pag.4)

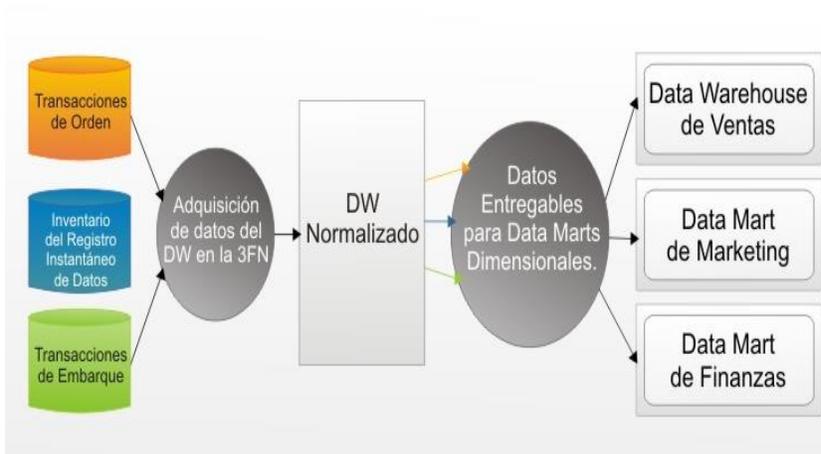
Por lo tanto, se llega a la conclusión que una data warehouse funciona como un almacén de datos. La información proviene de una o varias fuentes, siendo el proceso para transformar datos en información valiosa para la toma de decisiones.

## 1.2.2 Bases de datos multidimensionales.

Marysol Tamayo & Francisco Javier Moreno describen que “Dentro de la estructura de una base multifuncionales se pueden encontrar diversos almacenes de datos los cuales se encuentran organizados en hechos, medidas y dimensiones” (J. T. M., 2006)

- a) **Hechos:** Es el objeto analizar, posee atributos de tipo cuantitativo llamados de hechos o de síntesis, sus valores se obtienen generalmente por la aplicación de una función estadística que resume un conjunto de valores en un único valor.
- b) **Medidas:** En las tablas de hechos se almacenan las medidas de los datos. Estas medidas están formadas por dos componentes principales:
  - Una propiedad numérica o valor, como puede ser el precio de venta de un producto.
  - Una fórmula, de forma que se puedan combinar distintos valores de medidas
- c) **Dimensión:** representa cada uno de los ejes en un espacio multidimensional, en el que se obtienen las medidas de un hecho. Algunos ejemplos son: tiempo, producto, cliente, etc. Las dimensiones se utilizan para seleccionar y agrupar los datos en un nivel de detalle

deseado. Los componentes de una dimensión se denominan niveles y se organizan en jerarquías, y dimensión del tiempo (día, mes, año) como se muestra a continuación



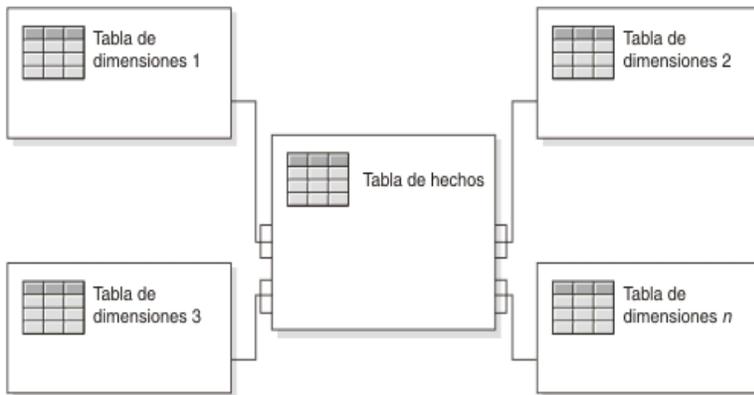
**Figura 2.** Esquema de organización de bases multidimensionales.

Nota. La figura representa la estructura de una base multifuncionales se pueden encontrar diversos almacenes de datos los cuales se encuentran organizados en hechos, medidas y dimensiones. Tomada de José Hernández Orallo. Introducción a la minería de datos 2004, Pearson educación libgenli compress.

Estas bases de datos se clasifican en estrella simple, copo de nieve, datamart y constelación.

### 1.2.2.1 Estrella simple.

Es un tipo de esquema de base de datos relacional que consta de una sola tabla de hechos central rodeada de tablas de dimensiones. Un esquema de estrella puede tener cualquier número de tablas de dimensiones. Las ramas situadas al final de los enlaces que conectan las tablas indican una relación de muchos a uno entre la tabla de hechos y cada tabla de dimensiones. (IBM, InfoSphere Data Architect, s.f.)

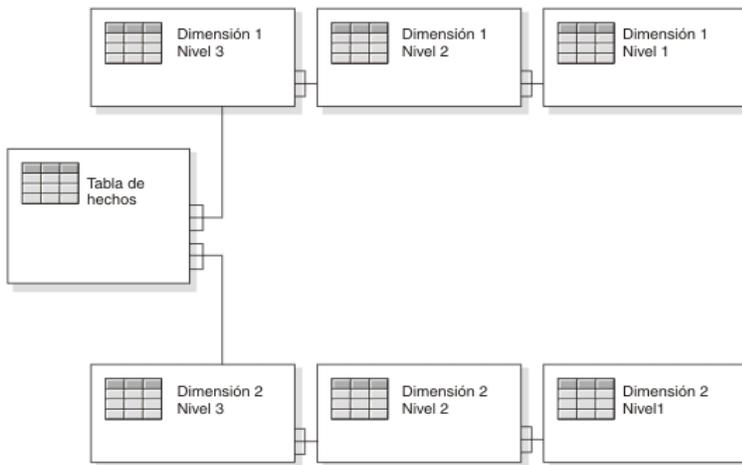


**Figura 3.** *Esquema de estrella simple con una sola tabla de hechos*

Nota: Esquema de estrella simple y su distribución. Tomada de <https://www.ibm.com/docs/es/ida/9.1.2?topic=schemas-star>

### 1.2.2.2. Copo de nieve.

El esquema de copo de nieve consta de una tabla de hechos que está conectada a muchas tablas de dimensiones, que pueden estar conectadas a otras tablas a través de una relación de muchos a uno. Las tablas de un esquema de copo de nieve generalmente se normalizan en el tercer formulario de normalización. Cada tabla de dimensiones representa exactamente un nivel en una jerarquía. (IBM, InfoSphere Data Architect, s.f.)

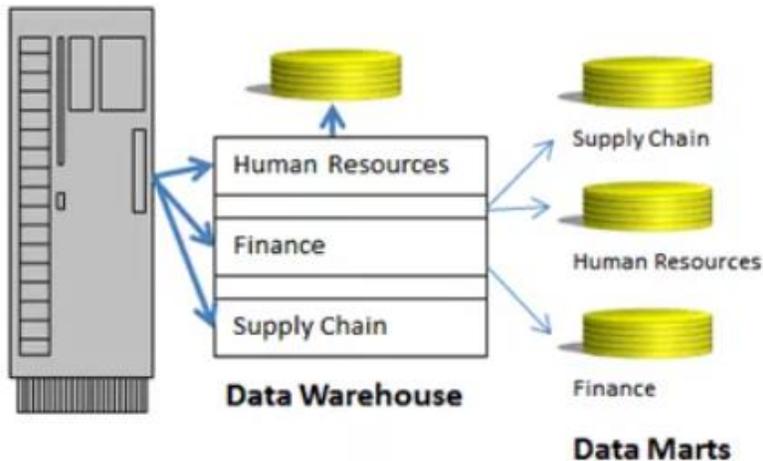


**Figura 4.** Esquema de copo de nieve con dos dimensiones y tres niveles cada una.

Nota: Esquema de copo de nieve con dos dimensiones, cada una con tres niveles. Un esquema de copo de nieve puede tener varias dimensiones y cada dimensión puede tener varios niveles. Tomada de <https://www.ibm.com/docs/es/ida/9.1.2?topic=schemas-snowflake>

### 1.2.2.3 Datamart.

Subconjunto del almacén de datos que se ocupa de un solo asunto. A menudo son construidos y administrados por un solo departamento comercial, normalmente toman datos de solo una pequeña cantidad de fuentes, que podrían ser sistemas operativos internos, un lago de datos, un depósito de datos centralizado o fuentes externas. Están condensados y son menos complejos que los almacenes de datos, lo que los hace más fáciles de construir y mantener mostrado en la figura 5. Tehreem N. (2023). ¿Qué es un datamart?. Recuperado el 19 de diciembre de 2023 de <https://www.astera.com/es/type/blog/types-of-data-marts/?nonitro=1>



**Figura 5.** Esquema de organización datamart

Nota: Repositorio de datos específico del departamento para su negocio específico. (<https://www.astera.com/es/type/blog/types-of-data-marts/?nonitro=1>).

### 1.2.2.3.1 Objetivo.

Incluye una subsección de datos de toda la empresa, que es valiosa para un grupo de usuarios en particular en la organización. A diferencia de un almacén de datos que es costoso y complejo de crear, ofrece una alternativa rentable. También permite un acceso más rápido a los datos y es fácil de usar, ya que está diseñado con precisión de acuerdo con los

requisitos de los operadores y se centra en un solo departamento / área temática.

Puede ayudar a acelerar sus procesos corporativos, ya que lleva menos tiempo implementarlo que una data warehouse. También, incluye datos pasados para que sus analistas puedan determinar fácilmente las tendencias de los datos.

#### **1.2.2.3.2 Ventajas.**

- Permite un acceso más rápido recuperando un conjunto específico de datos para BI y de informes. Como resultado, ayuda a acelerar la inteligencia empresarial.
- Son más fáciles de implementar y más rentables que construir una empresa.
- De acuerdo con los requisitos de un grupo particular de usuarios que trabajan en un departamento específico.
- Son comparativamente más adaptables que un almacén de datos. Cualquier cambio en el modelo de datos se puede incorporar fácil y rápidamente al datamart debido a su tamaño más pequeño.
- Derechos de control de acceso granulares debido a la amplia partición y segmentación.
-

### 1.2.2.3.3 Desventajas.

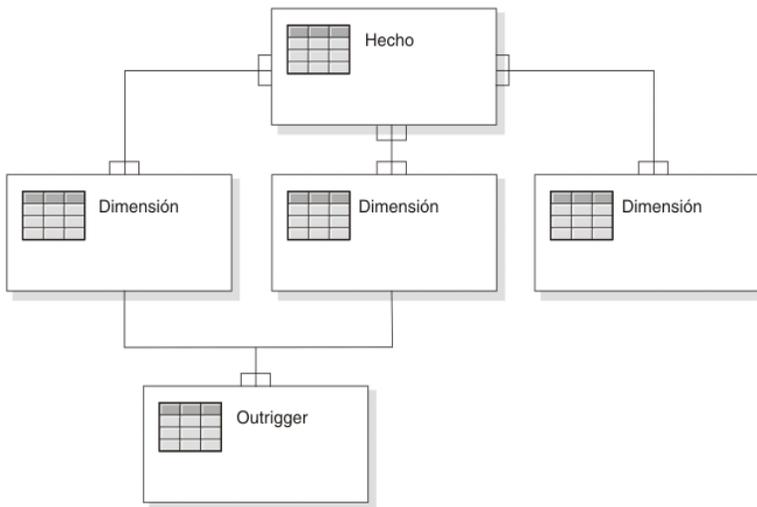
Uno de los problemas que presentan los almacenes de datos, es que al crecer el rendimiento de sus consultas el modelo centralizado deja de ser óptimo.

Datamart se clasificar en tres tipos principales:

1. **Centro de datos dependiente:** Un centro de datos dependiente le permite combinar todos sus datos comerciales en un solo almacenamiento de datos, ofreciéndole los beneficios típicos de la centralización.
2. **Centro de datos independiente:** Se puede crear una despensa de datos independiente sin utilizar el almacén de datos central. Se recomienda principalmente para unidades o grupos más pequeños dentro de una organización.
3. **Centro de datos híbrido:** Mediante el uso de una despensa de datos híbrida, puede combinar datos de varios sistemas de origen operativo además de un almacén de datos. Estos son particularmente útiles cuando necesita una integración ad hoc, como agregar un nuevo grupo o productos a la empresa.

#### 1.2.2.4 Constelación.

“Los esquemas de constelación son esquemas de copo de nieve en los que sólo algunas de las tablas de dimensiones se han desnormalizado, el objetivo es eliminar las redundancias de las dimensiones” ([https://www.ibm.com/docs/es/ida/9.1.2?topic=sc\\_hemas-starflake](https://www.ibm.com/docs/es/ida/9.1.2?topic=sc_hemas-starflake)). Para normalizar el esquema, las jerarquías dimensionales compartidas se colocan en outriggers (Figura 6).



**Figura 6.** Esquema de constelación con un hecho y dos dimensiones que comparten un outrigger.

Nota: Se muestra la combinación de un esquema de estrella y un esquema copo de nieve. Tomado de <https://www.ibm.com/docs/es/ida/9.1.2?topic=schemas-starflake>

### **1.3 Herramienta de desarrollo.**

A continuación, se describen los modelos de datos y sus herramientas para su construcción y desarrollo de la arquitectura datamart.

#### **A) Modelo conceptual e-r.**

Este modelo está basado en una percepción del mundo real el cual consiste en un conjunto de objetos básicos, denominados entidades, y de las relaciones entre esos objetos, las entidades se describen en la base de datos, mediante un conjunto de atributos; una relación es una asociación entre varias entidades.

A su vez estas entidades dentro de los datos se describen mediante un conjunto de atributos. Por ejemplo, los atributos número-cuenta y saldo pueden describir una cuenta concreta de un banco y constituyen atributos del conjunto de entidades cuenta, mientras que una relación es una asociación

entre varias entidades. Por ejemplo, la relación impositora asocia un cliente con cada cuenta que tiene. Silverschatz, K, S. (2006) Fundamentos de bases de datos, España: Mc Graw Hill. Pág. 13

La estructura lógica general de la base se puede expresar gráficamente mediante un diagrama e-r, que se compone por los siguientes componentes:

- ✓ Rectángulo, el cual representa conjuntos de entidades.
  - ✓ Elipses, representa los atributos.
  - ✓ Rombos, muestra conjuntos de relaciones entre miembros de varios conjuntos de entidades.
- Líneas, su función es unir los atributos con los conjuntos de entidades entre sí, y también los conjuntos de entidades con las relaciones.

#### **a. Programa Día.**

Día es una aplicación informática de propósito general para la creación de diagramas, creada originalmente por Alexander Larsson, y desarrollada como parte del proyecto GNOME . Está concebido de forma modular, con diferentes paquetes de formas para diferentes necesidades. En este caso

se usó la versión dia.exe 0.97.2 utilizando las herramientas para la construcción de diagramas E-R. (Día programa. Wikipedia la enciclopedia libre (27 jul 2022). Dia (programa) - Wikipedia, la enciclopedia libre)

## **B) Modelo lógico relacional.**

El modelo lógico contiene representaciones de entidades, atributos, relaciones, identificadores exclusivos, subtipos, súper tipos y restricciones entre relaciones de datos, un modelo lógico también puede contener objetos de modelo o hacer referencia a uno o más modelos. Una vez definidas las relaciones y los objetos lógicos en un modelo de datos, puede utilizar el entorno de trabajo para transformarlo en una representación física específica de base de datos en forma de modelo de datos físicos.

Los objetos del modelo lógico siempre están contenidos en un objeto de paquete raíz, pero puede añadir paquetes adicionales bajo el paquete raíz para agrupar objetos similares. Modelos de datos dimensionales lógicos (08-03-21-IBM). Modelos de datos dimensionales lógicos. Tomado de IBM <https://www.ibm.com/docs/es/ida/9.1.2?topic=design-logical-dimensional-data-models>

## **b. MySQL Workbench.**

“MySQL Workbench es una herramienta visual de diseño de bases de datos que integra desarrollo de software, administración de bases de datos, diseño de bases de datos, gestión y mantenimiento para el sistema de base”  
([https://es.wikipedia.org/wiki/MySQL\\_Workbench](https://es.wikipedia.org/wiki/MySQL_Workbench))

## **C) Modelo Físico.**

Dentro del proyecto se aplicará el modelo físico, siendo una representación del modelo para la construcción de la base de datos (código SQL). Usando esta herramienta para generar el código a implementar.

## **1.4 Sistema de base de datos.**

**Sistema de base de datos:** Conjunto de bases de datos a las cuales se les aplica cierto proceso para la toma de decisiones. Los elementos principales de este sistema de base de datos son:

**Hardware:** Se refiere a los dispositivos de almacenamiento físico, en los que puede estar una base de datos. Ejemplos: Disco duro, USB, CD, DVD, etc. Y por lo tanto se necesita de ellos para tener una base de datos.

**Software.** Se refiere a todo aquel conjunto de programas que interactúan con la base de datos; y los cuales se dividen en tres grupos importantes.

**Sistema Operativo:** Esto es el ambiente en el cual se desenvuelve la base de datos. Ejemplos: Linux, Unix, Windows, Mac-Os, etc.

**Manejadores de bases de datos:** Son los programas que controlan y administran los que diferentes tipos de bases de datos. Ejemplos: Oracle, DB2, informix, PostgreSQL, Firebird, SQL Server, etc.

**Herramientas CASE:** Las herramientas CASE son programas que ayudan a una tarea en específico de la informática, en este caso será del modelo de base de datos.

**Usuarios.** Son todas aquellas personas que interactúan con la base de datos.

**Datos.** Dependiendo el modelado de datos que se escoja, es la estructura que nosotros visualizaremos la información. Estos modelos de datos se analizarán en la parte de evolución de las bases de datos.

## 1.5 Sistema gestor de base de datos (SGBD).

Es un conjunto de programas que permiten la implantación, acceso y mantenimiento de la base de datos. Siendo una serie de rutinas de software interrelacionadas, encargadas de realizar tareas específicas.

Funciones:

- **Definir.** Utiliza un lenguaje de definición de datos (LDD) para especificar la estructura, restricciones y tipos de datos.
- **Construir:** Almacenar los datos en un medio controlado por SGBD.
- **Manipular.** Se deben de poder ejecutar consultas de eliminación, actualización, inserción (LMD) o por un lenguaje de consultas.
- **Compartir.** El sistema debe permitir el acceso a varios usuarios a los cuales deben poder acceder al mismo tiempo.

### 1.5.1 SGBDR – MariaDB.

“MariaDB es un sistema de gestión de bases de datos derivado de MySQL con licencia GPL (General Public License). Es desarrollado por Michael (Monty) Widenius —fundador de MySQL—, la fundación MariaDB y la comunidad de

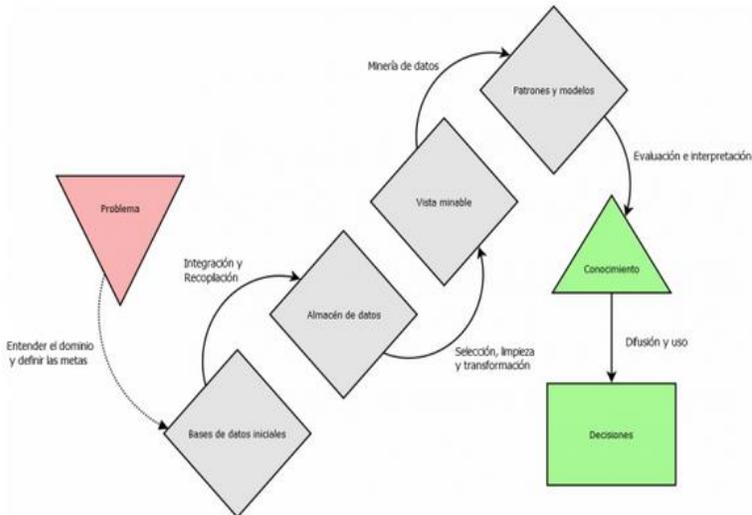
desarrolladores de software libre.<sup>1</sup> Introduce dos motores de almacenamiento nuevos, uno llamado Aria —que reemplaza a MyISAM— y otro llamado XtraDB —en sustitución de InnoDB—. Tiene una alta compatibilidad con MySQL ya que posee las mismas órdenes, interfaces, API y bibliotecas, siendo su objetivo poder cambiar un servidor por otro directamente” para la gestión del proyecto se utiliza Server versión: 10.4.24 - MariaDB. ([MariaDB - Wikipedia, la enciclopedia libre. https://es.wikipedia.org/wiki/MariaDB](https://es.wikipedia.org/wiki/MariaDB))

## **CAPÍTULO 2: METODOLOGÍA.**

### **2. Metodología KDD.**

Para el desarrollo de este proyecto se utiliza una metodología sistemática e iterativa llamada KDD, esta nos ayudara al proceso de descubrir conocimiento de las bases de datos. Tomando como referencia a los escritores Hernández Orallo José, Ramírez Quintana M. José y Ferri Ramírez César, complementada por cinco fases. Que se explican a continuación.

Antes de iniciar el proceso es necesario preparar una visión general, planificando las tareas que se van a realizar y de qué manera. Para ejecutar este proceso KDD se debe entender el entorno del proceso de conocimiento para que este se realice correctamente, a su vez conocer el contexto de cada organización, el problema en particular que puede determinar qué información se necesitará en la siguiente fase mostrado en la Figura 7.



**Figura 7.** Metodología KDD.

Nota: Estructura y elementos del proceso KDD. Tomado de <https://diagramasuml.com/que-es-el-kdd-o-proceso-de-descubrimiento-de-conocimiento/>

## 2.1 Fases de la metodología.

El desarrollo de la metodología KDD consta de diversas fases, esto varía según cada autor, por ejemplo, la agrupación, aunque en esencia todos cubren la misma metodología. Las fases son las siguientes:

### 2.1.1 Fase de integración y recopilación:

En esta fase se reconocen, y reúnen los datos con los que se trabajará en el futuro para su análisis, estos datos pueden obtenerse de dos fuentes diferenciadas:

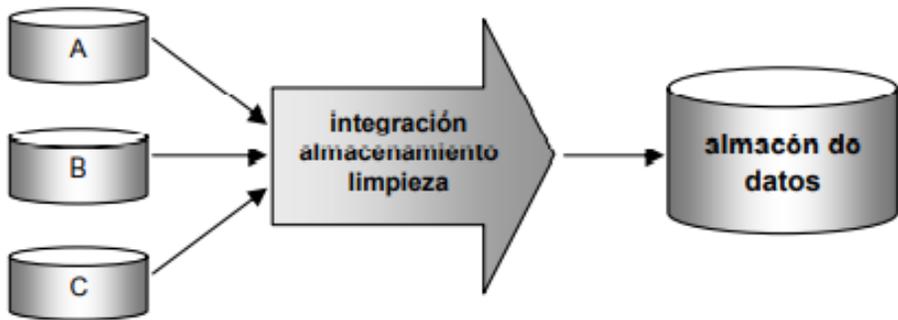
- a) **Fuentes internas:** Los datos son más sencillos de obtener y representa las bases de datos transaccionales propias de la organización, utilizadas tradicionalmente para el procesamiento transaccional en línea u OLTP (On-Line Transaction Processing).
- b) **Fuentes externas:** Representan la información que no pertenece al ámbito de las operaciones de la empresa, pero son importantes para el análisis. Muchos casos habrá que adquirir datos desde bases de datos públicas como por ejemplo puede ser el censo, los datos demográficos o climatológicos o desde otras bases de datos privadas a la que la organización tenga acceso.

Por otro lado, para la creación de la metodología se debe de contar con un almacén de datos, puede ser un conjunto de datos históricos, internos o externos, y descriptivos de un contexto o área de estudio, se integran y organizan de tal forma que permiten aplicar eficientemente herramientas para resumir,

describir y analizar los datos con el fin de ayudar en la toma de decisiones. Su propósito es la recuperación de información, realizar análisis e informes y suele ser utilizado por usuarios con capacidad de decisión, como pueden ser ejecutivos o analistas.

Recordemos que un almacén de datos es un repositorio de información coleccionada desde varias fuentes, almacenadas bajo un esquema unificado que normalmente reside en un único emplazamiento, que será de utilidad en el análisis de información de cierta área.

Existen varias formas de mezclar las distintas bases de datos para crear el repositorio, una posibilidad es simplemente hacer una copia de las bases de datos integrantes, obviamente, esta aproximación limita las ventajas para acceder a bases de datos heterogéneas por ello, generalmente los almacenes de datos se construyen vía un proceso de integración y almacenamiento en un nuevo esquema integrado. En la Figura 8, se muestra el proceso de integración de un almacén de datos para tres fuentes de datos originales (A, B y C).



**Figura 8.** *Proceso de integración de un almacén de datos.*

Nota: Descripción de repositorio de información coleccionada desde varias fuentes, almacenada bajo un esquema unificado que normalmente reside en un único emplazamiento. Tomado de José Hernández Orallo. Introducción a la minería de datos 2004, Pearson educación libgenli compress.pdf. pag.21

José Hernández Orallo menciona que “El objetivo de esta fase es crear un almacén de datos que contenga la información que será utilizada. En ocasiones durante la construcción de este almacén de datos también se realizan tareas propias de limpieza e integración propias de la siguiente fase. No son dos fases muy diferenciadas temporalmente, pero es conveniente visualizarlas individualmente” (José Hernández Orallo. Introducción a la minería de datos 2004, Pearson educación libgenli compress.pdf. pag.21).

## **2.1.2 Fase de selección, limpieza y transformación.**

En esta fase su principal objetivo es la mejora de la calidad del almacén de datos y trata de mostrar el conjunto de datos de la manera más apropiada para la siguiente fase que es la minería de datos.

Al momento de extraer la información antes mencionada da lugar a diversos tipos de imperfecciones, estas imperfecciones serán necesariamente resueltas mediante una limpieza minuciosa.

La información elaborada de esta fase recibe el nombre de vista minable y será utilizada por la siguiente fase, la fase de minería de datos.

### **2.1.2.1 ETL.**

Durante el desarrollo de una base de datos existen una parte fundamental la cual es la integración, implica que todos los datos de diversas fuentes que son producidos por los diversos departamentos, secciones y aplicaciones, tanto internos como externos, deben ser consolidados en una instancia antes de ser agregados al Data Warehouse (DW), y deben por lo tanto ser analizados para asegurar su calidad y limpieza, entre otras cosas. A este proceso se le conoce como integración de datos,

y cuenta con diversas técnicas y subprocesos para llevar a cabo sus tareas. Una de esas técnicas son los procesos ETL: Extracción, transformación y carga de datos.

Si bien el proceso ETL es solo una de las muchas técnicas de la integración de datos, el resto puede agruparse muy bien en sus diferentes etapas. Es decir, en el proceso de extracción tendremos un grupo de técnicas enfocadas, por ejemplo, el tomar solo los datos indicados y mantenerlos en un almacenamiento intermedio; en el proceso de transformación estarán aquellas técnicas que analizarán los datos para verificar que sean correctos y válidos, en el proceso de carga de datos se agruparán técnicas propias de la carga y de la actualización del Data Warehouse (DW).

La integración de datos resuelve diferentes tipos de problemas relacionados con las convenciones de nombres, unidades de medidas, codificaciones, fuentes múltiples, etc.

### **2.1.3 Fase de minería de datos.**

Definición: Es la extracción de información relevante de bases de datos robustas con la finalidad de detectar patrones, tendencias y relaciones para la toma de decisiones en un departamento es específico.

En esta fase se construye un modelo basándose en la vista minable que ha sido obtenida de la fase anterior, en ella se determinará el tipo de tarea, a su vez se elige el tipo de modelo y el algoritmo que más convenga. Este modelo es una representación simbólica y resumida de los datos analizados con anterioridad, que permitirá la extracción de conclusiones de manera cómoda y eficaz, determinando el tipo de modelo a aplicar, puede ser:

- a) **Predictivo:** Clasificación, regresión y series de tiempo.
- b) **Descriptivo:** Agrupamiento, reglas de asociación, análisis de correlación, etc.
- c) **Prescriptivo:** Simulaciones o estrategias para aplicar.

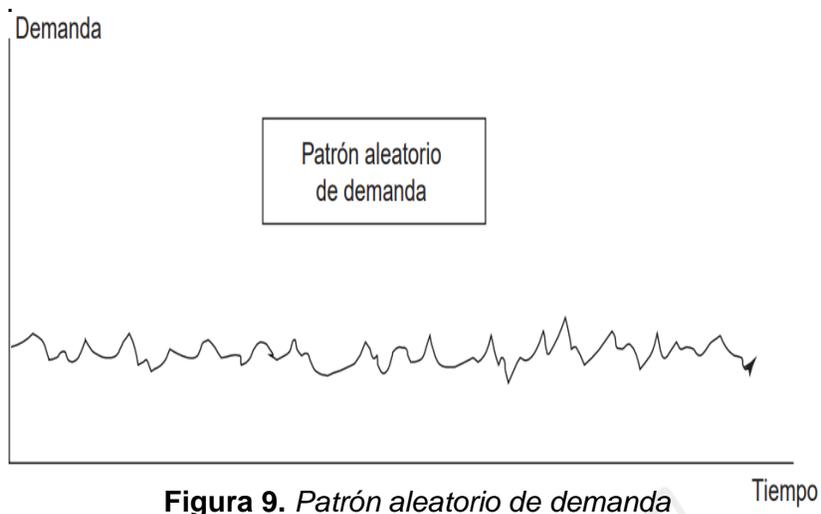
### **2.1.3.1 Series de tiempo.**

En las series de tiempo existen dos tipos fundamentales de pronósticos: cualitativos con dos subcategorías: de series de tiempo y causales, y cuantitativos con series de tiempo.

En las bases de datos los pronósticos de series de tiempo se encuentran entre los más utilizados por los paquetes de pronóstico vinculados con la proyección de demanda de productos, todos ellos parten, básicamente, de un supuesto común: que la demanda pasada sigue cierto patrón, y que si

este patrón puede ser analizado podrá utilizarse para desarrollar proyecciones para la demanda futura, suponiendo que el patrón continúa aproximadamente de la misma forma.

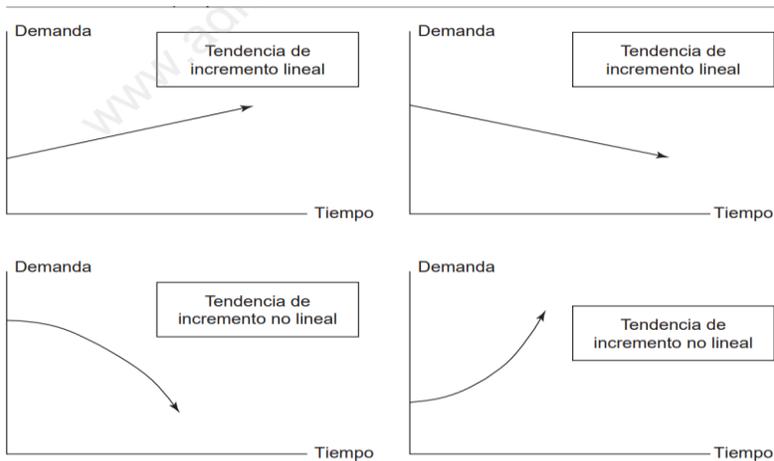
Casi todos los modelos de pronósticos de series de tiempo intentan capturar de manera matemática los patrones subyacentes de la demanda pasada, uno de ellos el patrón aleatorio, que parte del supuesto de que la demanda siempre posee un elemento aleatorio, esto significa lo que la mayoría de la gente sabe de forma intuitiva que el cliente que demanda bienes y servicios de una compañía, no lo hace de forma completamente uniforme y predecible. (Figura 9). (Stephen N. Chapman (2006). Planeación y control de la producción 2006, Pearson educación. México.pdf. pag.24)



**Figura 9.** Patrón aleatorio de demanda

Nota: Patrón aleatorio de demanda con respecto al tiempo de un determinado proceso. Tomado de Stephen N. Chapman. Planeación y control de la producción 2006, Pearson educación. México, 2006.pdf. pag.24

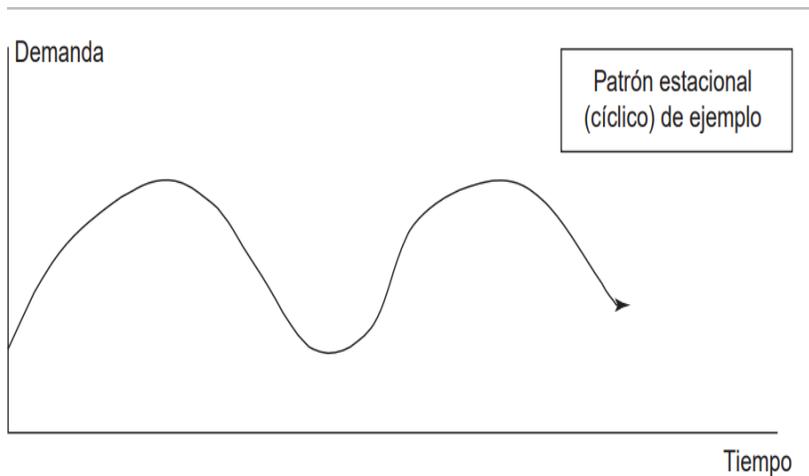
De la misma manera las tendencias pueden ser crecientes o decrecientes, y tener naturaleza lineal o no lineal, también conocido como patrones de tendencia (Figura 10).



**Figura 10.** Patrón de tendencia.

Nota: Gráficas con diferentes tipos de tendencia presentes en una serie de tiempo. Tomado de Stephen N. Chapman. Planeación y control de la producción 2006, Pearson educación. México, 2006.pdf. pag.24

Mientras que el patrón cíclico o estacional, denominado así porque sigue las estaciones del año, estos patrones en realidad son cíclicos, ya que pueden estar ligados o no a las estaciones del año. En consecuencia, los patrones cíclicos son aquellos que siguen cierto ciclo de demanda, creciente o decreciente (Figura 11).

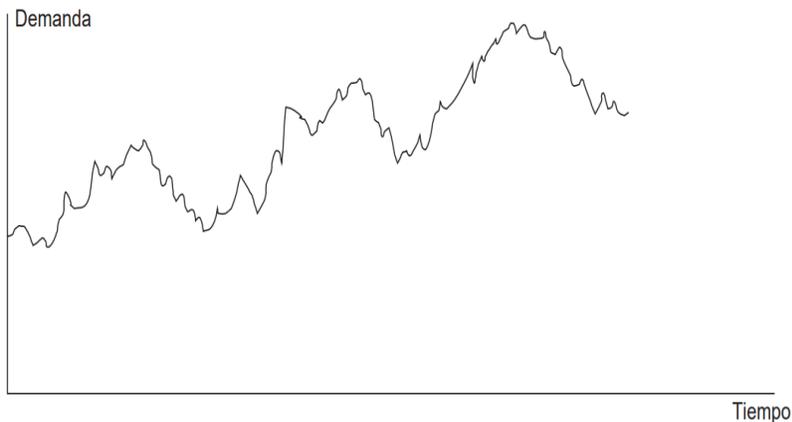


**Figura 11.** *Patrón cíclico o estacional.*

Nota: Muestra un patrón de tipo estacional de ventas. Tomado de Stephen N. Chapman. Planeación y control de la producción 2006, Pearson educación. México, 2006.pdf. pag.25

Si colocáramos un patrón aleatorio junto a un patrón de tendencia y uno estacional, obtendríamos un patrón de

demanda similar al patrón que enfrentan muchas compañías para sus productos o servicios. Por ejemplo, un patrón aleatorio estacional con una tendencia de incremento lineal (Figura 12).



**Figura 12.** *Patrón aleatorio estacional con tendencia de incremento lineal.*

Nota: Gráfica con comportamiento aleatorio con tendencia positiva. Tomado de Stephen N. Chapman. Planeación y control de la producción 2006, Pearson educación. México, 2006.pdf. pag.25.

Una vez que se han descrito los patrones básicos, se analizan algunos de los métodos de series de tiempo más simples que se han desarrollado para pronosticar la demanda, tomando en cuenta la existencia de dichos patrones. En el

primer conjunto de métodos de pronóstico se incluyen los métodos sencillos que se utilizan para intentar suavizar los patrones de demanda aleatorios suponiendo que no existen patrones de tendencia o estacionales, en este caso si no existen estos patrones en la demanda, el encargado de la proyección podría sentirse tentado a emplear la demanda real del último periodo como pronóstico para el siguiente. El problema con este método es que las organizaciones estarán continuamente incrementando o disminuyendo la producción para ajustarla al patrón aleatorio y, debido a la aleatoriedad, rara vez darían en el blanco, por este motivo los métodos de suavizado intentan, como el nombre lo indica, atenuar el patrón de demanda irregular, teniendo en cuenta que estos métodos implican importantes ventajas y desventajas que deben tomarse en cuenta; sí el método de suavizado es mínimo (es decir, si permite que permanezca casi toda la aleatoriedad), se obtendrán resultados muy poco estables, por el otro lado, si se realiza demasiado suavizado, los posibles cambios reales en la demanda no serán capturados en el pronóstico, generando inconsistencias.

### 2.1.3.2 Promedios móviles simples.

Según Stephen N. Chapman define a los promedios móviles simples “El promedio matemático de los últimos periodos recientes de la demanda real” (Stephen N. C. Control de la Producción.2006. Pág.26) La ecuación general para obtenerlos tiene la siguiente forma:

$$F_t = A_{t-n} + A_{t-n+1} + L + A_{t-1} n.$$

Dónde: **F** es el pronóstico.

**T** es el periodo de tiempo actual.

**F<sub>t</sub>** es el pronóstico para el periodo de tiempo actual.

**A<sub>t</sub>** es la demanda real en el periodo t, y n es el número de periodos que se utiliza.

### 2.1.3.3 Coeficiente de Correlación.

El coeficiente de correlación es una medida descriptiva de la intensidad de la relación entre dos variables x y y. Los valores del coeficiente de correlación siempre estarán entre -1 y +1. Un valor de +1 indica que las dos variables x y y, está perfectamente relacionadas en un sentido lineal positivo, es decir, todos los puntos de los datos se encuentran en una línea

recta que tiene pendiente positiva. Un valor  $-1$  indica  $x$  y  $y$  son perfectamente relacionados en un sentido lineal negativo con todos los puntos de los datos de la recta con pendiente negativa. Los valores del coeficiente de correlación cercanos a  $0$  indican que  $x$  y  $y$  no están relacionados linealmente.

#### **2.1.4 Fase de Evaluación e interpretación.**

La finalidad de esta fase es medir la calidad de los patrones descubiertos por un algoritmo de minería de datos, esta medida puede atañer a varios criterios, algunos de ellos bastante subjetivos, idealmente se pretende que los patrones descubiertos deben tener tres cualidades: ser precisos, comprensibles (es decir, inteligibles) e interesantes (útiles y novedosos) para que cumplan con el principal objetivo.

#### **2.1.5 Fase de difusión, uso y monitorización.**

Esta fase tiene como la finalidad de una aplicación correcta del conocimiento adquirido, en forma de modelo y en el dominio para el cual fue creado. Esto es usar el modelo de forma correcta por parte de los usuarios para el KDD.

## 2.2 Dashboard.

El dashboard es una herramienta de análisis de datos, esta nos sirve para representar la información gráfica de las métricas o KPI más importantes de manera tal que puedas hacer seguimiento a un proyecto de manera oportuna, eficaz y sencilla.

Al dashboard en español se le conoce también como “**cuadro de control**” o “**tablero de control**”, este se puede encontrar de diferentes maneras, según las métricas e información que se debe analizar; la información que se trabaja dependerá de las necesidades del proyecto, los objetivos, recursos y capacidad de trabajo y con ello tomar una buena decisión.

### 2.2.1 Pasos para la creación de dashboard:

- **Establece objetivos cuantificables:** En primer lugar, se debe definir cuáles son los objetivos, deben ser cuantitativos, es decir, contables, un dashboard rara vez puede representar información cualitativa.
- **Define tus KPI:** Definir y seleccionar los indicadores del rendimiento del trabajo (KPI) sobre los objetivos que

tienes marcados y es la forma de conocer si los estás cumpliendo. Los KPI son métricas, pero no todas las métricas pueden ser consideradas KPI, en muchas ocasiones, el principal problema es identificar y definir las KPI correctos, en función a las necesidades de información.

- **Separa y organiza los datos:** El tercer paso para hacer un buen dashboard es separar y organizar los datos que arrojan cada una de tus acciones.
- **Define un estilo gráfico para tu dashboard:** Cómo vas a visualizar tu información, ya sea para uso personal o para presentar ante un equipo de trabajo.

### **2.2.2 Ventajas y desventajas.**

A) **Ventajas:** Amplio acceso a la información, mejora en la toma de decisiones, establece indicadores con los resultados arrojados y automatización de la información.

B) **Desventajas:** El panel de datos no refleja totalmente la estrategia de la empresa.

## 2.2.3 Herramientas de Desarrollo.

Para la elaboración del dashboard se requieren diversos programas, que son fundamentales para la implementación de esta herramienta, a continuación, se muestra el proceso de instalación.

### a) Instalación del programa Día.

Descargar el programa desde el siguiente link:

<http://dia-installer.de/shapes/ER/index.html.es>

Al concluir la descarga, abrir el archivo y saldrá el primer cuadro de diálogo donde empezará la instalación (Figura 13) selecciona siguiente y se genera otro cuadro donde pide el idioma a utilizar; en este caso elegir “español” (Figura 14).



Figura 13. Instalar el programa.

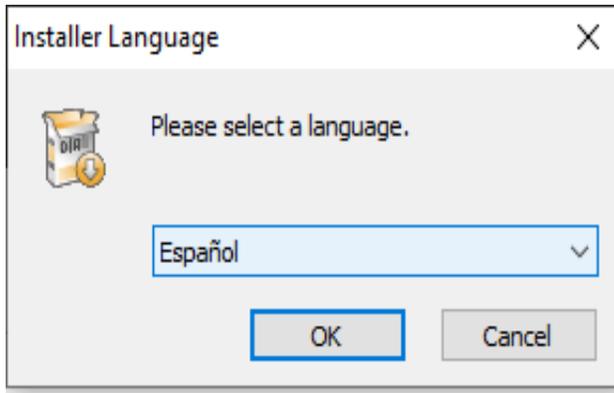


Figura 14. Seleccionar el idioma.

Posteriormente seleccionar los componentes a instalar (Figura 15), marcar la casilla traducciones, en el siguiente cuadro de diálogo (Figura 16) elegir la casilla instalar, elegir el directorio donde se va a guardar el programa que es c:\Program Files (x86) Día.

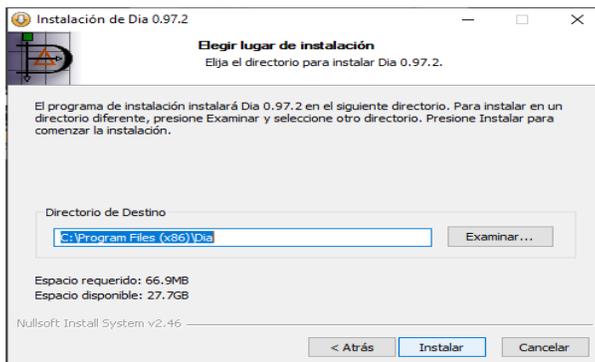


Figura 15. Selección de componentes.

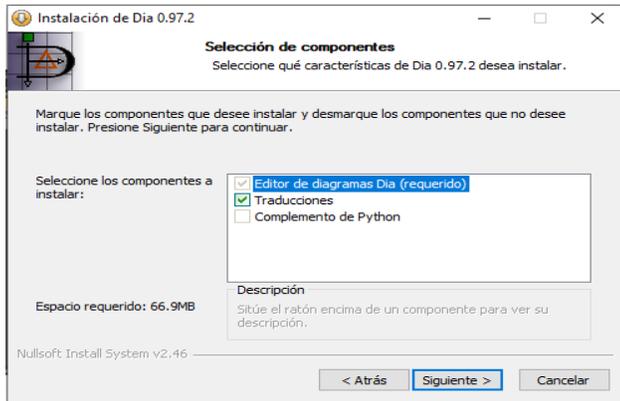


Figura 16. Elegir ubicación de instalación.

Para concluir la instalación dar click en la casilla terminar (Figura 17).

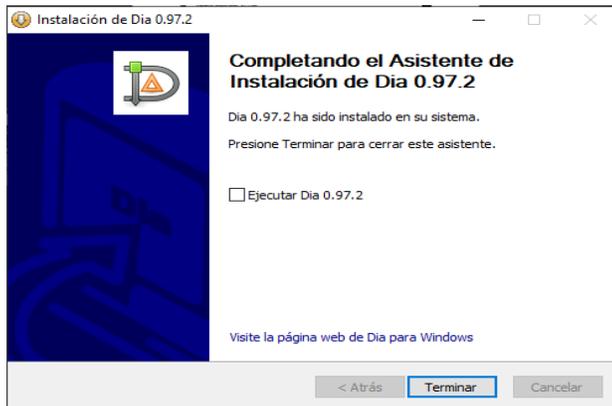


Figura 17. Completando el asistente de instalación.

## b) Instalación del programa XAMPP.

Descargar el programa desde el link:

<https://www.apachefriends.org/index.html>

Se selecciona la casilla next (Figura 18), en el siguiente cuadro de diálogo seleccionar nuevamente next (Figura 19). En este elegir los componentes a instalar, en este caso son todos, ahora en el cuadro de diálogo installation folder, seleccionar la carpeta donde se guardará el programa instalado c:\xampp (Figura 20).

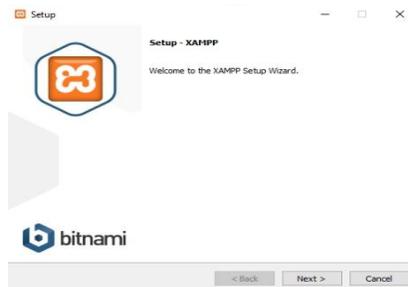


Figura 18. Instalación.

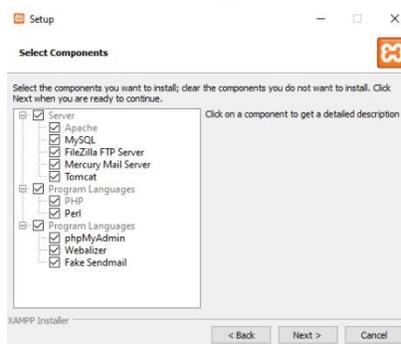


Figura 19. Seleccionar componentes.

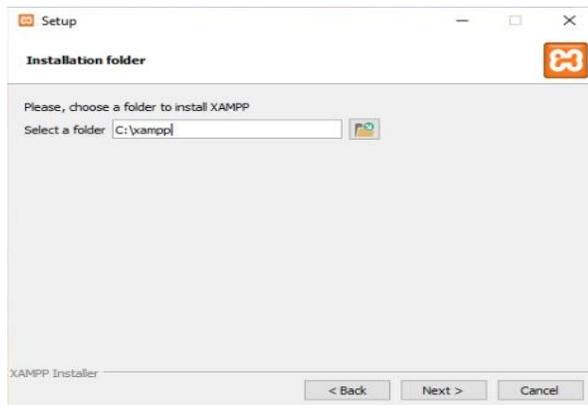


Figura 20. Seleccionar folder.

En la próxima ventana de diálogo (Figura 21) seleccionar el idioma a utilizar “inglés”, elegir siguiente, en el próximo cuadro se mostrará el proceso de instalación (figura 22), y para terminar en la próxima ventana (Figura 23) seleccionar la casilla finish.

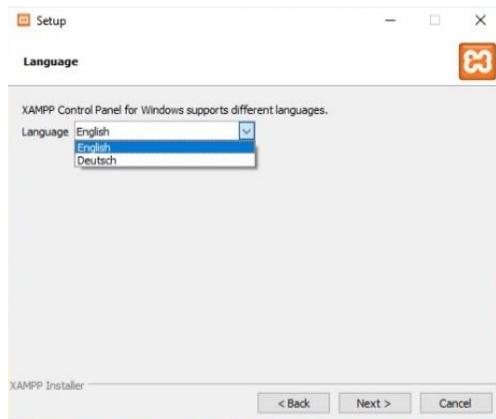


Figura 21. Elegir idioma.



Figura 22. Instalación.



Figura 23. Finalizar instalación.

#### d) Instalación del programa MySQL.

Descargar el programa desde el siguiente link:

<https://www.mysql.com/products/workbench/>

En la primera ventana seleccionar la casilla next (Figura 24) y posteriormente saldrá el cuadro de diálogo denominado Setup Type. En este caso elegir el tipo de instalación que será “complete” (Figura 25) le damos next, en la venta Ready to install the Program seleccionar la casilla install (Figura 26).

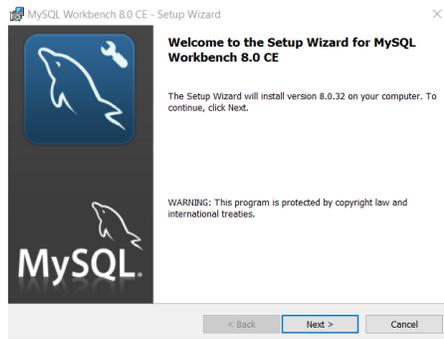


Figura 24. Iniciar instalación.

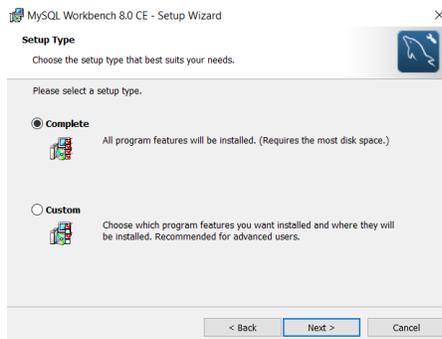


Figura 25. Elegir tipo de instalación.

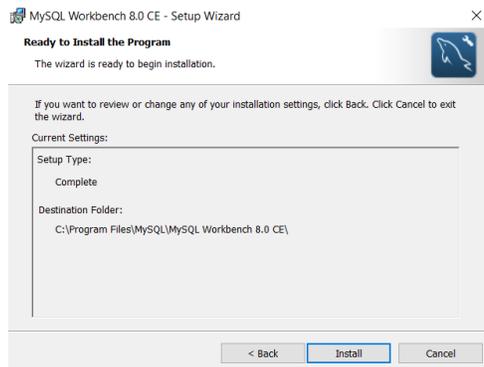


Figura 26. Instalación.

En la ventana destination folder (Figura 27) seleccionar la carpeta donde se va a instalar el programa y le damos click a la casilla next, en la siguiente ventana (figura 28) se mostrara el proceso de instalación, al final aparecera el último cuadro de diálogo en el cual elegir la casilla finish para terminar el proceso (Figura 29).

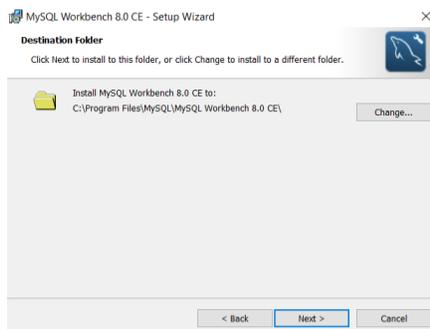


Figura 27. Seleccionar folder.

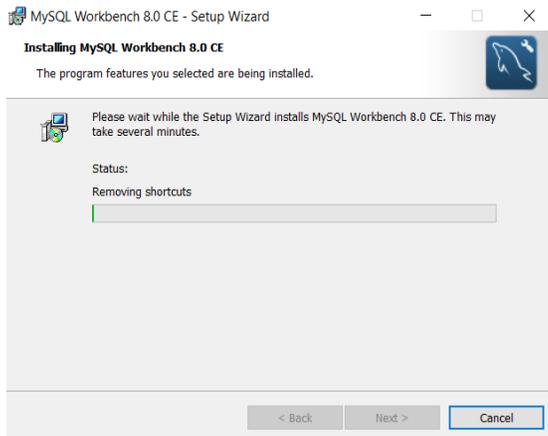


Figura 28. Instalando.

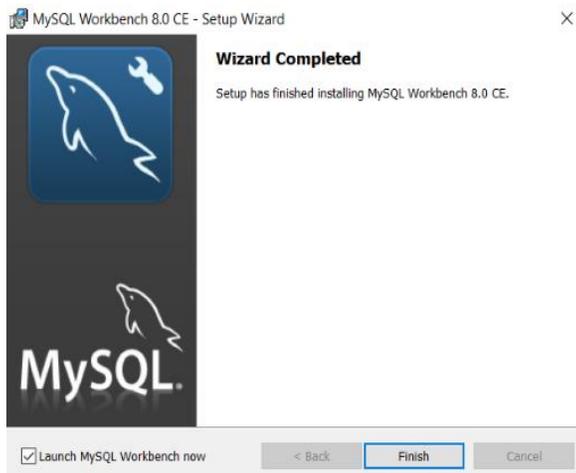


Figura 29. Instalación completa.

## e) Descarga del conector odbc.

Descargar el programa desde el link:

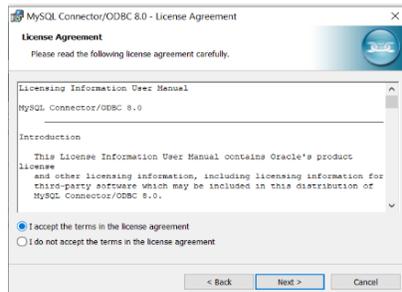
<https://dev.mysql.com/downloads/file/?id=515808>

En el primero cuadro de diálogo seleccionar la casilla next (Figura 30), en la siguiente venta elegir por default el botón y aceptar la licencia (Figura 31), posteriormente saldrá otra ventana donde elegir el tipo de instalación en este caso elegimos “typical” y damos click en la casilla next (Figura 32).



Figura 30. Iniciar instalación.

Figura 31. Aceptar la licencia.



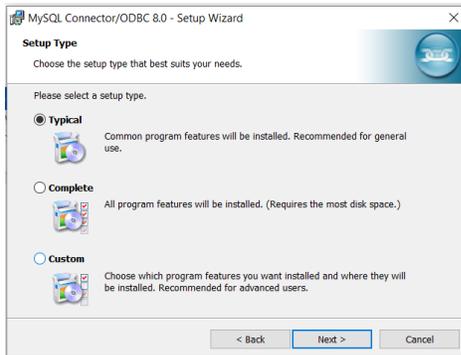


Figura 32. Elegir instalación.

En la ventana “Ready to install the program” seleccionar la casilla install, y para finalizar en el cuadro de diálogo (figura 33) seleccionar “finish”, con esto concluye la instalación del conector.

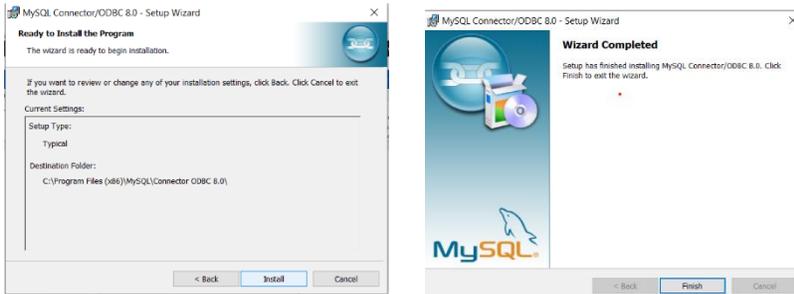


Figura 33. Instalar el conector y finalizar.

**f) Descarga el programa Power BI.**

Descargar el programa desde el siguiente link:

<https://www.microsoft.com/es-es/download/details.aspx?id=58494>

En el cuadro de diálogo (Figura 34) se selecciona la casilla “ejecutar”, en la siguiente ventana elegir el idioma “español” (Figura 35), en el próximo cuadro de diálogo elegir los “términos y condiciones del contrato de licencia” (Figura 36), seleccionando la casilla siguiente.

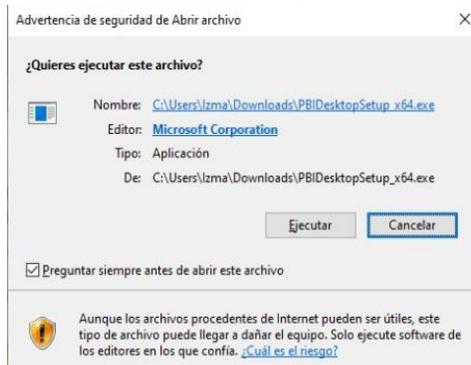


Figura 34. Ejecutar archivo.

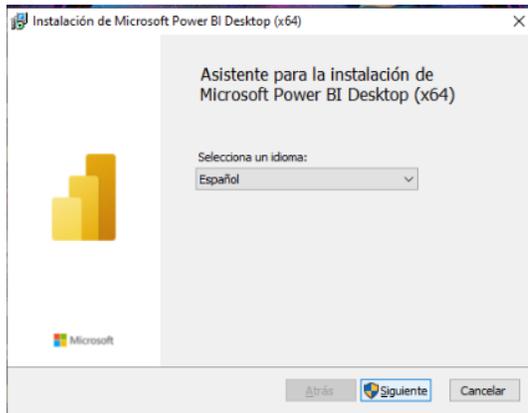


Figura 35. Seleccionar idioma.

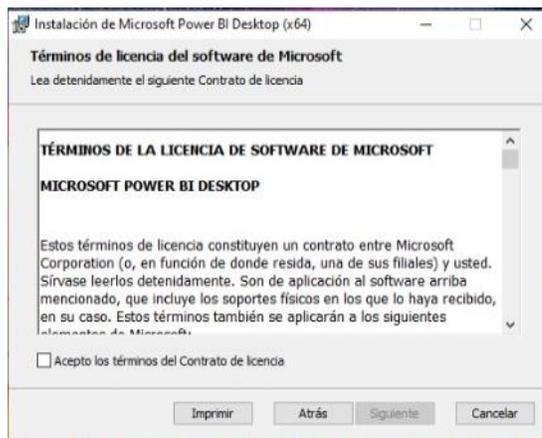


Figura 36. Términos y condiciones.

Posteriormente seleccionar la carpeta donde se guardará el programa y enseguida la casilla siguiente, al final saldrá otra ventana en la cual escogemos la casilla finalizar (Figura 37).

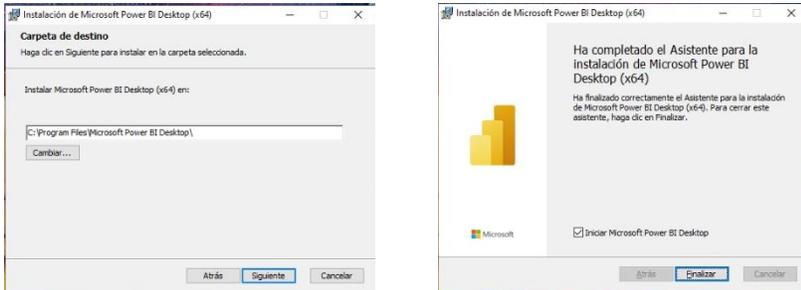


Figura 37. Seleccionar carpeta para finalizar instalación.

### **g) Descarga el programa Javajdk.**

En el siguiente link se descarga el programa Javajdk:

<https://www.techspot.com/downloads/5553-java-jdk.html>

En la ventana de diálogo (Figura 38) dar click en la casilla next, posteriormente seleccionare el destino del folder donde será instalado el programa (Figura 39) elegir next, en el cuadro de diálogo "Progress", esperar a que concluya la instalación. (Figura 40).



Figura 38. Configuración.

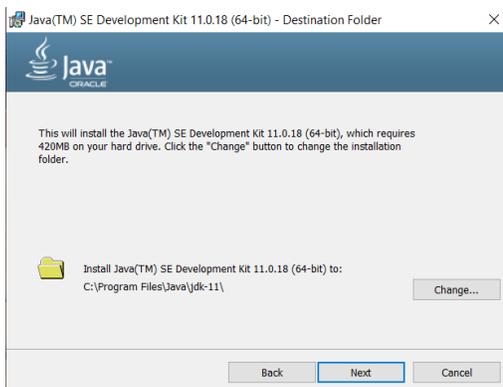


Figura 39. Seleccionar folder.



Figura 40. Instalando.

En la venta de diálogo (Figura 41) dar click en la casilla close, después validar en el sistema mediante la consola cmd si el programa se instaló correctamente, en este caso quedó la instalación en optimas condiciones. (Figura 42).



Figura 41. Instalación completa.

```
Símbolo del sistema
Microsoft Windows [Versión 10.0.19041.1004]
(c) Microsoft Corporation. Todos los derechos reservados.

C:\Users\Ismael>java --version
java 11.0.19 2021-01-17 LTS
Java(TM) SE Runtime Environment 18.9 (build 11.0.19-17-LTS-195)
Java HotSpot(TM) 64-Bit Server VM 18.9 (build 11.0.19-17-LTS-195, mixed mode)

C:\Users\Ismael>
```

Figura 42. Validar la instalación.

### h) Descarga el programa: Talend Open Studio.

Descargar el programa desde el siguiente link:

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

En el primer cuadro de diálogo (Figura 43) seleccionar la opción next, en la siguiente ventana (Figura 44) elegir la casilla next.

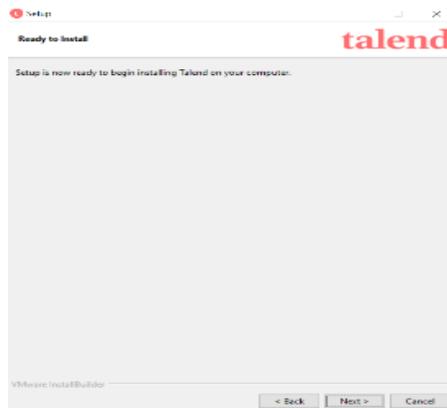


Figura 43. Comenzar la instalación.

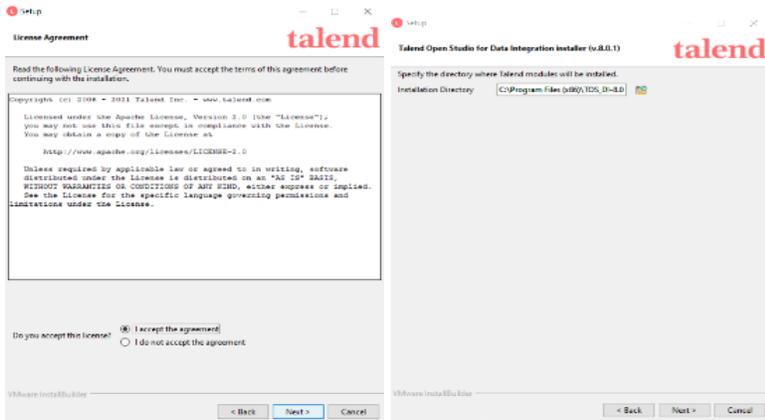


Figura 44. Aceptar la licencia y seleccionar el folder.

En la venta de diálogo “Setup” dar click en la casilla next, nuevamente seleccionar next en el cuadro de diálogo (Figura 45) y en la última ventana seleccionar la casilla finish (Figura 46) con esto se finaliza el proceso de instalación del programa Talend Open Studio.



Figura 45. Setup.

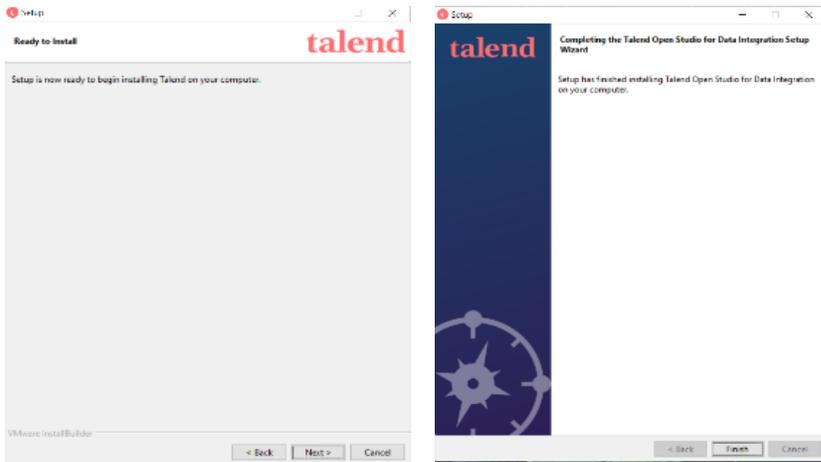


Figura 46 Instalando el programa y finalizar.

## **CAPÍTULO 3. CASO PRÁCTICO DISEÑO DE LA SOLUCIÓN.**

### **3. Contexto del negocio.**

Actualmente la empresa promotora y marketing trabaja de manera tradicional en el manejo de bases de datos donde aún no se integran nuevas e innovadoras tecnologías, provocando el rezago en el manejo de data. Esto ha generado problemas, en la consulta, análisis y gestión de la información con la inexistencia de una base de datos. Al elaborar reportes ineficientes e incongruentes, generando conflictos con los clientes, por la inconsistencia de la información, lo que provoca una salida errónea y una baja calidad de los resultados, evitando llegar a una buena toma de decisiones al momento requerido y también a causa de nula familiarización de nuevas herramientas.

En este caso teórico-práctico se implementaron técnicas y herramientas de BI (business intelligence), con la finalidad de mejorar los procesos de elaboración de reportes, esto nos ayudará a tener un mejor control sobre la información que se genera día a día en la empresa, con el objetivo de una mayor eficiencia en la recopilación de datos, mejor coherencia y mayores facilidades para compartir resultados.

También se aplicarán técnicas de minería de datos, con el fin de obtener una proyección relevante para la organización y se pueda tener una mejor toma de decisiones.

### 3.1 Propuesta de solución.

La propuesta para esta problemática básicamente fue estructurar una base de datos en la cual se concentró toda la información proporcionada por el personal de supervisión en formato Excel, posteriormente la carga de datos a la base se realizó con el programa Talend Open Studio, al subir la data se elabora una conexión al visualizador de datos Power Bi con el cual se realiza el dashboard y la minería de datos con la finalidad de apoyar en la toma de decisiones para el área de ventas del año 2023 mostrado en la Figura 47.

### 3.2 Arquitectura de solución.

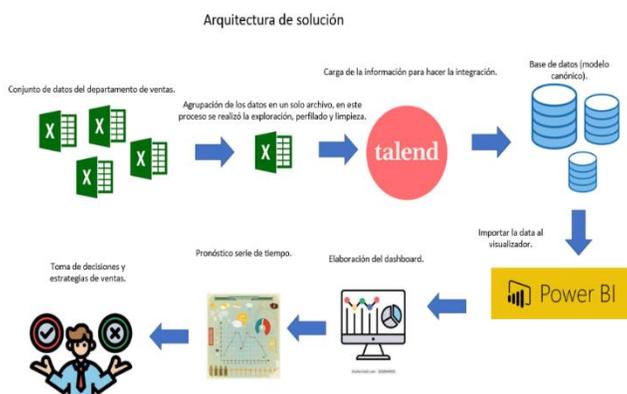


Figura 47. Estructura de solución

Nota: Procesos dentro de la arquitectura de solución, partiendo de la obtención de los datos hasta el dashboard. Creación propia.

### **3.3 ETL.**

El proceso ETL (extraer, transformar, cargar), se utiliza para recopilar datos de varios orígenes, transforma los datos según las reglas del negocio y los carga en un almacén de datos.

Se recopilan datos del departamento de ventas del año 2023, los cuales se utilizan en formato Excel:

- Producto
- Descripción
- Región
- Vende
- Fecha
- Tienda
- Distribuidor

Al realizar la exploración de datos encontramos inconsistencias, es decir, datos faltantes en el archivo producto, posiblemente a causa de una captura errónea de datos por el personal, mientras que en los demás no existen datos duplicados, sin faltantes y estaban formateados (Figura 48).

1	idproduct	descripcion	precio
2	m1	Tira Gold P220 (Subempaque c/50)	20.29
3	m10	3M? Grippy Sanding Cloth, 35110, 5.5 inch x 4.5 inch, P400	515.98
4	m11	Compuesto Pulidor 946 ML	859.13
5	m12	Pistola Grip Sander, 55 MM (2 in) y 75 MM (3 in)	3521.85
6	m13	Cinta Para Moldura	295
7	m14	Rollo Hookit 3M Blue G80 .1580M	761.72
8	m15	Disco Hookit? Clean Sanding Cubitron II 6" P400 (Subempaque c/50)	27.14
9	m16	Imperial Wetordry 9 x 11" P600 (Subempaque c/50)	15.46
10	m17	Disco Hookit? Gold Film 6", P1000 (Subempaque c/100)	19.92
11	m18	Disco Hookit? Gold Film 6", P1200 (Subempaque c/100)	19.92
12	m19	Disco Hookit? Clean Sanding 6" P240 (Subempaque c/50 discos)	27.17
13	m2	Cera liquida premium 946 ML	306.55
14	m20	Rollo Hookit? Cubitron II 40+	990.14
15	m21	Paquete de Fibra Gris	773
16	m22	Masking tape	370
17	m23	Disco Hookit Blue G320 6"	13
18	m24	Respaldo Hookit? manual para Disco 6" Media Luna	259
19	m25	Adhesivo AUTOMIX? para fibra de vidrio 400ml	1346
20	m26	Garlopa Amarilla Hookit? 2.75" x 5"	500.64
21	m27	Rollo Hookit 3M Blue G180 .1580M	649.32
22	m28	Borla de Esponja para Pulir Blanca 8" (para uso con el Respaldo 5718)	1013
23	m29		34.72
24	m3	Cinta Masking Tape 203 3/4	28
25	m30		36.09
26	m31		27.14
27	m32	Tela Pegajoza azul	28.8
28	m33	Abrillantador Econpack	34.72
29	m34	Abrillantador Manual Econpack	52.02
30	m35	Cera Liguada Econpack	45.09
31	m36	Cinta Doble Cra	31.06
32	m37	Compuesto Pulidor Econpack	27.14

**Figura 48.** *Recopilación de datos del departamento de ventas del año 2023.*

Nota: Recopilación de datos en el área de ventas 2023, datos proporcionados por la empresa para analizar el comportamiento de las mismas. Creación propia.

Esto se resolvió mediante la búsqueda por ID en el catálogo de productos proporcionados por la empresa. Una vez resuelto, se procedió a la integración mediante el uso del programa Talend Open Studio para el proceso ETL.

Lo primero a realizar, es crear una hoja donde se va a diseñar el ETL (Figura 49).

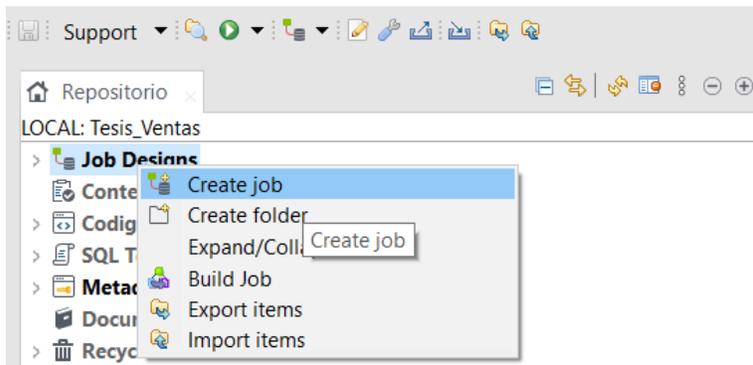


Figura 49. Creación de hoja donde se va a diseñar el ETL.

Al obtenerlo se genera un cuadro de diálogo (Figura 50), en este caso se agrega el nombre, el objetivo y se finaliza el trabajo, como se muestra en el support (Figura 51).

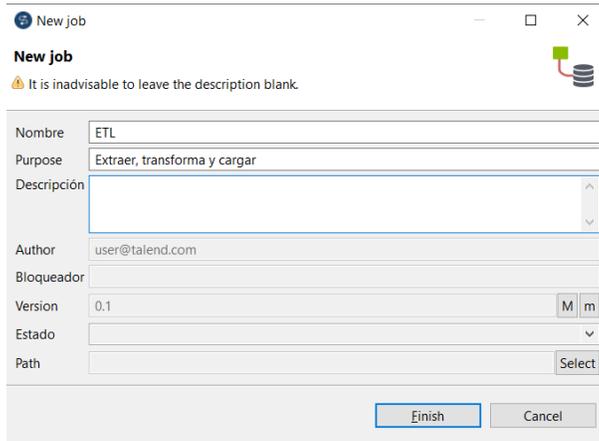


Figura 50. Nuevo trabajo.

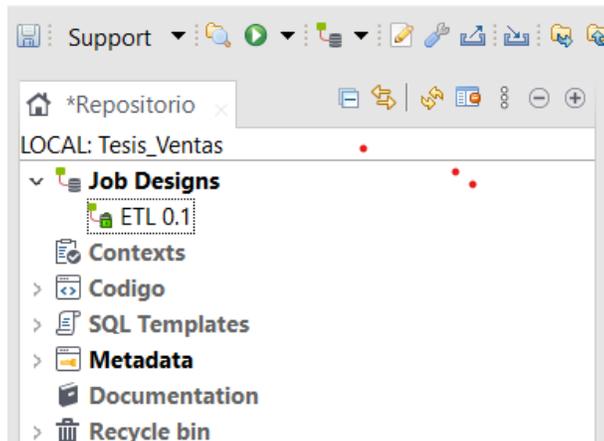


Figura 51. Configuración.

Posteriormente se carga la base de datos (Figura 52) donde se agregará la información, generando una conexión en DB connections (Figura 53).

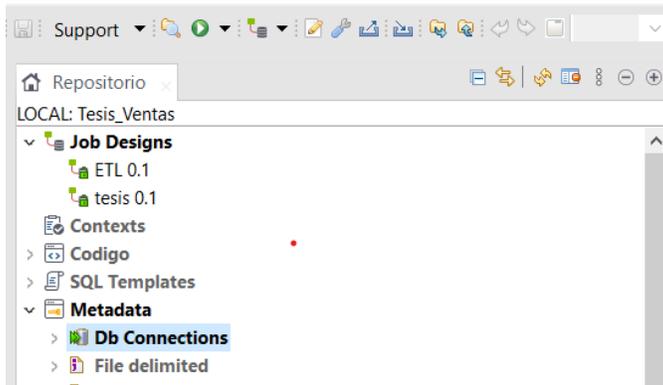


Figura 52. Seleccionar repositorio.

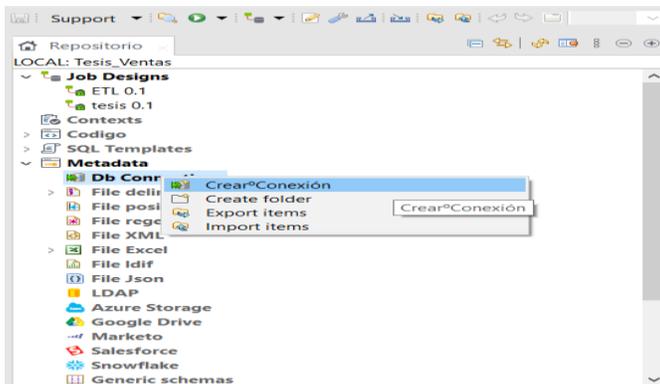


Figura 53. Crear conexión a la base de datos.

En la conexión (Figura 54) se coloca nombre y el objetivo de la base de datos, posteriormente se agrega información relevante de la base de datos como la versión, en este caso es MaríaDB, el login: root, el servidor: localhost, el puerto: 3306 y por último el nombre de la base: “tesis”, y se finaliza el proceso (Figura 55).

Nombre	TESIS
Purpose	FTI
Author	user@talend.com
Bloqueador	
Version	0.1
Estado	
Path	Select

Figura 54. Conexión a la base de datos.

DB Type	MySQL
Db Version	MariaDB
String of Connection	jdbc:mariadb://localhost:3306/tesis?noDatetimeStringSync=true&enabledTLSProto...
Login	root
Contraseña	
Servidor	localhost
Puerto	3306
Database	tesis
Additional parameters	noDatetimeStringSync=true&enabledTLSProtocols=TLSv1.2,TLSv1.1,TLSv1

Figura 55. Selección de la base de datos.

A continuación, se extraen los esquemas o las tablas (Figura 56) que se utilizarán para la conexión, posteriormente elegir extraer esquema en el cuadro de diálogo (Figura 57) y para finalizar seleccionar la casilla next.

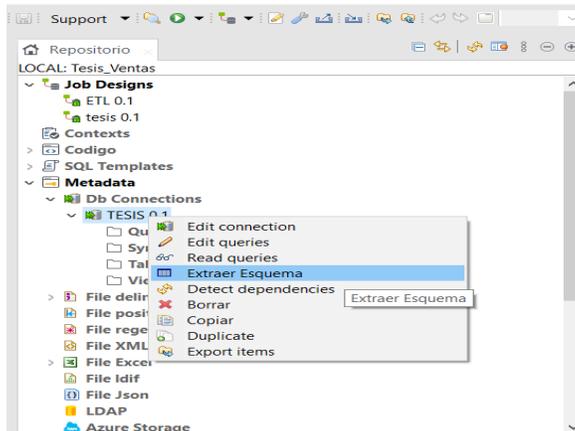


Figura 56. Extraer esquema.

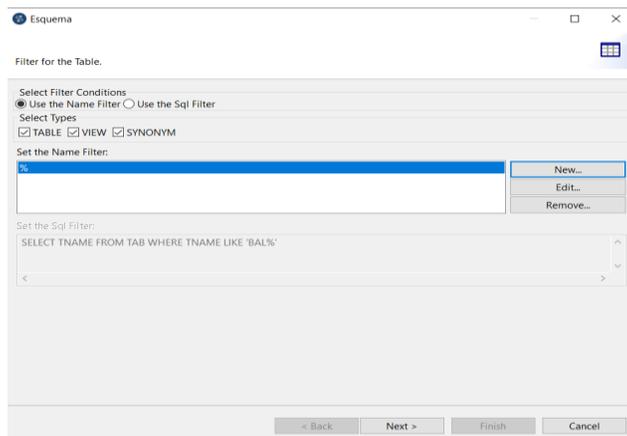


Figura 57. Esquema.

En el cuadro de diálogo seleccionar todas las casillas de la tesis, las cuales son las tablas (dimensiones) de la base de datos, y dar siguiente (Figura 58).

En la conexión del nuevo esquema seleccionar finalizar, cabe mencionar que en este recuadro se pueden cambiar las características del BD, pero en este caso, ya están definidas y no se tiene la necesidad de realizar dichos cambios (Figura 59).

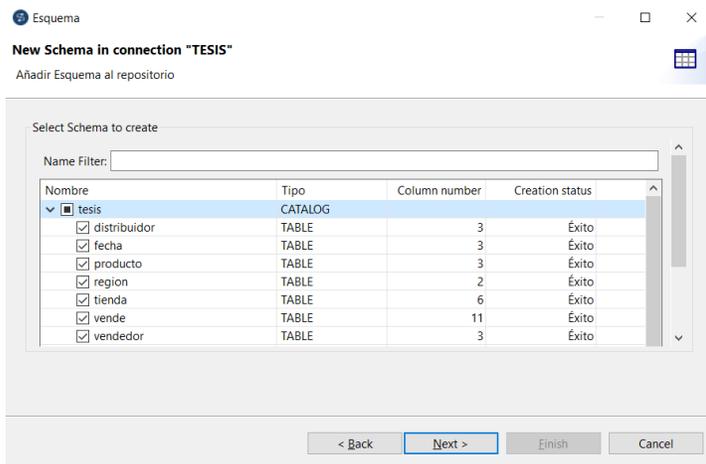


Figura 58. Seleccionar tablas.

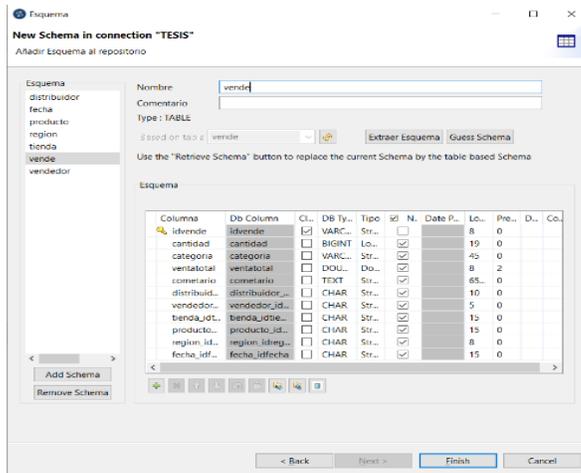


Figura 59. Nuevo esquema.

Como se muestra en los esquemas, se encuentran cargados en nuestra DB connections (Figura 60).

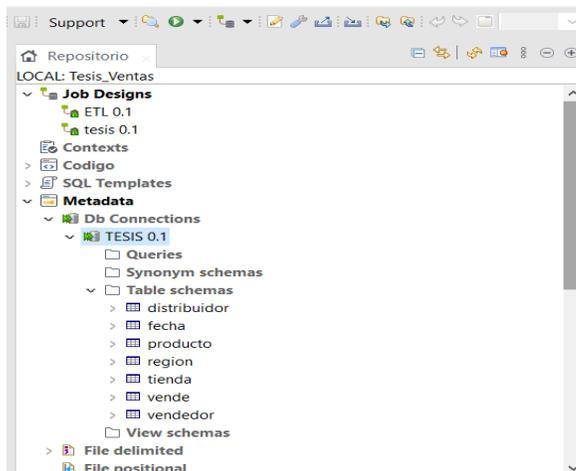


Figura 60. Esquema tesis 0.1.

En el proceso de importación de información fuente de datos, se inicia por la dimensión Distribuidor. Seleccionar créate file delimited (Figura 61) ya que el archivo tiene un formato CSV, como se observa en la figura 62 step 1 of 4 agregar el nombre del archivo que se va a importar y elegir siguiente.

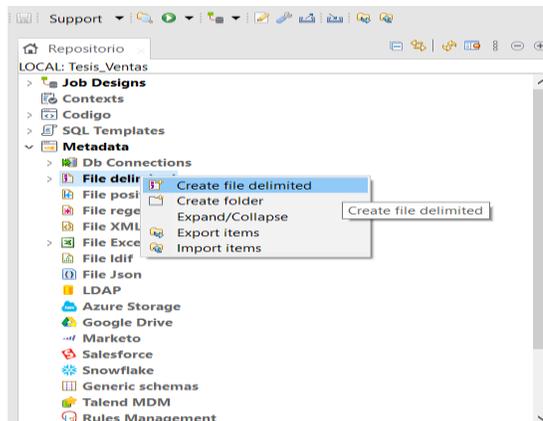


Figura 61. Crear folder.

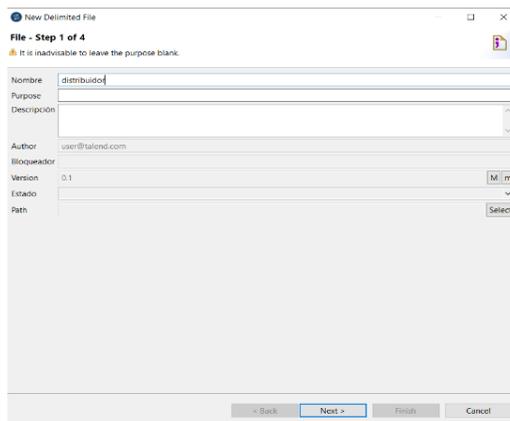


Figura 62. Nuevo archivo delimitado.

En el cuadro de diálogo step 2 of 4 muestra la ubicación donde se guarda el fichero, seleccionar la casilla next (Figura 63), luego se elige el tipo de codificación ventana de diálogo step 3 of 4 que tiene los datos en este caso es UTF-8 y le dar next (Figura 64).

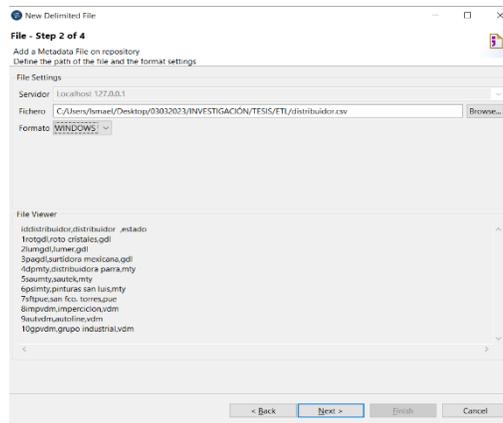


Figura 63. Seleccionar carpeta.

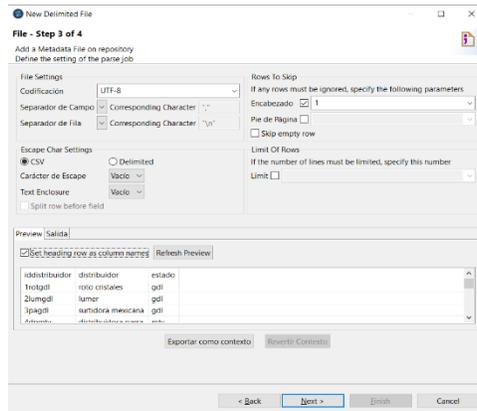


Figura 64. Seleccionar componentes.

Para finalizar se validan los datos para cada atributo en el cuadro de diálogo step 4 of 4, se selecciona y se concluye el proceso de importación (Figura 65).

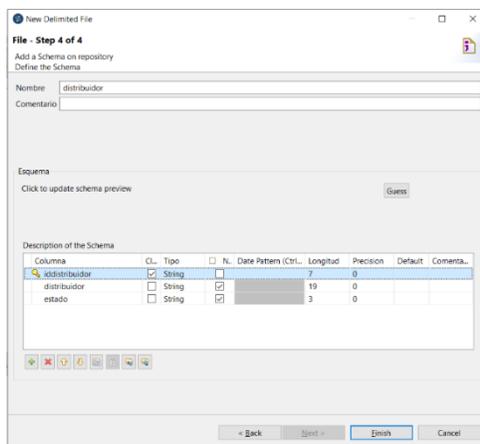


Figura 65. Seleccionar componentes de la tabla distribuidor.

Se repite los pasos para cada dimensión de la base de los datos que se denominan "esquemas", hasta tener todo cargado como se muestra en la Figura 66.

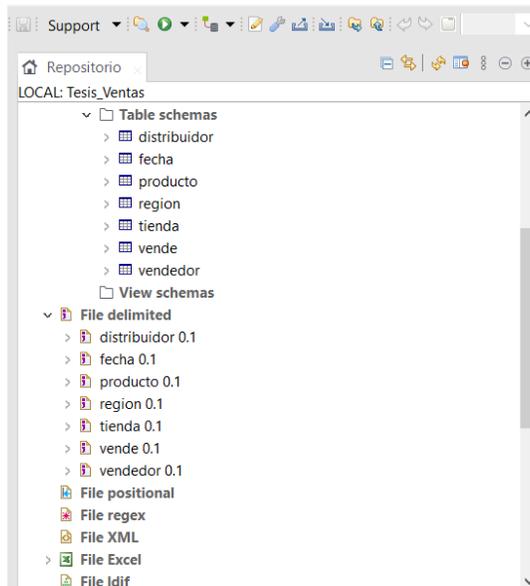


Figura 66. Esquemas de la base de datos.

Por otro lado, se comienza a realizar el proceso de carga de la información, incorporando el primer componente que será la relación a la hoja de trabajo, seleccionar la entrada tFileinputDelimited ya que los archivos que se cargaron están en formato CSV (Archivo delimitado por comas) (Figura 67).

Se agrega otro componente, pero ahora de la base de datos que se generó en MySQL y en este caso seleccionar la salida tDBinput(MySQL), (Figura 68).

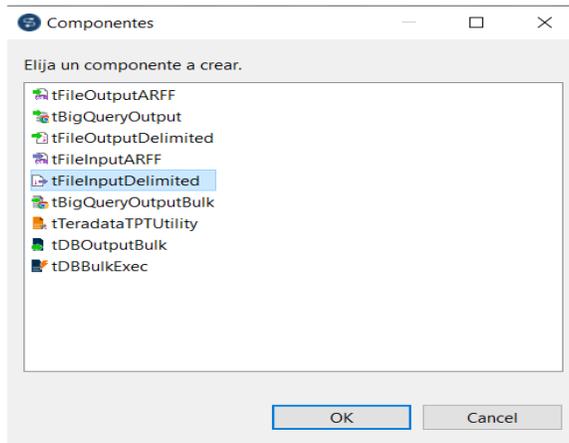


Figura 67. Componente por crear.

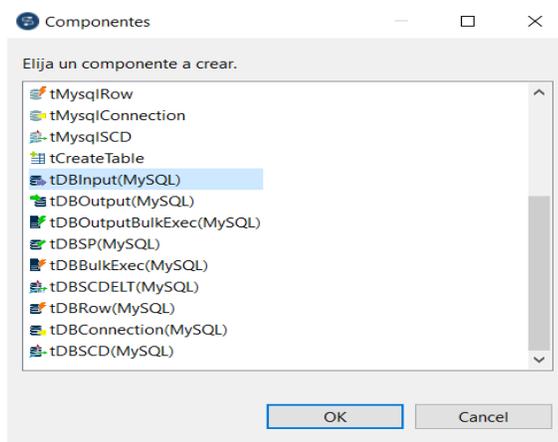


Figura 68. elegir componente.

El palette de la herramienta ETL (Extract Transform Load) a utilizar para cargar la información, eligiendo tMap sirve

para la unión de los datos(mapeo) y tLogRow para verificar que no haya error en la carga (Figura 69).

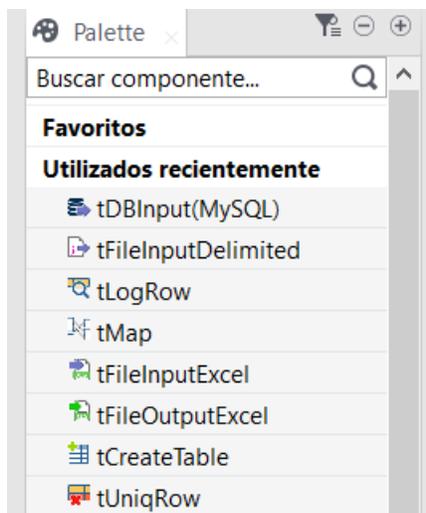


Figura 69. Palette.

Se muestra la hoja de trabajo con los componentes que se cargaron, para posteriormente unirlos (Figura 70).

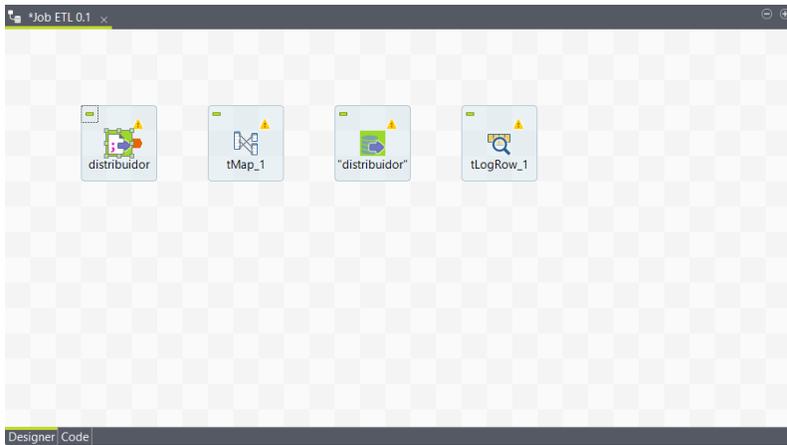


Figura 70. Hoja de trabajo.

Para visualizar las conexiones dar click derecho, escoger fila (Figura 71), posteriormente Main como se muestra en la Figura 72, esto permitirá ver las conexiones realizadas.

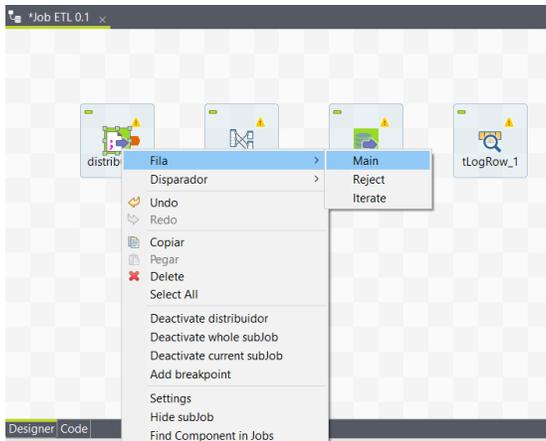


Figura 71. Conectar componentes.

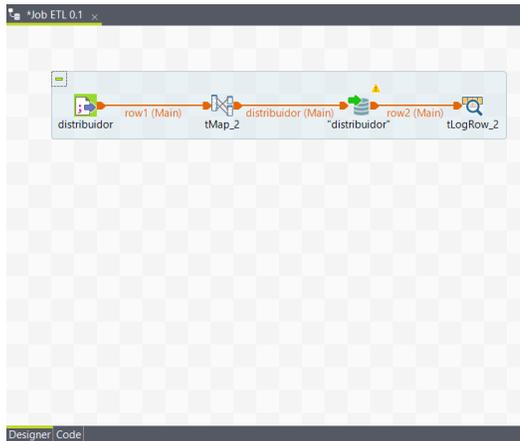


Figura 72. Componentes mapeados.

Se procede a validar la información con la herramienta tMap, esta nos ayuda a cargar los datos de los archivos a la base de datos con sus respectivas llaves, formatos y tamaño de datos.

Con el editor de esquemas se da formato y valida que los datos sean iguales, en este caso son iguales, ahora seleccionar apply, ok para finalizar. (Figura 73).

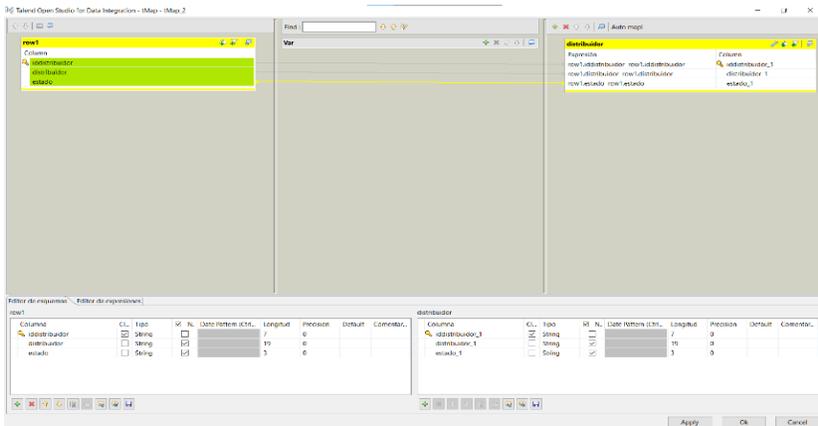


Figura 73. Editor de esquemas.

Se repite el proceso con todas las dimensiones de la base de datos que se van a cargar, lo que dará como resultado la siguiente hoja de trabajo (Figura 74).



Figura 74. Mapeo de todos los esquemas.

Se muestra el tablero donde se va a ejecutar el trabajo, se procede a correr la integración a la base de datos, seleccionar la casilla Run (Figura 75).

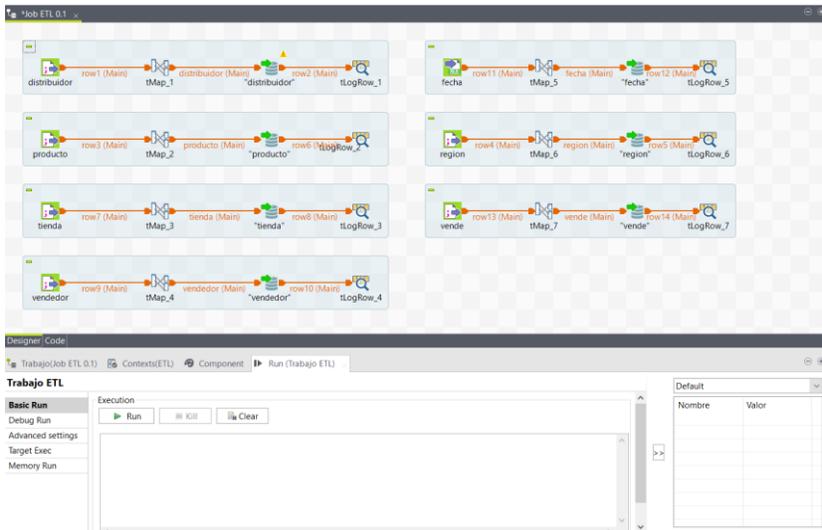


Figura 75. Proceso de carga de los datos.

En la Figura 76, muestra que el trabajo ha corrido sin ningún problema, así como el tablero del proceso ETL (Extract Transform Load), validar que se haya cargado en la base de datos de MaríaDB.

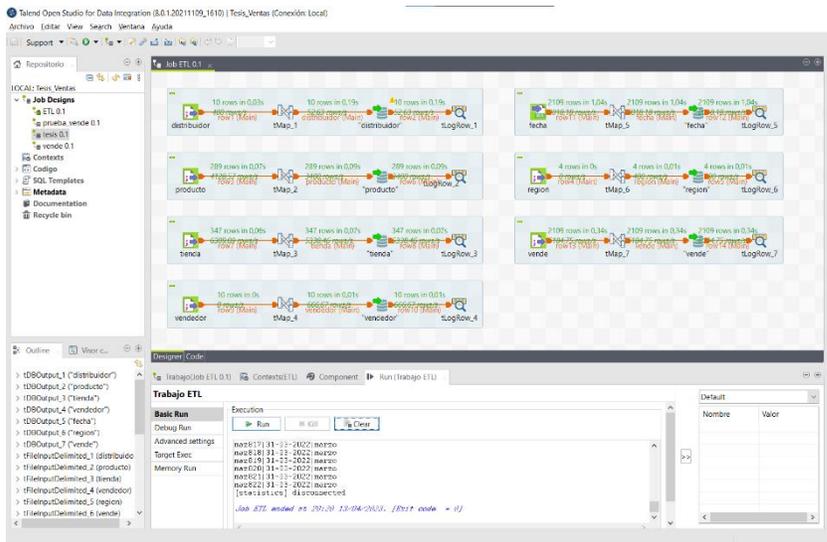


Figura 76. Trabajo ETL finalizado.

En la validación del Proceso ETL, se revisa en la base de datos de MaríaDB que se observen las filas de cada relación correspondiente con el número de filas que cargó en la hoja de trabajo, confirmando que el proceso fue correcto (Figura 77).

```

MaríaDB [tesis]: SELECT *FROM distribuidor;
-----
| idistribuidor | distribuidor | estado |
-----
| 189vdm       | grupo industrial | vdm |
| 189gdl       | roya cristales | gdl |
| 21ungdl      | luner          | gdl |
| 189gdl       | surtidora mexicana | gdl |
| 46wtr        | distribuidora parrá | wty |
| 189aunty     | sautok         | wty |
| 189alnty     | pinturas san luis | wty |
| 78fpe        | san fco. torres | pue |
| 81aupdm      | impericlon     | vdm |
| 81autdm      | autoline       | vdm |
-----
10 rows in set (0.000 sec)

MaríaDB [tesis]: SELECT *FROM region;
-----
| idregion | estado |
-----
| gdl      | gdl |
| wty2     | wty |
| pue3     | pue |
| vdm4     | vdm |
-----
4 rows in set (0.011 sec)

MaríaDB [tesis]: SELECT *FROM vendedor;
-----
| idvendedor | nombre | telefono |
-----
| gdl33      | daniel morales | 333012797 |
| gdl34      | david rojas | 333330734 |
| gdl36      | alna delia mtr. | 36335022 |
| wty81     | gabriela luna casa | 8181050416 |
| wty82     | diego rivera | 811220535 |
| wty83     | alejandra zaragoza | 8132600010 |
| pue22     | berenice gallardo | 2220310032 |
| vdm88     | pedro mayen | 554235406 |
| vdm57     | enrique apillar | 5590154857 |
| vdm88     | laura pino | 5573007588 |
-----
10 rows in set (0.016 sec)

```

Figura 77. Validación de la carga de datos en MaríaDB.

### 3.4 Diseño de la base de datos.

#### B) Creación de las entidades.

Los datos explorados y limpios se crea un esquema multidimensional de tipo datamart de acuerdo con las características del conjunto de datos.

#### C) Generación de los atributos para cada entidad.

Características propias de la entidad, que posee atributos básicos que la caracterizan. Los atributos se modelan como columnas de la entidad; la forma de diferenciarlas es por medio de atributos, y cada una de ellas debe tener por lo menos uno diferente.

Los atributos para este caso son:

- **Atributos de la dimensión venden:** idvende, cantidad, categoría, venta total, comentario, id distribuidor, idvendedor, idtienda, idproducto, idregion, idfecha.
- **Atributos de la tabla vendedor:** id\_vendedor, nombre y teléfono.
- **Atributos de la tabla producto:** id\_producto, descripción y precio.
- **Atributos de la tabla tienda:** id\_tienda, nombre, encargado, estado, dirección y teléfono.
- **Atributos de la tabla distribuidor:** id\_distribuidor, distribuidor y estado.
- **Atributos de la tabla región:** id\_region y estado.
- **Atributos de la tabla fecha:** id\_fecha, fecha y mes.

#### **D) Constraints.**

La integridad de una base de datos se asegura mediante reglas y restricciones; Como restricciones nos encontramos con llave primaria, llave foránea, valor no nulo, único valor y de revisión.

#### **E) Integridad de la entidad (Primary Key).**

La integridad de la entidad afecta a las llaves primarias, lo que establece la totalidad de entidades que todas las tablas deben tener, como una llave primaria y que las columnas elegidas como llaves deben ser únicas y no nulas.

#### **F) Llaves primarias y foráneas.**

Llave primaria (primary key) es el atributo que identifica de manera única a un registro o renglón de la tabla, tiene un valor único y diferente para cada fila de una tabla, no puede contener valores nulos.

**Llaves primarias:** id\_vende, id\_distribuidor, id\_vendedor, id\_fecha, id\_region, id\_tienda y id\_producto.

#### **G) Integridad referencial (Foreign Key).**

La integridad referencial afecta a las llaves foráneas, teniendo en cuenta que es el valor que hace énfasis a la llave primaria de otra relación (tabla). Esta integridad indica que los valores de la llave foránea en la relación hijo deben corresponderse con los valores de la primary key en la relación padre o bien ser nulos, si se admiten nulos. Es decir, que uno puede tener registros “huérfanos” en la tabla hijo o, dicho de otra forma, que no tengan alguna relación con un registro en la tabla padre.

Llave secundaria (foreign key) es la llave primaria de una tabla y al mismo tiempo forma parte de otra tabla como atributo.

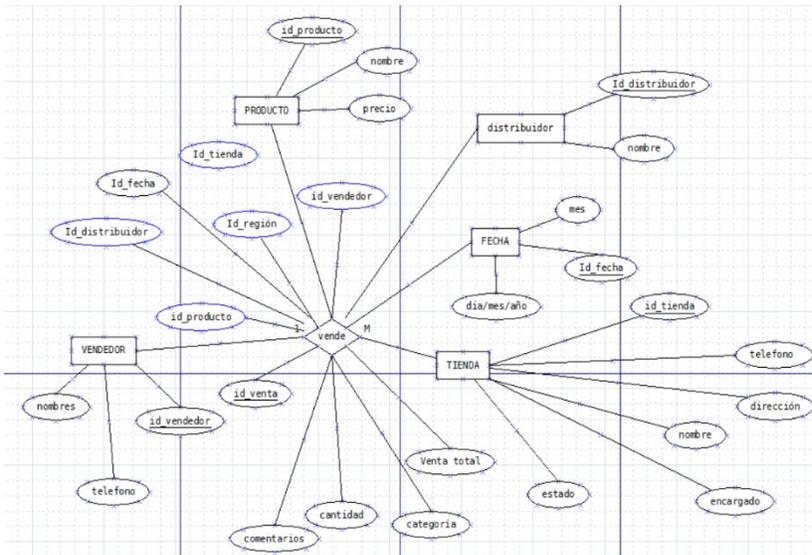
#### **H) Llaves foráneas.**

id\_producto, id\_tienda, id\_region, id\_vendedor, id\_fecha, id\_distribuidor.

### **3.5 E-R (Entidad-Relación).**

Para el diseño de una base de datos relacional se tiene que pasar por tres etapas básicas: el modelado conceptual, lógico y físico. A continuación, nos enfocaremos en el **modelado conceptual o Entidad-Relación**, siendo un modelo de datos semántico.

En este caso teórico-práctico se construyeron las características generales de la base de datos, usando el software "Día", herramienta de diseño conceptual para realizar el modelado representando las entidades, propiedades y relaciones: los diagramas E/R (Figura 78).



**Figura 78.** Modelado de entidades, propiedades y relaciones.

Nota: Modelado de entidades en el cual se visualizan las dimensiones, tabla de hechos de la base de datos. Creación propia.

Es importante mencionar que para el modelo E/R las entidades se representan con un rectángulo, los atributos se presentan con una elipse y las relaciones con una línea recta en la cual se muestra el tipo de relación, en este caso son

relaciones de uno a muchos por el tipo de base de datos (Figura 78).

Retomando el modelo E/R (entidad-relación) los vínculos o relaciones se representan por medio de un rombo, en el cual contiene el verbo de acción que relacionará a dichas entidades.

Se muestra las entidades, atributos y relaciones del diseño de la base de datos las cuales utilizaremos en el modelo datamart.

### **3.6 Relacional.**

En el modelo relacional se convierten las entidades en relaciones, los vínculos en asociaciones y se implementan llaves primarias.

El modelo relacional se ocupa de tres aspectos principales de la información:

1. *La estructura de datos.*
2. *La manipulación de datos.*
3. *La integridad de datos.*

En el modelo E/R se crean las entidades, los atributos y las relaciones, dicha información se exporta a MariaDB para

crear llaves primarias, foráneas y el tipo de dato para cada relación.

### **Tabla Producto.**

- Idproducto (PK) tipo de dato CHAR de longitud (15).
- Descripción VARCHAR de longitud (100).
- Precio Double (6,2).

### **Tabla Distribuidor.**

- Iddistribuidor CHAR (10).
- Distribuidor VARCHAR (45).
- Estado VARCHAR (45).

### **Tabla fecha.**

- Idfecha CHAR (15).
- Fecha Date.
- Mes CHAR (10).

### **Tabla región.**

- Idregion CHAR (8).
- Estado VARCHAR (45).

### **Tabla tienda.**

- Idtienda CHAR (15).
- Nombre VARCHAR (100).

- Encargado VARCHAR (45).
- Estado VARCHAR (45).
- Dirección VARCHAR (100).
- Teléfono BIGINT (10).

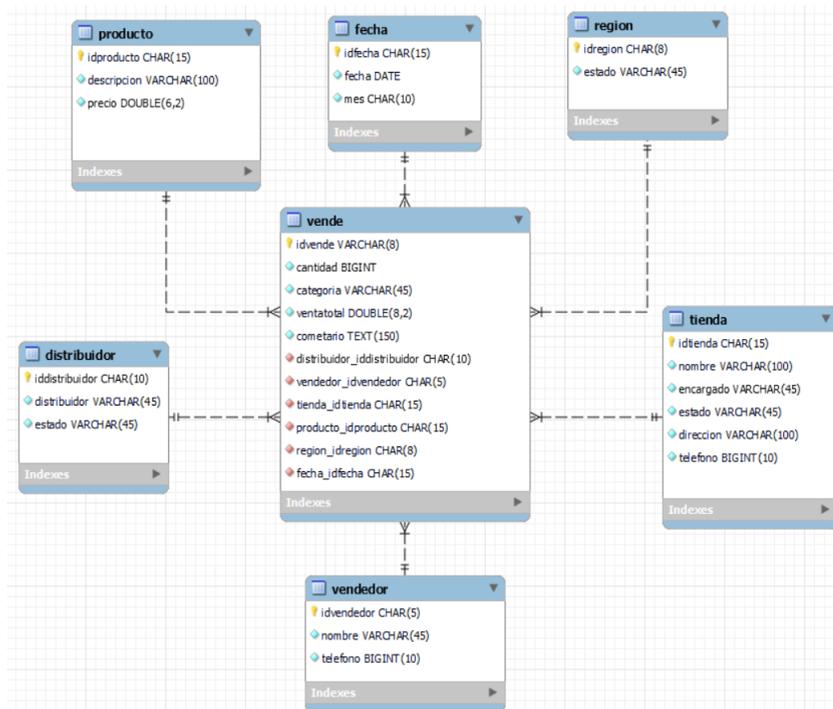
**Tabla vendedor.**

- Idvendedor CHAR (5).
- Nombre VARCHAR (45).
- Teléfono BIGINT (10).

**Tabla vende.**

- Idvende VARCHAR (8).
- Cantidad BIGINT (100).
- Categoría VARCHAR (45).
- Ventatotal DOUBLE (8,2).
- Comentario TEXT (150).
- Distribuidor\_iddistribuidor CHAR (10).
- Vendedor\_idvendedor CHAR (5).
- Tienda\_idtienda CHAR (15).
- Producto\_idproducto CHAR (15).
- Region\_idregion CHAR (8).
- Fecha\_idfecha CHAR (15).

El modelo relacional en MySQL Workbench permite observar las entidades, atributos, relaciones y modelos de la base de datos multidimensional (Figura 79).



**Figura 79.** Modelo relacional en MySQL Workbench

Nota: Estructura de la base de datos (Modelo relacional), en el cual se visualizan las dimensiones la tabla de hechos en una base de datos. Creación propia.

En este caso práctico en primer lugar se realizó el esquema relacional en MySQL Workbench y después se generó el código que se emplea para crear la base de datos en MaríaDB que es el SGBD que se estará usando. Se utilizó una opción en Workbench para poder generar un plugin que tiene incluido el programa MySQL Workbench para poder realizar el código, como se mostrará en los siguientes pasos.

Paso 1. Dirigirse a herramientas y seleccionar Forward Engineer como se muestra en la (Figura 80).

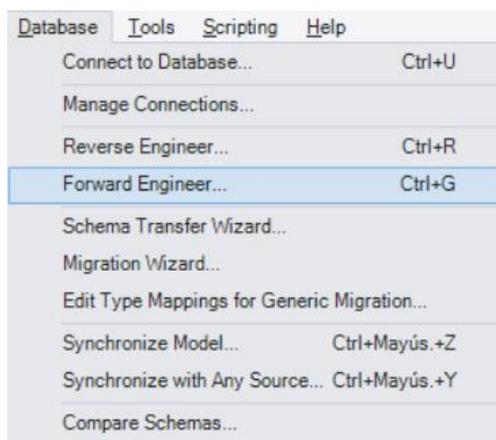


Figura 80. Database.

En la ventana de diálogo “connection options” seleccionar la casilla next (Figura 81) posteriormente elegir el cuadro de diálogo “set options for database to be created” escoger la casilla next (Figura 82).

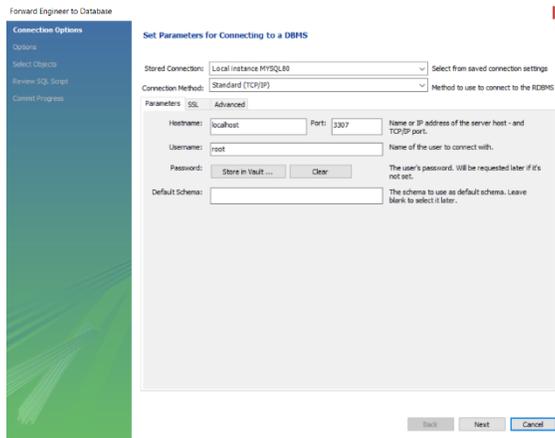


Figura 81. Opción de conexión.

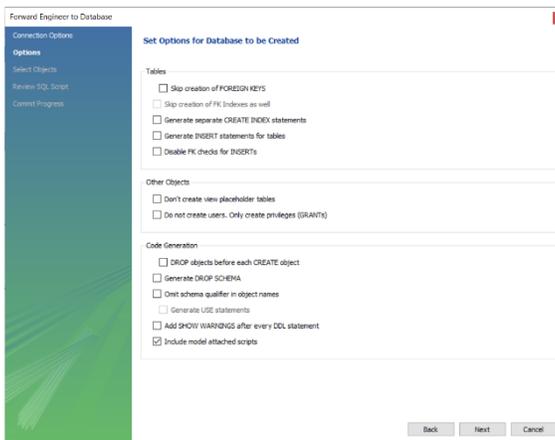


Figura 82. Seleccionar campos.

En la ventana de diálogo select objects to forward engineer seleccionar la casilla next (Figura 83), después

generar el código que se va a utilizar, elegir la casilla next para continuar (Figura 84).

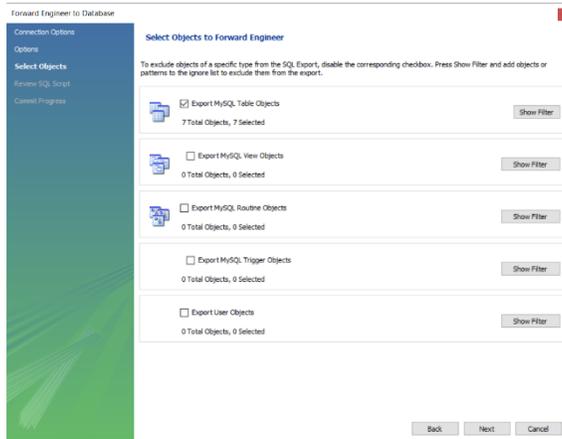


Figura 83. Seleccionar objetos.



Figura 84. Código creado.

El código generado en la Figura 85-86, se realizó con el plugin que usamos de MySQL que viene descrito en los pasos anteriores. Este plugin viene instalado por default, es funcional para generar códigos con tablas realizadas en el modelo relacional (Figura 79).

```

Forward Engineer to Database
Review the SQL Script to be Executed
This script will now be executed on the DB server to create your databases.
You may make changes before executing.

51 -- Table 'mydb`.`distribuidor`
52 -----
53 CREATE TABLE IF NOT EXISTS `mydb`.`distribuidor` (
54   `distribuidor` CHAR(10) NOT NULL,
55   `distribuidor` VARCHAR(45) NOT NULL,
56   `telefono` VARCHAR(45) NOT NULL,
57   PRIMARY KEY (`distribuidor`)
58 )
59 ENGINE = InnoDB;
60
61
62 -- Table 'mydb`.`region`
63 -----
64 CREATE TABLE IF NOT EXISTS `mydb`.`region` (
65   `region` CHAR(10) NOT NULL,
66   `telefono` VARCHAR(45) NOT NULL,
67   PRIMARY KEY (`region`)
68 )
69 ENGINE = InnoDB;
70
71
72 -- Table 'mydb`.`vendedor`
73 -----
74 CREATE TABLE IF NOT EXISTS `mydb`.`vendedor` (
75   `vendedor` CHAR(10) NOT NULL,
76   `telefono` VARCHAR(45) NOT NULL,
77   `vendedor` VARCHAR(45) NOT NULL,
78   `telefono` VARCHAR(45) NOT NULL,
79   `direccion` VARCHAR(255) NOT NULL,
80   `telefono` VARCHAR(45) NOT NULL,
81   PRIMARY KEY (`vendedor`)
82 )
83 ENGINE = InnoDB;

```

Figura 85. Entidades y atributos de las tablas.

```

Forward Engineer to Database
Review the SQL Script to be Executed
This script will now be executed on the DB server to create your databases.
You may make changes before executing.

86 -- Table 'mydb`.`vendedor`
87 -----
88 CREATE TABLE IF NOT EXISTS `mydb`.`vendedor` (
89   `vendedor` VARCHAR(45) NOT NULL,
90   `telefono` BIGINT NOT NULL,
91   `vendedor` VARCHAR(45) NOT NULL,
92   `vendedor` DOUBLE(10) NOT NULL,
93   `vendedor` TEXT(10) NOT NULL,
94   `distribuidor_distribuidor` CHAR(10) NOT NULL,
95   `vendedor_distribuidor` CHAR(10) NOT NULL,
96   `telefono_distribuidor` CHAR(10) NOT NULL,
97   `telefono_distribuidor` CHAR(10) NOT NULL,
98   `region_distribuidor` CHAR(10) NOT NULL,
99   `telefono_distribuidor` CHAR(10) NOT NULL,
100  PRIMARY KEY (`vendedor`),
101  INDEX `fk_vendedor_distribuidor_id` (`distribuidor_distribuidor` ASC) VISIBLE,
102  INDEX `fk_vendedor_distribuidor_id` (`vendedor_distribuidor` ASC) VISIBLE,
103  INDEX `fk_vendedor_distribuidor_id` (`telefono_distribuidor` ASC) VISIBLE,
104  INDEX `fk_vendedor_distribuidor_id` (`vendedor_distribuidor` ASC) VISIBLE,
105  INDEX `fk_vendedor_distribuidor_id` (`region_distribuidor` ASC) VISIBLE,
106  INDEX `fk_vendedor_distribuidor_id` (`telefono_distribuidor` ASC) VISIBLE,
107  CONSTRAINT `fk_vendedor_distribuidor`
108  FOREIGN KEY (`distribuidor_distribuidor`)
109  REFERENCES `mydb`.`distribuidor` (`distribuidor`)
110  ON DELETE NO ACTION
111  ON UPDATE NO ACTION,
112  CONSTRAINT `fk_vendedor_distribuidor`
113  FOREIGN KEY (`vendedor_distribuidor`)
114  REFERENCES `mydb`.`vendedor` (`vendedor`)
115  ON DELETE NO ACTION
116  ON UPDATE NO ACTION,
117  CONSTRAINT `fk_vendedor_distribuidor`

```

Figura 86. Código generado.

Seleccionar los objetos a crear en el código, elegir en el cuadro de diálogo close (Figura 87).

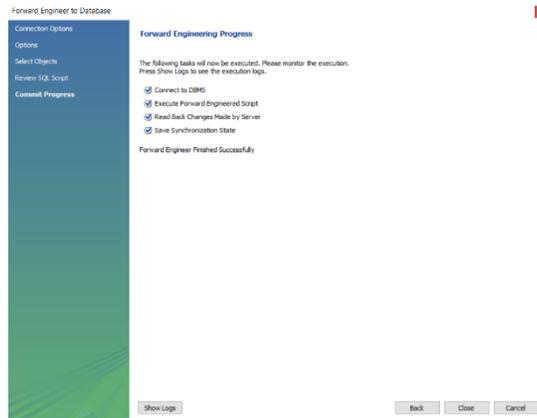


Figura 87. Seleccionar objetos y finalización del proceso.

### 3.7 Modelado Físico.

En el diseño de la base de datos, se aborda la etapa del modelado físico. Primeramente, se genera el código en MySQL Workbench para el SGBD.

Se abre XAMPP, se selecciona START (Figura 88) para correr el programa, en los módulos Apache y MySQL, no se presentó ninguna inconveniente (Figura 89), por consecuencia el SGBD funciono adecuadamente, seleccionar el recuadro Shell para abrir el modo comando.

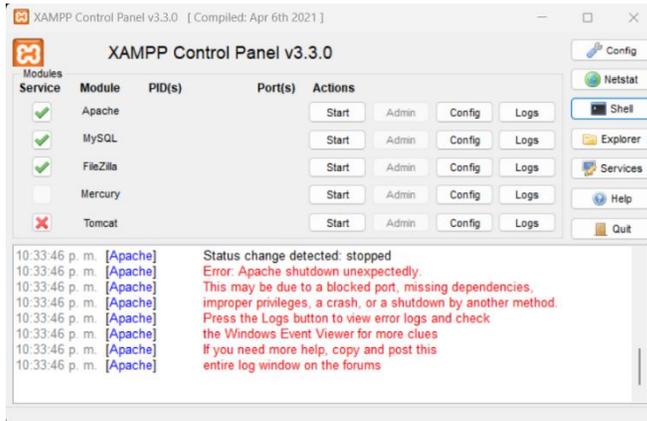


Figura 88. Panel de control XAMPP.

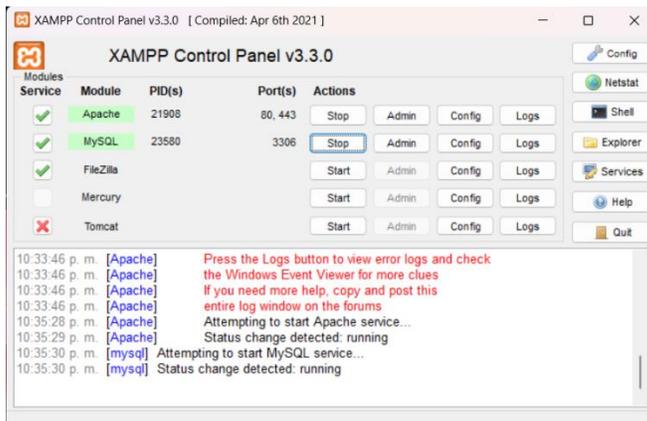
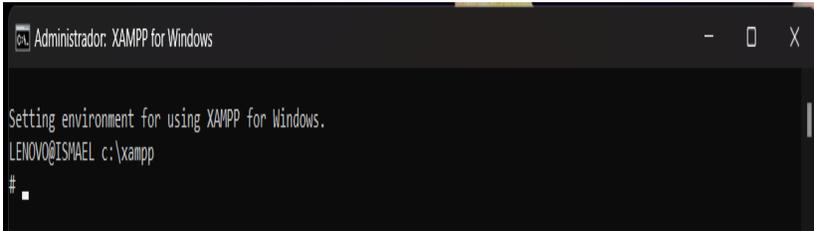


Figura 89. Correr los servicios de XAMPP.

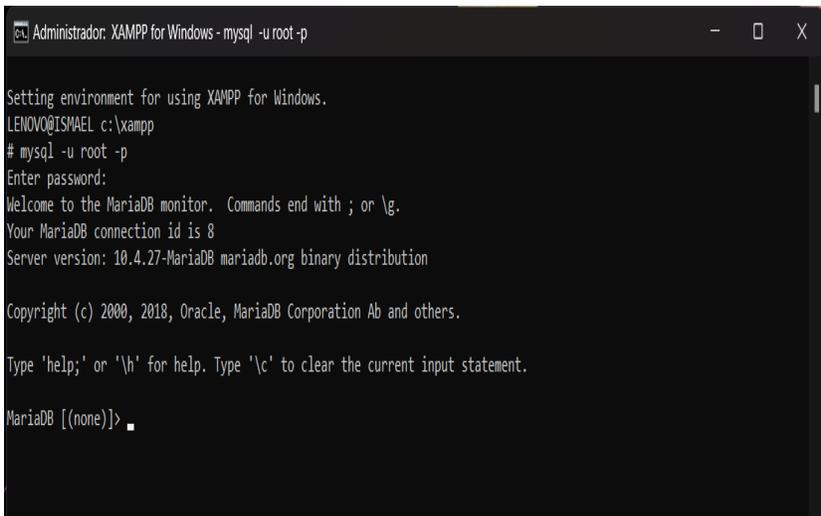
Se abre el modo consola (Figura 90).



```
Administrador: XAMPP for Windows
Setting environment for using XAMPP for Windows.
LENOVO@ISMAEL c:\xampp
#
```

Figura 90. Modo consola MaríaDB.

Escribir la sintaxis “mysql -u root -p” para poder correr el SGBD (Figura 91).



```
Administrador: XAMPP for Windows - mysql -u root -p
Setting environment for using XAMPP for Windows.
LENOVO@ISMAEL c:\xampp
# mysql -u root -p
Enter password:
Welcome to the MariaDB monitor. Commands end with ; or \g.
Your MariaDB connection id is 8
Server version: 10.4.27-MariaDB mariadb.org binary distribution

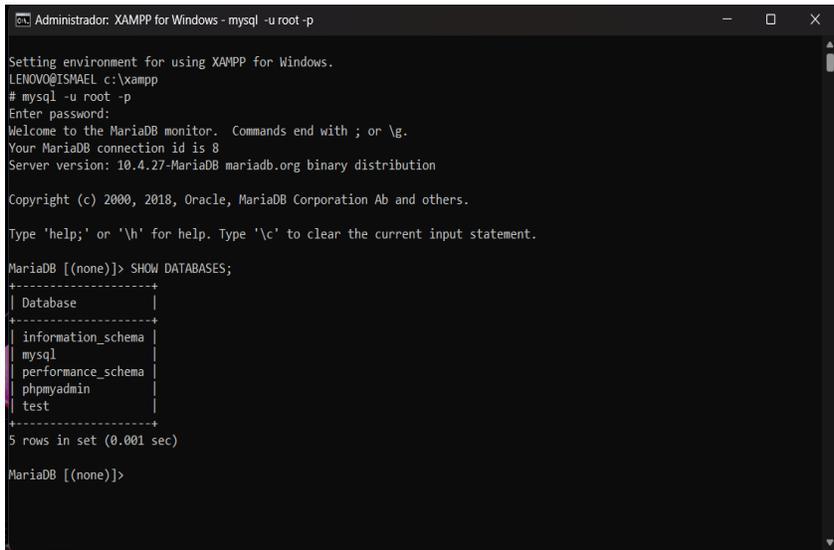
Copyright (c) 2000, 2018, Oracle, MariaDB Corporation Ab and others.

Type 'help;' or 'h' for help. Type 'c' to clear the current input statement.

MariaDB [(none)]>
```

Figura 91. Administrador XAMPP.

Escribir “SHOW DATABASES” (Figura 92), dar enter para poder visualizar las bases de datos que están creadas en MariaDB.



```
Administrador: XAMPP for Windows - mysql -u root -p
Setting environment for using XAMPP for Windows.
LENOVO@ISMAEL c:\xampp
# mysql -u root -p
Enter password:
Welcome to the MariaDB monitor.  Commands end with ; or \g.
Your MariaDB connection id is 8
Server version: 10.4.27-MariaDB mariadb.org binary distribution

Copyright (c) 2000, 2018, Oracle, MariaDB Corporation Ab and others.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

MariaDB [(none)]> SHOW DATABASES;
+-----+
| Database |
+-----+
| information_schema |
| mysql |
| performance_schema |
| phpmyadmin |
| test |
+-----+
5 rows in set (0.001 sec)

MariaDB [(none)]>
```

Figura 92. Bases de datos en MaríaDB.

Crear la base de datos “Tesis” (Figura 93).

```
Administrador: XAMPP for Windows - mysql -u root -p

Setting environment for using XAMPP for Windows.
LENOVO@ISMAEL c:\xampp
# mysql -u root -p
Enter password:
Welcome to the MariaDB monitor.  Commands end with ; or \g.
Your MariaDB connection id is 9
Server version: 10.4.27-MariaDB mariadb.org binary distribution

Copyright (c) 2000, 2018, Oracle, MariaDB Corporation Ab and others.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

MariaDB [(none)]> SHOW DATABASES;
+-----+
| Database |
+-----+
| information_schema |
| mysql |
| performance_schema |
| phpmysqladmin |
| test |
+-----+
5 rows in set (0.001 sec)

MariaDB [(none)]> CREATE DATABASE tesis;
```

Figura 93. Código para generar la base de datos.

La base de datos se ha creado exitosamente (Figura 94).

```
Administrador: XAMPP for Windows - mysql -u root -p

Setting environment for using XAMPP for Windows.
LENOVO@ISMAEL c:\xampp
# mysql -u root -p
Enter password:
Welcome to the MariaDB monitor.  Commands end with ; or \g.
Your MariaDB connection id is 9
Server version: 10.4.27-MariaDB mariadb.org binary distribution

Copyright (c) 2000, 2018, Oracle, MariaDB Corporation Ab and others.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

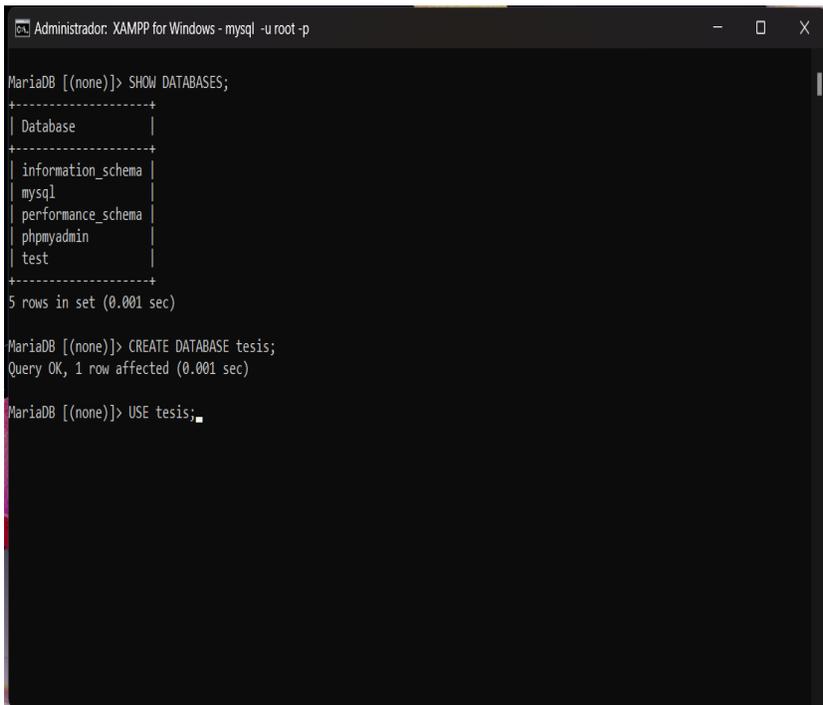
MariaDB [(none)]> SHOW DATABASES;
+-----+
| Database |
+-----+
| information_schema |
| mysql |
| performance_schema |
| phpmysqladmin |
| test |
+-----+
5 rows in set (0.001 sec)

MariaDB [(none)]> CREATE DATABASE tesis;
Query OK, 1 row affected (0.001 sec)

MariaDB [(none)]> .
```

Figura 94. Creación de la base de datos.

Escribir el código USE tesis; para poder usar la base de datos creada (Figura 95).



```
Administrador: XAMPP for Windows - mysql -u root -p
MariaDB [(none)]> SHOW DATABASES;
+-----+
| Database |
+-----+
| information_schema |
| mysql |
| performance_schema |
| phpmyadmin |
| test |
+-----+
5 rows in set (0.001 sec)

MariaDB [(none)]> CREATE DATABASE tesis;
Query OK, 1 row affected (0.001 sec)

MariaDB [(none)]> USE tesis;_
```

Figura 95. Base de datos creada.

Se ejecuta el código que se generó en MySQL Workbench. (figura 96).

```
Administrador: XAMPP for Windows - mysql -> root ->
MariaDB [(none)]> CREATE DATABASE tesis;
Query OK, 1 row affected (0.001 sec)

MariaDB [(none)]> USE tesis;
Database changed
MariaDB [tesis]> CREATE TABLE IF NOT EXISTS 'tesis'. 'vendedor' (
->   'idvendedor' CHAR(5) NULL,
->   'nombre' VARCHAR(45) NULL,
->   'telefono' BINARY(10) NULL,
->   PRIMARY KEY ('idvendedor'))
-> ENGINE = InnoDB;
Query OK, 0 rows affected (0.012 sec)

MariaDB [tesis]>
MariaDB [tesis]> CREATE TABLE IF NOT EXISTS 'tesis'. 'producto' (
->   'idproducto' CHAR(15) NULL,
->   'descripcion' VARCHAR(100) NULL,
->   'precio' DOUBLE(6,2) NULL,
->   PRIMARY KEY ('idproducto'))
-> ENGINE = InnoDB;
Query OK, 0 rows affected (0.006 sec)

MariaDB [tesis]>
MariaDB [tesis]> CREATE TABLE IF NOT EXISTS 'tesis'. 'fecha' (
->   'idfecha' CHAR(15) NULL,
->   'fecha' DATE NULL,
->   'mes' CHAR(10) NULL,
->   PRIMARY KEY ('idfecha'))
-> ENGINE = InnoDB;
Query OK, 0 rows affected (0.006 sec)

MariaDB [tesis]>
MariaDB [tesis]> CREATE TABLE IF NOT EXISTS 'tesis'. 'distribuidor' (
->   'iddistribuidor' CHAR(10) NULL,
->   'distribuidor' VARCHAR(45) NULL,
->   'estado' VARCHAR(45) NULL,
->   PRIMARY KEY ('iddistribuidor'))
-> ENGINE = InnoDB;
Query OK, 0 rows affected (0.005 sec)

MariaDB [tesis]>
MariaDB [tesis]> CREATE TABLE IF NOT EXISTS 'tesis'. 'region' (
->   'idregion' CHAR(3) NULL,
->   'estado' VARCHAR(45) NULL,
->   PRIMARY KEY ('idregion'))
-> ENGINE = InnoDB;
Query OK, 0 rows affected (0.005 sec)
```

Figura 96. Código de la base de datos.

Se tuvo que cambiar la restricción de integridad referencial ON DELETE NO ACTION por un ON DELETE SET NULL, esto realizó para que permitiera integrar los datos.

La creación de la base de datos esta lista (Figura 97), ya que los Query corrieron de forma correcta.

```

Administrador: XAMPP for Windows: mysql -u root -p
MariaDB [tesis]:
MariaDB [tesis]: CREATE TABLE IF NOT EXISTS `tesis`.`vende` (
-> `idvende` VARCHAR(4) NULL,
-> `cantidad` DOUBLE NULL,
-> `categoria` VARCHAR(45) NULL,
-> `ventatotal` DOUBLE(8,2) NULL,
-> `comentario` TEXT(150) NULL,
-> `distribuidor_iddistribuidor` CHAR(10) NULL,
-> `vendedor_idvendedor` CHAR(5) NULL,
-> `tienda_idtienda` CHAR(15) NULL,
-> `producto_idproducto` CHAR(15) NULL,
-> `region_idregion` CHAR(3) NULL,
-> `fecha_idfecha` CHAR(15) NULL,
-> PRIMARY KEY (`idvende`),
-> CONSTRAINT `fk_vende_distribuidor`
-> FOREIGN KEY (`distribuidor_iddistribuidor`)
-> REFERENCES `tesis`.`distribuidor` (`iddistribuidor`)
-> ON DELETE SET NULL
-> ON UPDATE SET NULL
-> CONSTRAINT `fk_vende_vendedor1`
-> FOREIGN KEY (`vendedor_idvendedor`)
-> REFERENCES `tesis`.`vendedor` (`idvendedor`)
-> ON DELETE SET NULL
-> ON UPDATE SET NULL
-> CONSTRAINT `fk_vende_tienda1`
-> FOREIGN KEY (`tienda_idtienda`)
-> REFERENCES `tesis`.`tienda` (`idtienda`)
-> ON DELETE SET NULL
-> ON UPDATE SET NULL
-> CONSTRAINT `fk_vende_producto1`
-> FOREIGN KEY (`producto_idproducto`)
-> REFERENCES `tesis`.`producto` (`idproducto`)
-> ON DELETE SET NULL
-> ON UPDATE SET NULL
-> CONSTRAINT `fk_vende_region1`
-> FOREIGN KEY (`region_idregion`)
-> REFERENCES `tesis`.`region` (`idregion`)
-> ON DELETE SET NULL
-> ON UPDATE SET NULL
-> CONSTRAINT `fk_vende_fecha1`
-> FOREIGN KEY (`fecha_idfecha`)
-> REFERENCES `tesis`.`fecha` (`idfecha`)
-> ON DELETE SET NULL
-> ON UPDATE SET NULL)
-> ENGINE = InnoDB;
Query OK, 0 rows affected (0.020 sec)

MariaDB [tesis]:

```

Figura 97. Creación de la base de datos en MaríaDB.

### 3.8 Dashboard.

El dashboard o tablero de control se realizó con Power BI (Business Intelligence).

#### a) Vistas.

Este programa permite visualizar los datos, cabe mencionar que las vistas creadas en María DB visualizaron datos importantes del dashboard.

En esta vista se calculó un promedio de venta en piezas por mes (Figura 98).

```
Administrador: XAMPP for Windows - mysql -u root -p
MariaDB [tesis]> CREATE VIEW prompzaxmes AS SELECT F.mes AS 'mensual',AVG(V.cantidad) AS 'promedio' FROM vende AS V INNER JOIN fecha AS F ON V.fecha_idfecha=F.idfecha GROUP BY F.mes;
Query OK, 0 rows affected (0.007 sec)
MariaDB [tesis]>
```

Figura 98. Venta promedio mensual por piezas.

Se realizó una vista para poder obtener la venta total (Figura 99).

```
Administrador: XAMPP for Windows - mysql -u root -p
MariaDB [tesis]> CREATE VIEW ventatotal AS SELECT SUM(ventatotal) AS 'venta total' FROM vende;
Query OK, 0 rows affected (0.006 sec)
MariaDB [tesis]>
```

Figura 99. Venta total.

Venta por mes, esta vista permite visualizar las ganancias por periodos (Figura 100).

```
Administrador: XAMPP for Windows - mysql -u root -p
MariaDB [tesis]> CREATE VIEW ventaxmes AS SELECT F.mes AS 'mes',SUM(ventatotal) AS 'total' FROM vende AS V INNER JOIN fecha AS F ON V.fecha_idfecha=idfecha GROUP BY mes;
Query OK, 0 rows affected (0.005 sec)
MariaDB [tesis]>
```

Figura 100. Venta por mes.

Se elabora una vista donde muestra la venta por región (Figura 101).

```
Administrador: XAMPP for Windows - mysql -u root -p
MariaDB [tesis]> CREATE VIEW ventaxregion AS SELECT R.estado AS 'estado', SUM(V.ventatotal) AS 'total' FROM vende AS V I
INNER JOIN region AS R ON V.region_idregion=R.idregion GROUP BY R.idregion;
Query OK, 0 rows affected (0.006 sec)

MariaDB [tesis]>
```

Figura 101. Venta por región.

Se genera una vista para visualizar la venta por tienda (Figura 102).

```
Administrador: XAMPP for Windows - mysql -u root -p
MariaDB [tesis]> CREATE VIEW ventaxtienda AS SELECT T.idtienda,T.nombre,V.ventatotal FROM vende AS V INNER JOIN tienda
AS T ON V.tienda_idtienda=T.idtienda GROUP BY T.idtienda ORDER BY V.ventatotal ASC;
Query OK, 0 rows affected (0.005 sec)

MariaDB [tesis]>
```

Figura 102. Venta por tienda.

Se realizó una vista en la cual muestra las ventas por vendedor (Figura 103).

```
Administrador: XAMPP for Windows - mysql -u root -p

MariaDB [tesis]> CREATE VIEW ventaxvendedor AS SELECT vendedor.idvendedor AS 'idvendedor',vendedor.nombre AS 'nombre',SUM(vende.ventatotal) AS 'total' FROM vende AS vende INNER JOIN vendedor AS vendedor ON vendedor_idvendedor=idvendedor GROUP BY vendedor.nombre ORDER BY total ASC;
Query OK, 0 rows affected (0.004 sec)

MariaDB [tesis]>
```

Figura 103. Venta por vendedor.

Se inicia el programa Power Bi, se realiza el dashboard o tablero de control que se entregará a cliente para la comprensión de la información, para ello se importan los datos a power BI el cual muestra el interfaz del programa (Figura 104).

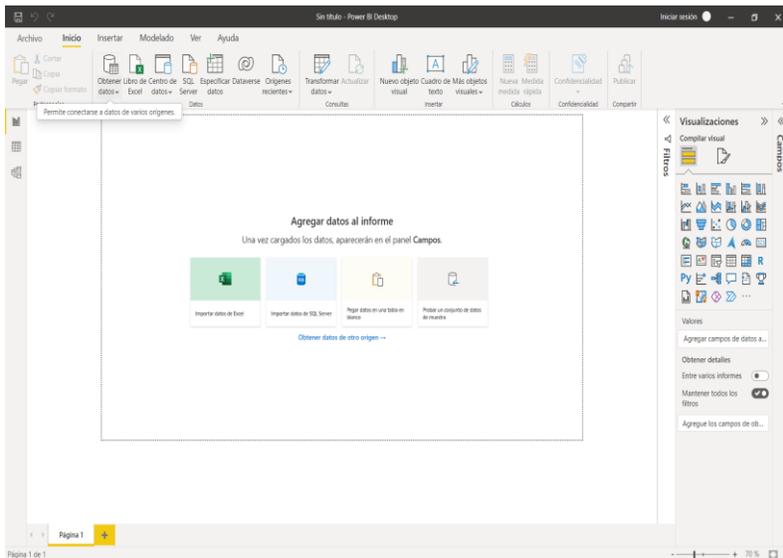


Figura 104. Importación de datos a Power Bi.

En la importación de la base de datos, seleccionar la herramienta obtener, permitiendo acceder a cualquier SGBD para poder visualizar (Figura 105).

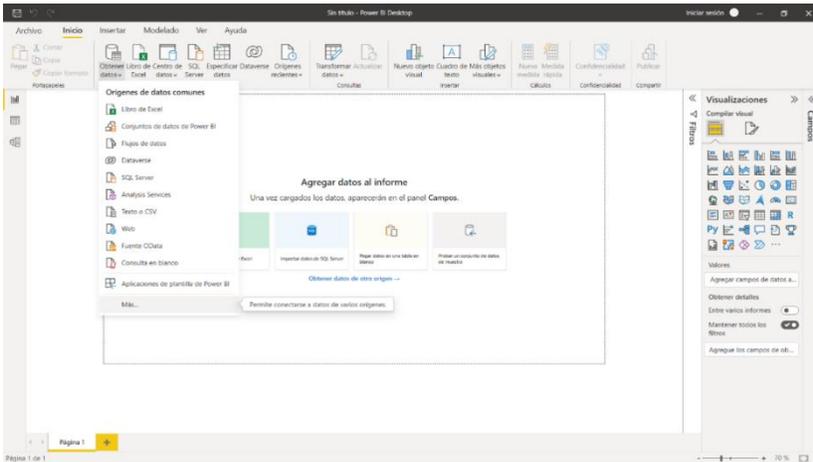


Figura 105. Orígenes de datos.

Seleccionar el origen de la base de datos que en este caso es MySQL (Figura 106).

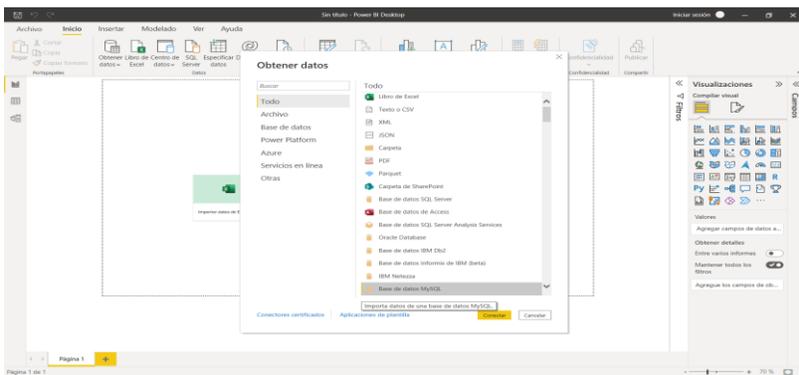


Figura 106. Obtener datos.

Colocar el nombre del servidor que es localhost:3306 y el nombre de la base de datos en este caso es tesis (Figura 107).

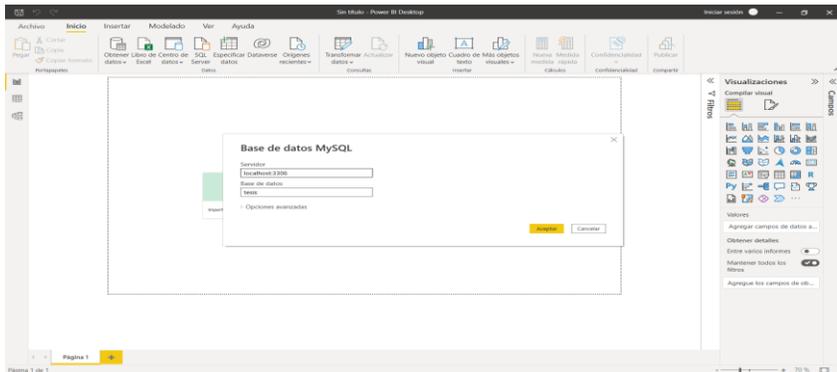


Figura 107. Obtener datos de MySQL.

Seleccionar todas las tablas para identificar las visitas realizadas (Figura 108).

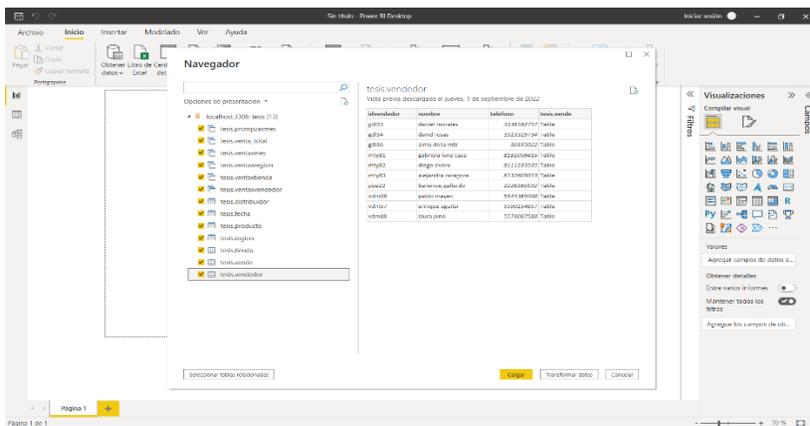


Figura 108. Seleccionar tablas a cargar.

## Cargar las tablas a Power BI (Figura 109).

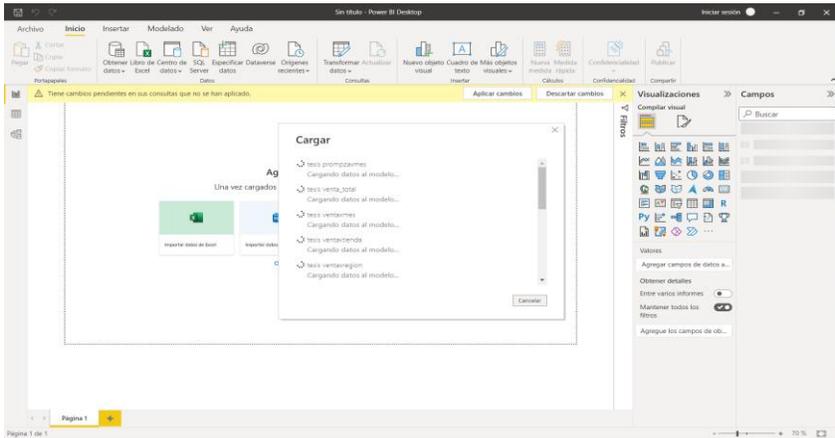


Figura 109. Carga de tablas.

Como se visualiza las tablas se cargaron de forma correcta y estas se pueden observar en la parte derecha del interfaz (Figura 110).

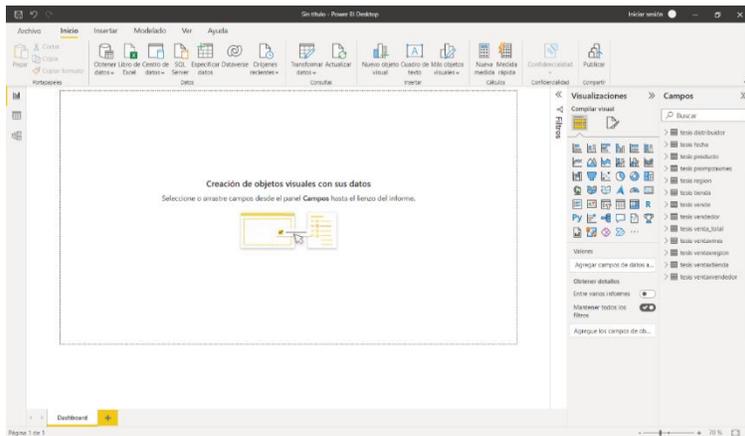


Figura 110. Power BI Desktop.

### 3.9 Modelado de datos.

Para iniciar el modelado de datos, ubicar la opción propiedades y seleccionar la visualización del modelo, ésta muestra todas las tablas realizadas en MariaDB, las vistas, se crean las relaciones tal como se hizo en MySQL Workbench (Figura 111).

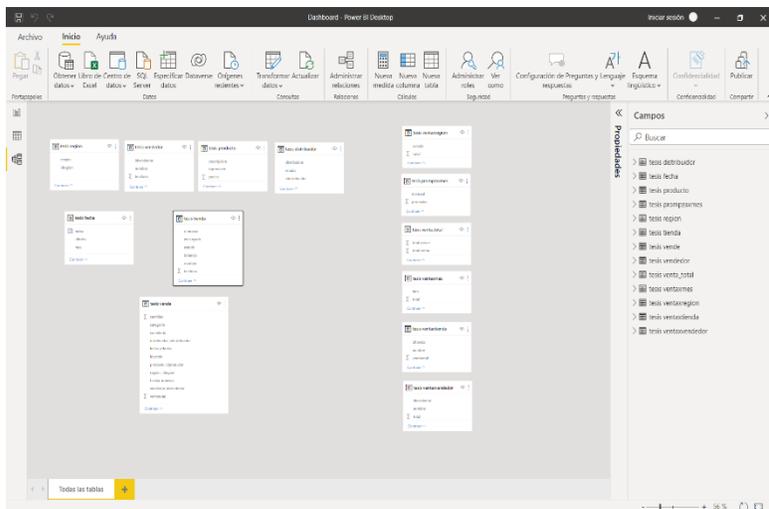


Figura 111. Modelo de datos.

Se generan las relaciones de las tablas en Power Bi (Figura 112), dichas relaciones se producen para poder usar de forma correcta la base de datos, ya que estas uniones logran

que las tablas al insertar visualizaciones sean más dinámicas e interactúan entre sí.

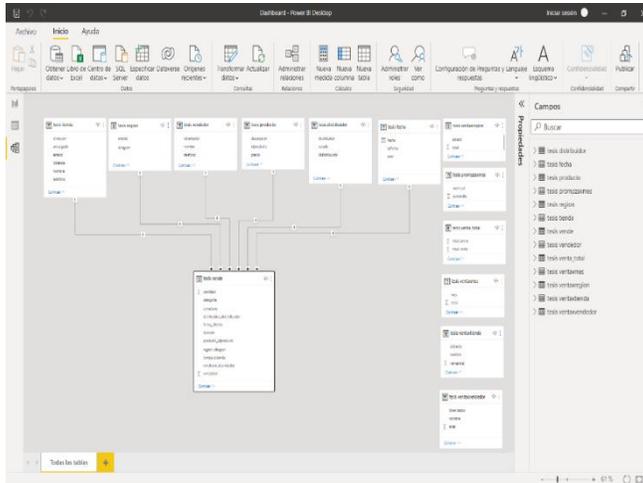


Figura 112. Relación de tablas en Power Bi

### 3.10 Construcción del dashboard.

En la creación de esta herramienta se agregan las visualizaciones de segmentación de datos (Figura 113), la cual ayuda a ver los elementos de la base de datos, filtrando y/o segmentando por periodo.

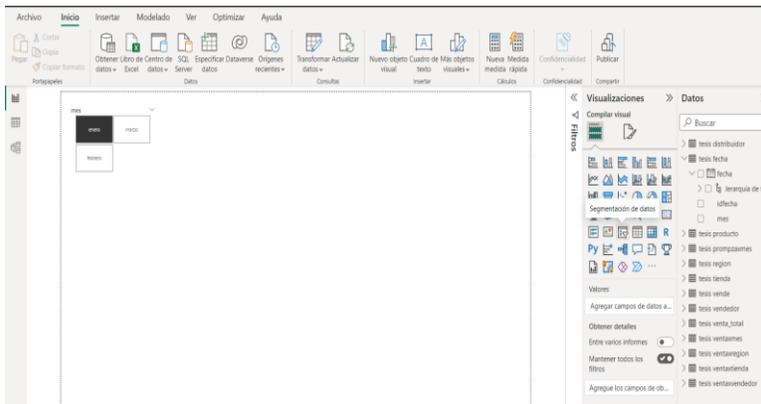


Figura 113. Segmentación de datos.

Posteriormente se implementó el uso de una tarjeta de varias filas (Figura 114), en la cual se observarán las ventas en pesos/piezas y fue creada con una vista (Figura 115), que se realizó con MaríaDB.

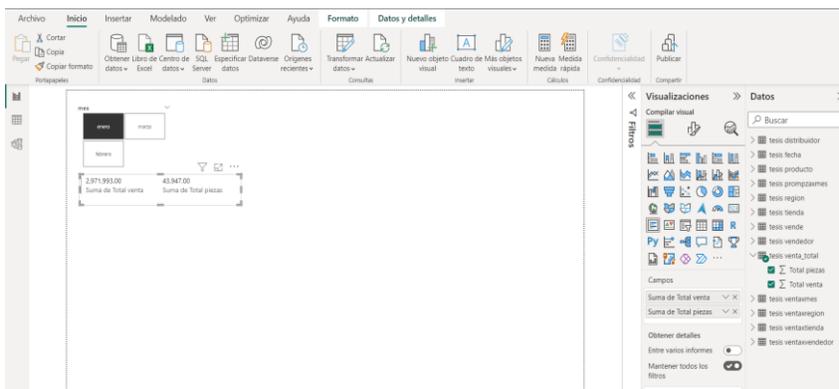


Figura 114. Visualización de datos de tarjeta.

```

Administrador: XAMPP for Windows - mysql -u root -p
MariaDB [tesis]> CREATE VIEW venta_total AS SELECT SUM(ventatotal) AS 'Total venta', SUM(cantidad) AS 'Total piezas' FROM vende;
Query OK, 0 rows affected (0.026 sec)

```

Figura 115.venta total.

Continuando con el proceso se usa gráfico de cascada (Figura 116), que permite ver las ventas por mes en pesos y el total de las ventas, esta operación se realizó con otra vista (Figura 117).

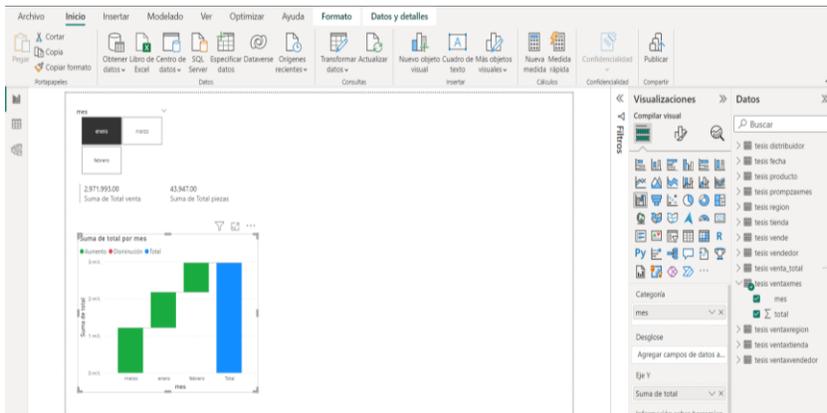


Figura 116. Uso de gráficos de cascada.

```

Administrador: XAMPP for Windows - mysql -u root -p
MariaDB [tesis]> CREATE VIEW ventaxmes AS SELECT F.mes AS 'mes',SUM(ventatotal) AS 'total' FROM vende AS V INNER JOIN fecha AS F ON V.fecha_idfecha=idfecha GROUP BY mes;
Query OK, 0 rows affected (0.005 sec)

MariaDB [tesis]>

```

Figura 117. Venta total.



Se implementa el gráfico de columnas apiladas y de líneas, donde se visualizan las ventas por vendedor tanto en pesos como en piezas (Figura 120).

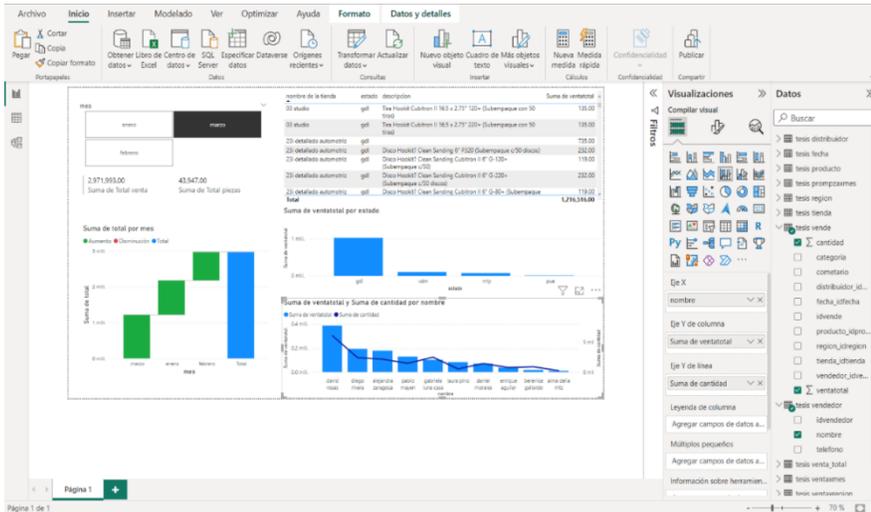


Figura 120. Gráfico de columnas apiladas.

Para la última visualización, se emplea una tarjeta de varias filas (Figura 121) que utiliza los campos de una vista creada con anterioridad (figura 122), mostrando las ventas por mes.

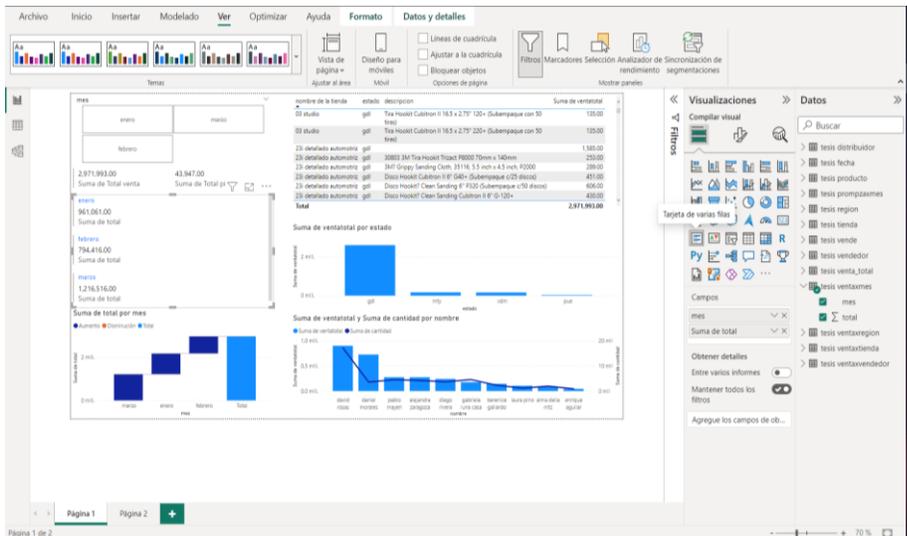


Figura 121. Tarjeta de varias filas.

```

Administrador: XAMPP for Windows - mysql -u root -p
MariaDB [tesis]> CREATE VIEW ventaxmes AS SELECT F.mes AS 'mes',SUM(ventatotal) AS 'total' FROM vende AS V INNER JOIN fecha AS F ON V.fecha_idfecha=idfecha GROUP BY mes;
Query OK, 0 rows affected (0.005 sec)

```

Figura 122. Vista venta por mes.

Se finaliza el dashboard, con él se visualizan las ventas generales de un trimestre, esto permitirá la toma de decisiones de forma más certera.

Se le dará formato para presentar al cliente el dashboard como se observa en la figura 123.



Figura 123. Dashboard finalizado.

Para la creación del KPI, se obtiene la información del reporte de ventas trimestral del año 2023 del departamento de ventas, así objetivos los generados por el cliente, estos tienen que ser medibles, cuantificables y reales.

Lo primero que se realiza, es la obtención de las ventas totales en pesos, piezas y mensuales, estas se comparan contra el objetivo, con la información se grafican KPI de crecimiento (Figura 124).

Mes	Venta Real	Venta Objetivo	%
<b>Enero</b>	\$961,061.00	\$1,000,000.00	96%
<b>Febrero</b>	\$794,416.00	\$900,000.00	88%
<b>Marzo</b>	\$1,216,516.00	\$1,200,000.00	101%

Mes	Venta Real Pzas.	Venta Objetivo Pzas.	%
<b>Enero</b>	9697.00	10000.00	97%
<b>Febrero</b>	15693.00	17000.00	92%
<b>Marzo</b>	18557.00	18000.00	10.%

Figura 124. Ventas trimestrales del año 2023 del departamento de ventas

En la barra de herramientas, seleccionar la opción especificar datos, aparecerá el recuadro: Crear tabla, como se observa en la Figura 125.

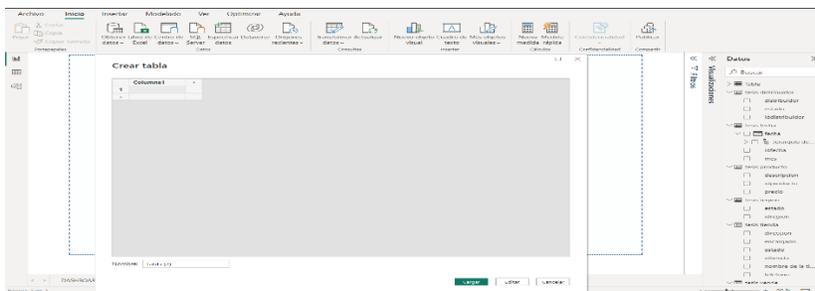


Figura 125. Elaboración del KPI.

Se integran los datos de las ventas y objetivo como se muestra en la Figura 126, se inicia crear la tabla venta en pesos, nombrándola “KPI \$”.

Crear tabla

Mes	Venta Real	Venta Objetivo	%	
1 Enero	\$961,061.00	\$1,000,000.00	96%	
2 Febrero	\$794,416.00	\$900,000.00	88%	
3 Marzo	\$1,216,516.00	\$1,200,000.00	101%	
+				

Nombre:

Carpet Editar Cancelar

DASHBOARD KPI \$ +

Página 2 de 2

Datos

Visualizaciones

Filtros

Buscar

- Tabla
- tesis distribuidor
  - distribuidor
  - estado
  - idistribuidor
- tesis fecha
  - fecha
  - jerarquia de...
  - isfecha
  - mes
- tesis producto
  - descripcion
  - idproducto
  - precio
- tesis region
  - estado
  - idregion
- tesis tienda
  - direccion
  - encargado
  - estado
  - is tienda
  - nombre de la s...
  - telefono
- tesis usuario

Figura 126. Tabla KPI pesos.

Se vuelve hacer el mismo proceso, pero ahora será para la venta en piezas, como se muestran los datos en la Figura 127.

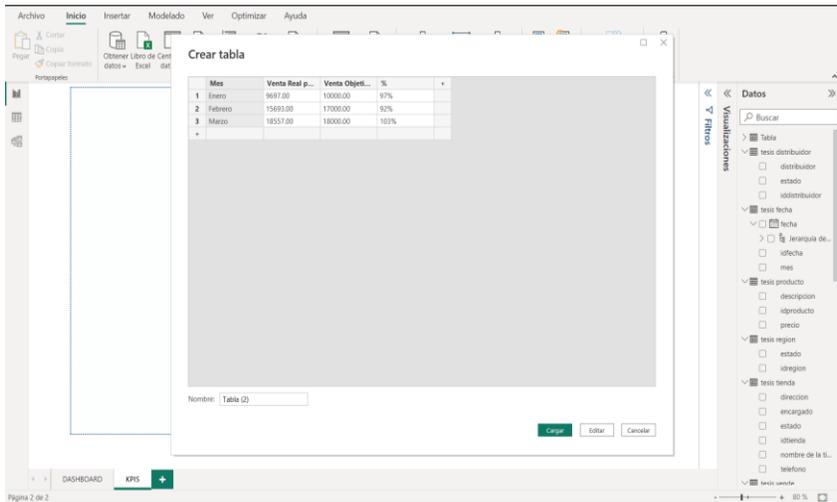


Figura 127. Tabla KPI piezas.

Se crearon las tablas KPI \$ y KPI piezas (Figura 128).

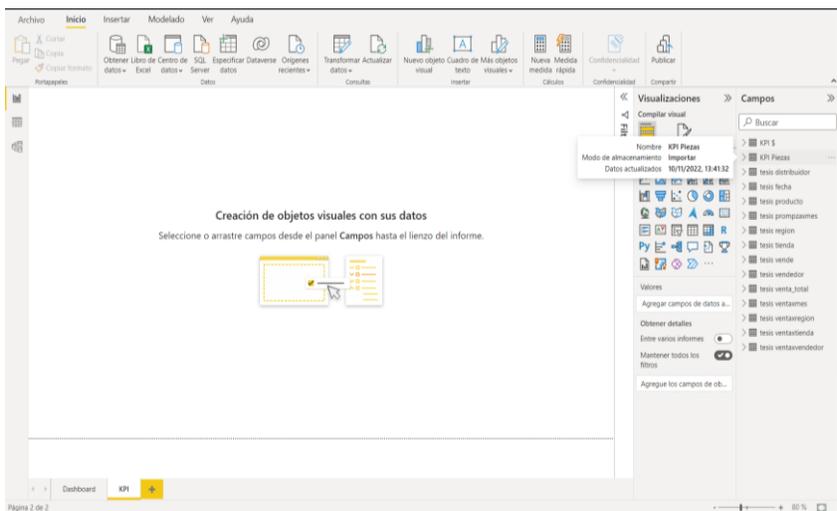


Figura 128. Creación de las tablas KPI.

Seleccionar el gráfico KPI con los campos: venta en pesos y venta en piezas (Figura 129), agregar una segmentación con el campo mes y dos medidores, estos pasos ya han sido descritos en las secciones pasadas.

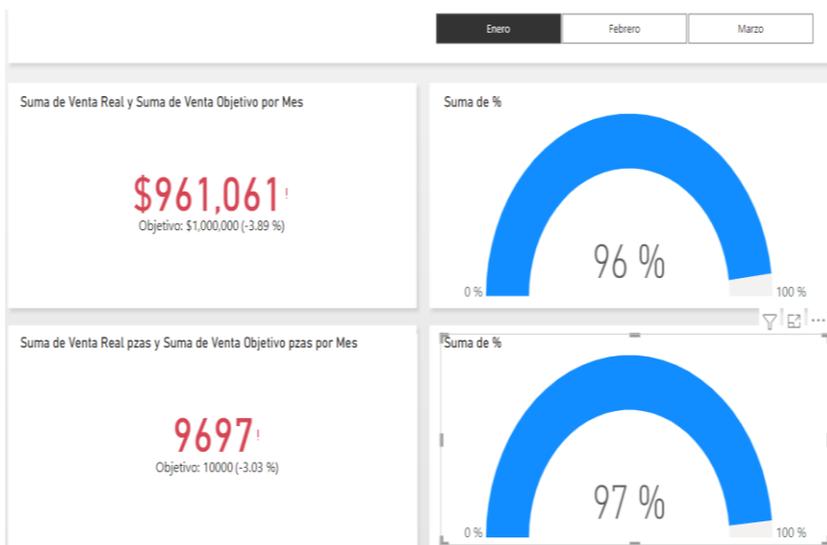


Figura 129. Selección de gráfico KPI con los campos venta en pesos y venta en piezas.

Con esta visualización podemos ver el alcance de la venta vs objetivo por mes del año 2023.

Se muestra la visualización de los KPI con el formato que se ha venido utilizando (Figura 130), este ayudará a identificar de mejor manera los alcances de las ventas.



Figura 130. Indicadores de rendimiento mes de enero.

Como se muestra en el indicador, en el mes de enero se tuvo un alcance en la venta en pesos del 96% (-\$38.9 mil pesos) por debajo del objetivo esto equivale al -3.03%.

En la venta de piezas se obtuvo el 97% (9697 pzas.) quedando por debajo del objetivo al -3.03(303 pzas.).

Esto indica que el objetivo es real y alcanzable, facilitando el control sobre la información proporcionada.

Se procede a la revisión del mes de febrero, obteniendo una venta de \$794,416 mil pesos quedando por debajo del

objetivo al 88% (-\$105,584 pesos), esto indica que el mes de febrero no tuvo impacto sobre las ventas.

En la venta de las piezas tuvo un alcance del 92% (15,693 pzas.) quedando por debajo del objetivo en un 8% (1307 pzas) (Figura 131).



Figura 131. Indicadores de rendimiento mes de febrero.

En el mes de marzo el alcance en venta pesos del 101%(\$1,216 MDP) teniendo un excedente del +\$16,516 pesos (1%), como se observa la venta en piezas tuvo un alcance del

103%(18,557pzás) rebasando su objetivo el 3%(557pzás) (Figura 132).



Figura 132. Indicadores de rendimiento mes de marzo.

### 3.11 Modelos de minería de datos.

Coefficiente de correlación (Figura 133).



Figura 133. Coeficiente de correlación.

Promedio móvil simple (Figura 134).

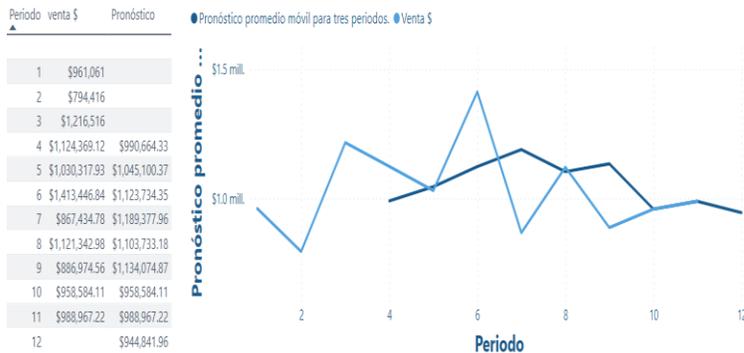


Figura 134. Pronóstico de la demanda.

Resultado del pronóstico:

La línea de pronóstico es más suave que la línea de la demanda, lo que prueba que el impacto de tomar un promedio mientras más periodos se utilicen para calcular el promedio móvil, el resultado será más suavizado. El pronóstico permanece rezagado en relación con la demanda real.

## RESULTADOS

Periodo: año y meses.

Con el dashboard que se generó se pudo visualizar los resultados de un trimestre de ventas del año 2023 en el cual la información obtenida es confiable gracias al modelo de base de datos relacional que se creó para este proyecto con la finalidad de tener información que cumpla con las dimensiones de calidad.

Para este trimestre se obtiene una venta total de \$2,971,33.00 MDP y en piezas fue de 43,947.00.

En el mes de enero tuvo una venta de \$961,061.00 y su objetivo fue de \$1,000,000 el alcance vs objetivo fue del 96%, nos faltó \$38,939 para poder cumplir con la meta.

Las ventas en piezas fueron de 9697 y su objetivo es de 1000, esto indica que se obtuvo un 97% de alcance quedando por debajo de 303 piezas, casi cumpliendo la meta para este mes.

En el mes de febrero hubo una venta de \$794,416 y su objetivo fue de \$900,000, mostrando que se estuvo a un 88% cerca de cumplir con la meta quedando por debajo \$105,584.00.

La venta en piezas fue de 15,693 y su objetivo es de 17,000 piezas, se tuvo un 92% de alcance vs su objetivo (1,307 piezas).

El mes de marzo tuvo un mejor comportamiento en ventas ya que se cumplió con el objetivo, este generó una venta de \$1,216,516 con un objetivo de \$1,200,000, indicando que sobrepaso su objetivo por 1% (\$16,516).

La venta en piezas fue de 18557 y su objetivo es de 18,000 rebasando el objetivo por más del 3%(557 piezas), fue el mes con mejor comportamiento.

La venta por estado se comportó de la siguiente manera:



Figura 135. Ventas por estado.

Esto nos indica que ciudad de México y monterrey son los estados que impactaron en la venta total del trimestre. Los vendedores con mejor venta son:

<b>VENDEDOR</b>	<b>VENTA TRIMESTRAL</b>	<b>PIEZAS VENDIDAS</b>
<b>David Rosas</b>	\$903,595	17176.
<b>Daniel Morales</b>	\$728,551	3585.
<b>Pablo Mayen</b>	\$275,062	4481.
<b>Alejandra Zaragoza</b>	\$274,471	4219.
<b>Diego Rivera</b>	\$241,190	3681.
<b>Gabriela Luna</b>	\$175,619	4558.
<b>Berenice Gallarda</b>	\$144,545	2283.
<b>Laura Pine</b>	\$105,648	1140.
<b>Alma Delia Ortega</b>	\$80,278	1935.
<b>Enrique Aguilar</b>	\$43,034	889

Figura 136. Venta por vendedor.

Al realizar el coeficiente de correlación para la variable ventas y tiempo indica que no hay relación entre variables al

igual arrojo el mismo resultado para la variable piezas con respecto al tiempo, y esto se debe a que tiene poca afluencia de puntos en las gráficas que se realizaron (Figura 132).

Para la regresión lineal se presenta una ligera tendencia al alza, pero los datos están muy dispersos y estos arroja un coeficiente de determinación muy bajo, indicando que la proyección no es confiable ya que el comportamiento de las ventas está muy disperso (Figura 133).

## ANÁLISIS DE RESULTADOS.

El primer análisis que se realizó en el dashboard fue el de la venta total en pesos y en piezas (Figura 137).

2,971,993.00	43,947.00
Total venta	Total Piezas

Figura 137. Venta total.

Ahora se quiere ver las ventas de cada mes y piezas (Figura 138).

<b>enero</b>	
\$961,061	9697
Total venta	Total piezas
<b>febrero</b>	
\$794,416	15693
Total venta	Total piezas
<b>marzo</b>	
\$1,216,516	18557
Total venta	Total piezas

Figura 138. Venta trimestral enero-marzo.

Con el gráfico de cascada se visualiza las ventas por mes y el total de estas (Figura 139), muestra que el mes con mayor venta es marzo y posteriormente enero, el mes con menor venta fue febrero.

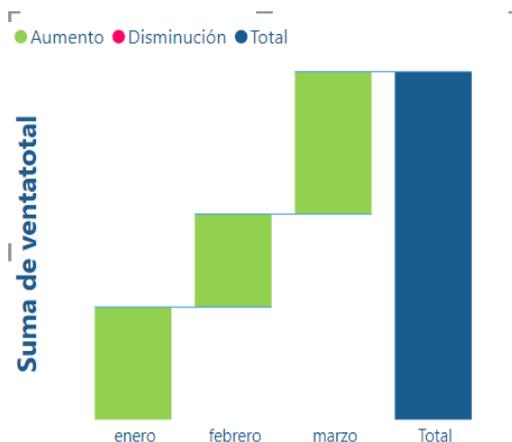


Figura 139. Gráfica de cascada venta en pesos.

En el siguiente análisis se visualiza la venta por región en los tres meses (Figura 140), este ayuda a percibir de mejor manera el comportamiento que se tuvo en ese trimestre, como se muestra la región con mayor venta es Ciudad de México y posteriormente le sigue Monterrey quedando Puebla rezagado.

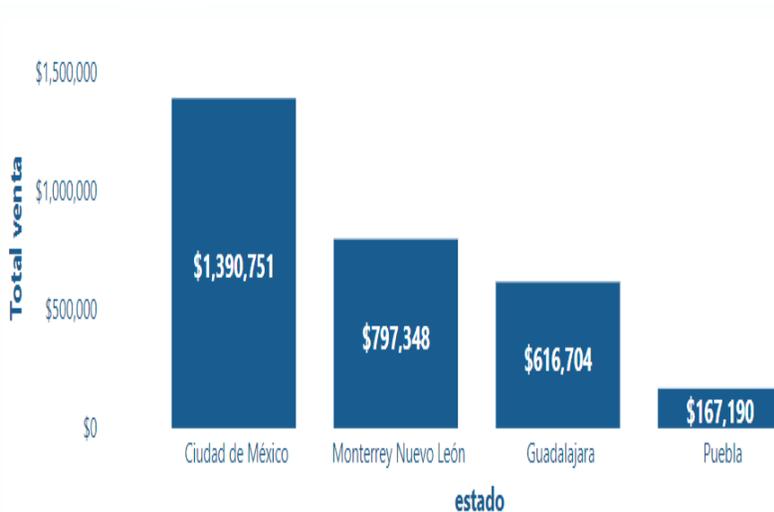


Figura 140. Gráfica de barras venta por estado.

El siguiente gráfico de columnas apiladas (venta \$) y de líneas (venta piezas) (Figura 141) se puede analizar las ventas de los vendedores tanto en pesos como en piezas a total trimestre.

Esto indica que el vendedor con mejor venta en ese trimestre fue David Rosas, seguido de Daniel Morales y Pablo Mayen como los primeros tres lugares.

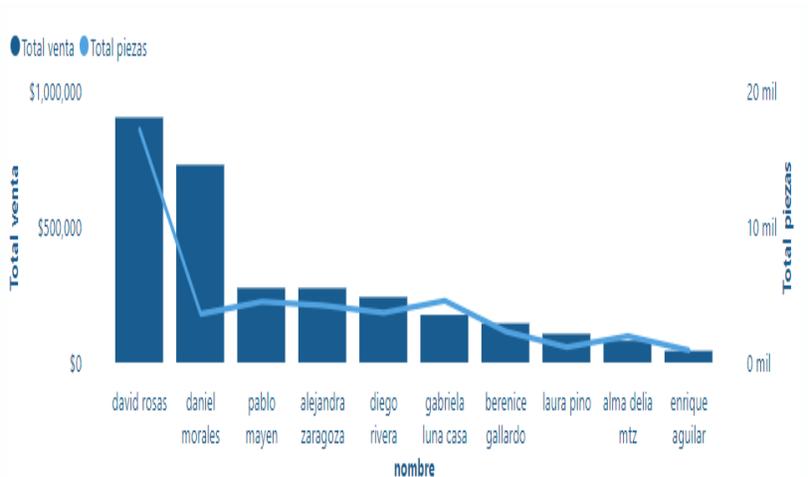


Figura 141. Gráfica de barras con línea de tendencia.

Se puede ver el alcance del objetivo de la venta del mes de enero tanto en piezas como en pesos (Figura 142) indica que no alcanzamos la meta en ninguno de los dos rubros, esto puede ser por el personal, ya que en este mes se contrataron a cuatro personas nuevas y por falta de capacitación y experiencia en esta área no pudieron cerrar sus ventas y esto generó que no se alcanzará el objetivo propuesto por el cliente.

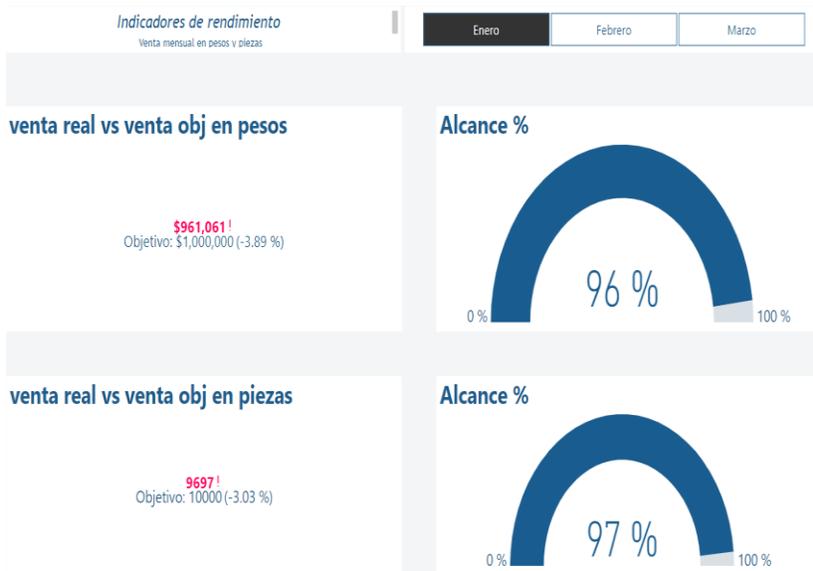


Figura 142. Indicadores de rendimiento en pesos.

El mes de febrero (Figura 143) tuvo un menor impacto en la venta y esto puede ser por falta de experiencia del nuevo personal, en este mes se dio dando capacitación a los vendedores esperando a que en el mes de marzo impacte sobre la venta.



Figura 143. Indicadores de rendimiento piezas.

Mes de marzo como se esperaba, las capacitaciones impactaron sobre la venta y en este mes se rebaso el objetivo de venta (Figura 144). Esto nos indica que dando capacitaciones acerca del producto genera un impacto positivo en los vendedores.

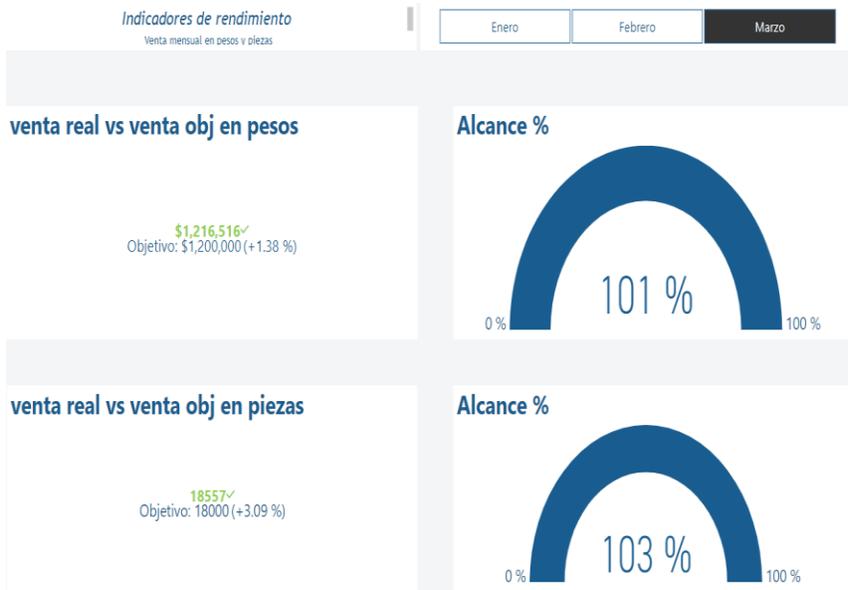


Figura 144. Indicadores de rendimiento pesos.

## DISCUSIÓN DE RESULTADOS.

Durante la elaboración del proyecto se identificaron varios aspectos, estos han generado conocimientos, llegando a lo siguiente:

- El modelo desarrollado en el proyecto ayudará a realizar bases de datos multidimensionales (datamart) para un área en específica, en este caso ventas, generando una optimización de la información.
- La interpretación de este conjunto de datos influye en la toma de decisiones acertadas a las necesidades de la empresa, traduciéndolos en ganancia o pérdidas.
- Mientras mayor sea la calidad de datos, estos serán más certeros, se incrementará la confianza en la compañía y su imagen, así también los costos disminuirán a largo plazo.
- La información que se obtiene se interpreta y se convierte en conocimiento y esta refleja el mundo real, el crecimiento de la compañía para afrontar los desafíos competitivos de la industria actual.
- El resultado de todo el análisis de las diversas variables y datos ayudarán a cumplir con las dimensiones de calidad, entre ellas la exactitud, consistencia, completos, válidos y la oportunidad.

- En la figura 145 se observa la optimización del proceso de elaboración de reporte y el tiempo se redujo de cinco horas con veintidós minutos a tres horas con seis minutos, esto confirma el propósito de la metodología KDD, retomando la información de mayor utilidad, a través de nuevos procesos de obtención y segregación de información.
- En la figura 146 se observa que las ventas del año 2023 aumentaron en comparación del 2022, esto se identifica con el análisis de datos realizado a través de la implementación de la metodología KDD.

**Estrategias que se van a implementar para este modelo:**

- ❖ Capacitar al equipo de ventas en habilidades de comunicación, negociación, manejo de objeciones y conocimiento del producto.
- ❖ Diseñar un proceso de ventas que guie a los vendedores desde el primer contacto con el cliente potencial hasta el cierre de ventas.

Proceso en excel			
Paso	Proceso	Descripción	Tiempo estimado (Hr/min)
1	Descargar	Obtener la información para la elaboración del dashboard.	2 Hr
2	Acomular	Reunir toda la información en un solo archivo.	1 Hr
3	Validar	Revisar minuciosamente la información obtenida.	30 Min
4	Limpieza	Homologar y eliminar información duplicada.	30 Min
5	Cargar	Integrar los datos a la plantilla de excel.	15 Min
6	Validar	Revisar los rangos de las tablas dinámicas y de los graficos.	50 Min
7	Actualizar	Actualizar las tablas dinámicas.	1 Min
8	Validar	Revisar que todas las gráficas jalen correctamente la información.	15 Min
9	Actualizar	Actualizar la pagina del dashboard en excel	1 Min
	Total		05:22

Proceso kdd			
Paso	Proceso	Descripción	Tiempo estimado (Hr/min)
1	Descargar	Obtener la información para la elaboración del dashboard.	2 Hr
2	Acomular	Reunir toda la información en un solo archivo.	1 Hr
3	Cargar	Subir la información a la base de datos para hacer la integración.	5 Min
4	Actualizar	Actualizar la pagina del dashboard que se genero en Power BI.	1 Min
	Total	Tres horas con seis minutos.	03:06

**Figura 145.** *Comparativo del proceso anterior vs la metodología KDD.*

Nota: la tabla comparativa muestra la optimización y eficiencia del proceso KDD proceso KDD en la elaboración de dashboard. Creación propia.

Año 2022				Año 2023			
Mes	Venta Real	Venta Objetivo	%	Mes	Venta Real	Venta Objetivo	%
Enero	\$ 833,490.00	\$ 950,000.00	88%	Enero	\$ 961,061.00	\$1,000,000.00	96%
Febrero	\$ 574,416.00	\$ 800,000.00	72%	Febrero	\$ 794,416.00	\$ 900,000.00	88%
Marzo	\$1,071,516.00	\$1,150,000.00	93%	Marzo	\$1,216,516.00	\$1,200,000.00	101%

**Figura 146.** *Comparativo de ventas vs año anterior.*

Nota: Los gráficos muestran el comparativo de ventas del año 2022 contra el año 2023. Creación propia.

## CONCLUSIONES.

- El diseño de la base de datos relacional que se realizó en este proyecto nos ayudó a tener una mejor calidad en los datos, esto con las restricciones que se tienen cuando se usa el modelo relacional con la finalidad de eliminar inconsistencias y redundancia en la información.
- Con el visualizador de datos se obtuvo información de manera clara y concisa a un público que no necesariamente tiene experiencia técnica en el campo, facilitando su interpretación y manejo para cumplir con los objetivos fijados.
- La minería de datos permite la identificación de tendencias y apatrones que brindan información valiosa para la implementación de decisiones estrategias y así lograr, como el

## BIBLIOGRAFÍA:

Anderson R., David y Sweeney J. D. (2012), Estadística para negocios y economía, México: Editorial Cengage Learning.

Base de datos Jerárquica. Historia de la bases de datos, <https://sites.google.com/site/sites/system/errors/WebspaceNotFound?path=/fsisbdd/home&pli=1>

Camacho M. (2017) Monosaurio y multiusuario. Recuperado el 08 de marzo de 2017.

Douglas A. L. (2006). Estadística aplicada a los negocios y la economía, México: Mc Graw Hill.

Elementos de una Base de Datos (2020) Dirección General de Cómputo y de tecnologías de Información y Comunicación, UNAM, 2020.

Esquemas de constelación (08-03-21-IBM) Esquemas de constelación - Documentación de IBM. <https://www.ibm.com/docs/es/ida/9.1.2?topic=schemas-starflake>

Esquema copo de nieve (2017) Esquemas de copo de nieve - Documentación de IBM. Recuperado el 08 de marzo de 2017.

<https://www.ibm.com/docs/es/ida/9.1.2?topic=schemas-snowflake>

Gamazo. A. G. Introducción a la Data Warehouse.

[https://openaccess.uoc.edu/bitstream/10609/136246/5/Disen%C2%BFo%20y%20construccio%C2%BFn%20de%20un%20almace%C2%BFn%20de%20datos\\_Mo%C2%BFdulo1\\_Introduccio%C2%BFn%20al%20Data%20Warehouse.pdf](https://openaccess.uoc.edu/bitstream/10609/136246/5/Disen%C2%BFo%20y%20construccio%C2%BFn%20de%20un%20almace%C2%BFn%20de%20datos_Mo%C2%BFdulo1_Introduccio%C2%BFn%20al%20Data%20Warehouse.pdf)

Hernández Orallo J. (2004). Introducción a la minería de datos, 2004. Pearson educación libgenli compress.

ISC. Hernández Cedano J. A. (2015) Modelo de Minería de Datos para identificación de patrones que influyen en el aprovechamiento académico. Recuperado en septiembre 2015, de <http://posgrado.lapaz.tecnm.mx/uploads/archivos/TesisHdzCedano.pdf>.

Levin & Rubin. Estadística Administración Economía, México: Prentice Hall, 7ma Edición.

María DB. Foundation, documentación (2009 – 2022) recuperado de MariaDB Foundation el 2009-2022. <https://mariadb.org/documentation/>

MARIA DB (2022). Recuperado el 18 de noviembre de 2022.  
<https://es.wikipedia.org/wiki/MariaDB>

Modelos de datos dimensionales lógicos (08-03-21-IBM).  
Modelos de datos dimensionales lógicos -  
Documentación de IBM.  
<https://www.ibm.com/docs/es/ida/9.1.2?topic=design-logical-dimensional-data-models>

Mora Rioja A. Base de datos Diseño y Gestión. Edit. Síntesis.  
2014.Pdf. pág. 2.  
<https://www.sintesis.com/data/indices/9788490770429.pdf>

Ramez, E. y Shamkant, B. N (2007). Fundamentos de  
sistemas de bases de datos, México: Pearson.

¿Qué es el KDD o Proceso de Descubrimiento de  
Conocimiento? (2019). DiagramasUML. Recuperado el  
04 de julio de 2019. <https://diagramasuml.com/que-es-el-kdd-o-proceso-de-descubrimiento-de-conocimiento/>

Silverschatz, K, S. (2006) Fundamentos de bases de datos,  
España: Mc Graw Hill.

Laudon, K. C. y Laudon, J. P. Sistemas de información  
gerencial, México: PEARSON EDUCACIÓN.

¿Qué es el data warehouse?. Blog de CEUPE  
<https://www.ceupe.com/blog/que-es-el-datawarehouse.html>

¿Qué es Power BI? (2022). ¿Qué es power BI? Recuperado el 10 de mayo de 2022. <https://powerbi.microsoft.com/es-mx/what-is-power-bi/Minería de datos>

¿Qué son la data mart? (2019). Tipos de Data Marts: Definición e Implementación | Astera. Recuperado el 21 de octubre de Astera Enabling Data-Driven Innovation. <https://www.astera.com/es/type/blog/types-of-data-marts/>

Regresión lineal. Documentos IBM. <https://www.ibm.com/mx-es/analytics/learn/linear-regression>

Tamayo M. & Moreno F. J.(2006). Comparing the MOLAP The ROLAP Storage Models [Versión electrónica]. <https://www.redalyc.org/pdf/643/64326317.pdf>

XAMPP (2022). Recuperado el 25 de noviembre de 2022. <https://es.wikipedia.org/wiki/XAMPP>