CONVOCATORIA | ESPECIALISTA EN GESTIÓN DE BASE DE DATOS

**Análisis y modelado de datos:**

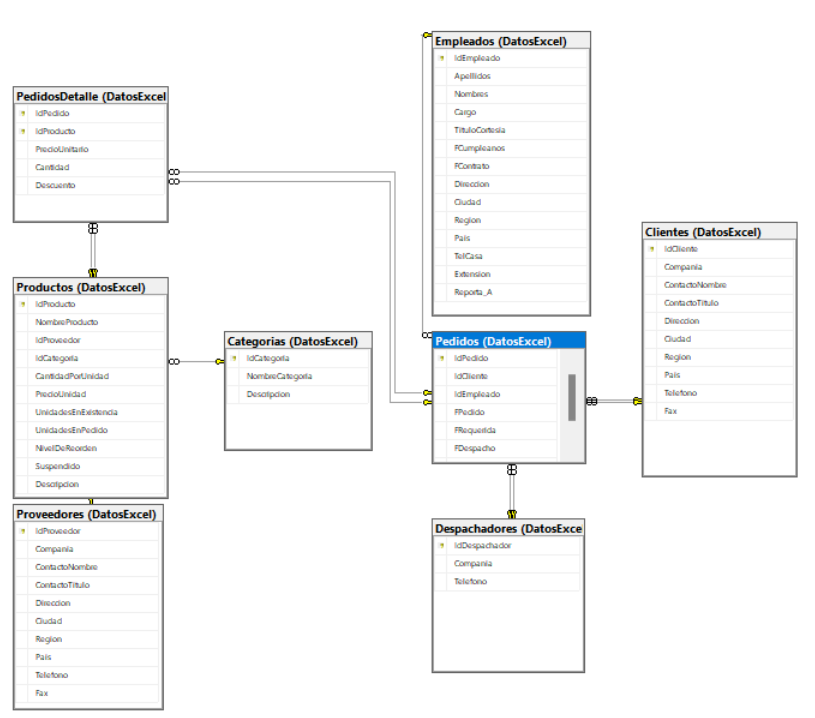
* ***Analice y estructure un modelo de datos relacional o no relacional que permita tener una visión multidimensional del proceso de pedidos.***

Para abordar el análisis y modelado del proceso de pedidos, se implementó un modelo de datos relacional que permite una visión multidimensional. Este modelo se estructura en dos esquemas dentro de SQL Server: DatosExcel y DW\_Pedidos.

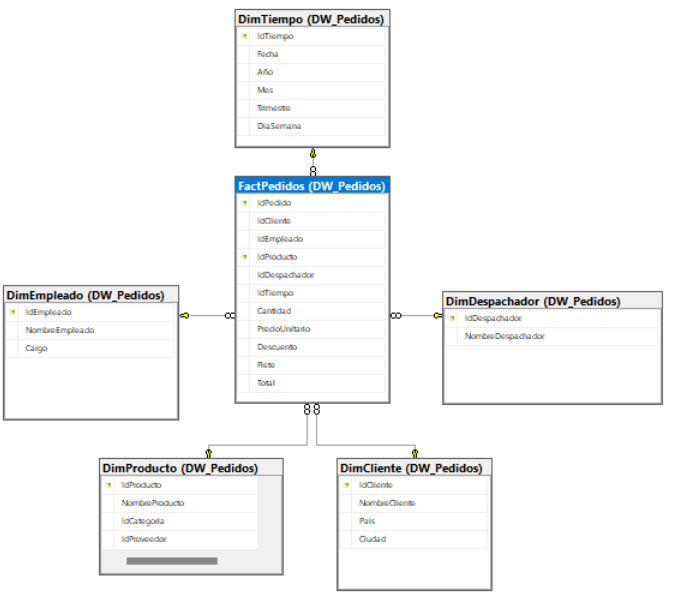
***Esquema DatosExcel:***

Este esquema contiene las tablas transaccionales directamente importadas del archivo Excel (Data.xlsx). Las tablas incluyen Pedidos, Clientes, Empleados, Productos, Despachadores, entre otras. Cada tabla está debidamente normalizada con sus respectivas claves primarias y foráneas, asegurando la integridad referencial.

Se diseñaron procedimientos almacenados para importar datos desde Excel a estas tablas y para limpiar los datos antes de una nueva importación.



***Esquema DW\_Pedidos:***

Este esquema está diseñado para soportar un ***Data Warehouse,*** facilitando análisis multidimensionales. Se crearon tablas de dimensiones (***DimCliente***, ***DimEmpleado***, ***DimProducto***, ***DimDespachador***, ***DimTiempo***) que organizan la información en categorías relevantes.

La tabla de hechos ***FactPedidos*** centraliza los datos de las transacciones, vinculando cada pedido con sus dimensiones asociadas, como cliente, empleado, producto, despachador, y tiempo. Además, se calculan métricas clave como cantidad, precio unitario, descuento, y flete.

Un procedimiento almacenado específico, CrearDatos\_DW, procesa los datos desde el esquema DatosExcel y los carga en el esquema DW\_Pedidos, estructurando la información para análisis posterior.

Visión multidimensional: Con esta estructura, el modelo de datos permite realizar análisis multidimensionales, como el análisis de ventas por cliente, producto, o empleado, y el seguimiento de la eficiencia operativa en términos de despachos y tiempos de entrega. También facilita la identificación de tendencias y patrones a lo largo del tiempo.

***Proceso ETL:*** El modelo incluye un proceso ETL que importa los datos desde un archivo Excel hacia las tablas transaccionales y luego los transforma (**se realizaron varias transformaciones clave para garantizar la correcta importación y manipulación de los datos. Estas incluyen la conversión de cadenas de texto en fechas para campos como FPedido, FRequerida y FDespacho**) y carga en el Data Warehouse. Este proceso se automatiza mediante procedimientos almacenados que aseguran la limpieza, la integridad y la precisión de los datos.

***Dimensiones de Tiempo:*** Esta vista (***Vw\_PedidosConTiempo***) la definí para facilitar la integración y consulta de los datos de pedidos junto con las dimensiones de tiempo.

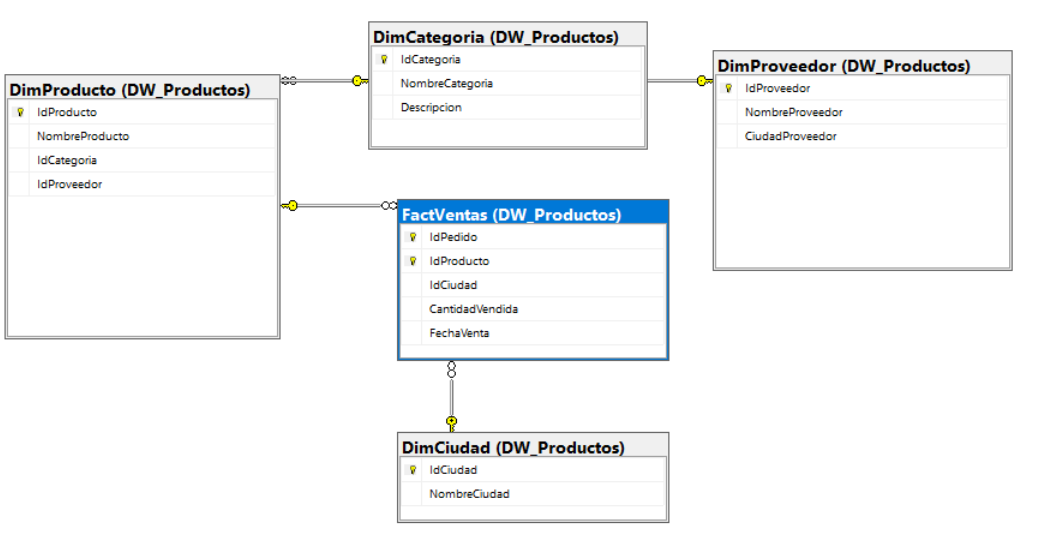
* ***Defina mediante un esquema los atributos destinados a medir la rotación del producto más vendido por ciudad, de los pedidos cuyo proveedor sea de Melbourne.***

Para abordar la medición de la rotación del producto más vendido por ciudad, enfocado en los pedidos cuyos proveedores sean de Melbourne, se ha definido un esquema de datos dentro de un nuevo esquema de base de datos llamado ***DW\_Productos***. Este esquema incluye tanto tablas de dimensiones como una tabla de hechos, diseñadas específicamente para soportar este análisis. He creado una vista llamada ***DW\_Productos].[Vw\_RotacionProductoPorCiudad]*** la cual tiene todas las rotaciones de productos y con esta con solo filtrar por ciudad podemos llegar a diseñar el informe solicitado.

* ***Complete y entregue la visualización de la información solicitada y del esquema diseñado.***

Se anexa PBIX en repositorio GIT





* ***Detalle las dimensiones y tabla de hechos que intervienen en el módulo de pedidos.***

Diagrama

Descripción generada automáticamente

Esquema General del Módulo de Pedidos

FactPedidos: Centraliza las transacciones relacionadas con los pedidos realizados.

DimCliente: Describe los clientes que realizan los pedidos.

DimEmpleado: Describe los empleados que gestionan los pedidos.

DimDespachador: Describe las compañías encargadas del despacho de los pedidos.

DimProducto: Describe los productos incluidos en los pedidos.

DimTiempo: Proporciona detalles temporales asociados a los pedidos.

* ***Identifique y justifique el modelo que más se ajuste (copo de nieve o estrella).***

Dado que he utilizado relaciones entre las tablas de hechos y dimensiones, y para mantener la integridad y escalabilidad del sistema, el modelo de copo de nieve es el que mejor se ajusta a este módulo de pedidos. Este modelo permite mantener una estructura de datos más normalizada y organizada, lo que es beneficioso para la escalabilidad y el mantenimiento a largo plazo, a pesar de que pueda incrementar la complejidad de las consultas.

* ***Identifique las posibles vulnerabilidades en la arquitectura de datos propuesta.***

Podría identificar muchas oportunidades de mejora de la arquitectura propuesta, toda vez que su alcance dependerá del ambiente en el que se esta trabajando y de la complejidad de los datos. Entre las vulnerabilidades potencias que identifico, se encuentra el inyección SQL, exposición de datos sensibles y por ultimo integridad de los datos (toda vez que la fuente de datos es un archivo de Excel que puede ser modificado)

* ***Describa que medidas de seguridad usaría para protección contra inyecciones SQL, políticas de acceso, auditorías, y asegura su cumplimiento.***

Para proteger la arquitectura de datos contra inyecciones SQL, implementar políticas de acceso seguras, realizar auditorías efectivas y asegurar el cumplimiento de estas medidas, se deben adoptar las siguientes estrategias y tecnologías:

**Protección Contra Inyecciones SQL**

a. Uso de Consultas Parametrizadas, Las consultas parametrizadas aseguran que los datos de entrada se traten como parámetros en lugar de código SQL, lo que previene la inyección de comandos maliciosos.

b. Uso de ORM (Object-Relational Mapping), Un ORM, como Entity Framework en .NET, abstrae las consultas SQL, lo que reduce la posibilidad de inyección al generar automáticamente consultas SQL seguras.

c. Validación y Saneamiento de Entradas

**Políticas de Acceso**

a. Control de Acceso Basado en Roles, Implementar una política de seguridad donde los usuarios se asignan a roles con permisos específicos. Cada rol tiene acceso solo a las operaciones y datos necesarios para sus funciones.

Políticas de Contraseñas Fuertes, Implementar políticas que requieran contraseñas seguras para acceder a la base de datos.

Autenticación Multifactor (MFA), Requiere múltiples formas de verificación antes de permitir el acceso a la base de datos.

**Auditorías**

Implementación de Auditoría en la Base de Datos, Configurar auditorías que registren quién accede, modifica, o elimina datos críticos en la base de datos.

Revisiones Periódicas de Auditoría, Las auditorías solo son útiles si se revisan regularmente para detectar comportamientos anómalos.

**Aseguramiento del Cumplimiento**

Documentación de Políticas y Procedimientos, Mantener una documentación clara de todas las políticas de seguridad, procedimientos de auditoría, y controles de acceso implementados.

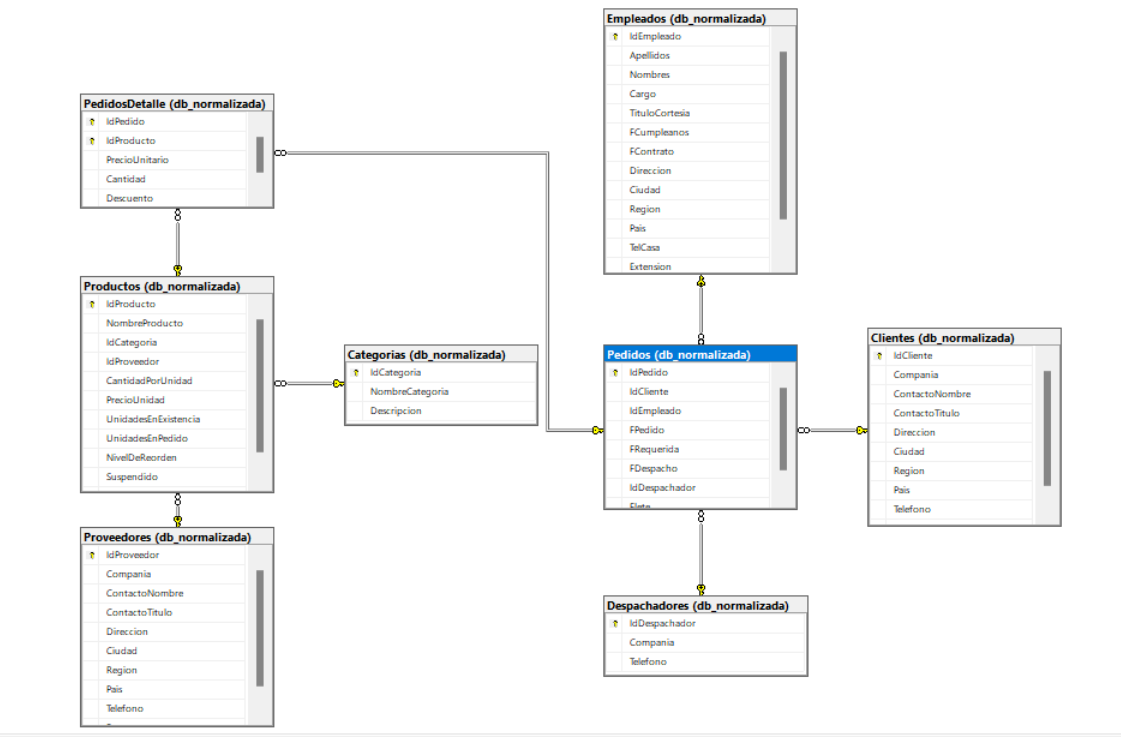
Capacitación y Concienciación, Capacitar a los usuarios y administradores sobre las mejores prácticas de seguridad y políticas específicas de la organización.

**Diseño de Modelo Relacional**

* ***Proponga el diseño de un modelo de datos relacional normalizado para la solución del punto anterior.***

Mi propuesta a la implementación de un modelo de tablas normalizado basado en las tablas que se importaron del archivo Excel, partiría de los principios de normalización para evitar redundancia de datos, asegurar la integridad referencial y facilitar el mantenimiento de la base de datos.

Primero identificar las tablas y relaciones y por ultimo normalizar el modelo en su tercera forma normal (3NF), lo que implica que todas las dependencias funcionales no triviales deben ser dependencias de clave.



* ***Plantee la solución que se debería implementar para la migración a un modelo de datos no relacional.***

Migrar a un modelo de datos no relacional implica un cambio significativo en cómo se estructuran y acceden los datos. Para esto se requiere una planificación cuidadosa para asegurar que las estructuras de datos y las consultas se adapten a las características y beneficios de las bases de datos no relacionales, como la flexibilidad, la escalabilidad horizontal y la capacidad de manejar datos no estructurados o semiestructurados.

Los pasos que yo implementaría en su orden serian:

Evaluar y Seleccionar la Base de Datos No Relacional a implementar.

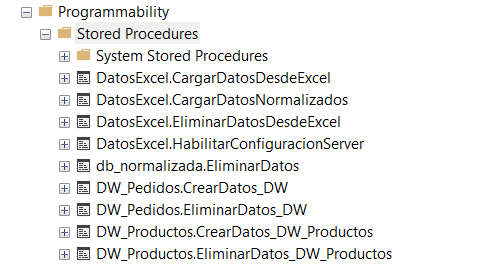
Seleccionar el modelo (Documentos – Columnas – Clave/Valor – Grafos)

Definir si se trabajara con documentos anidados (JSON son arrays que contengan subconsultas) o trabajar con tablas de columnas.

Definir los ETL de migración y desarrollar scripts que transformaran el sistema en un modelo NoSql

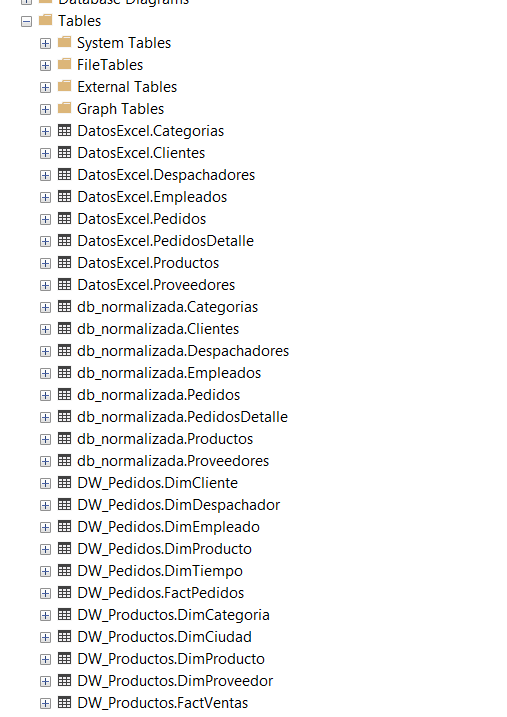
* ***ETL y Transformaciones:***
* ***Realice los programas de carga necesarios mediante una ETL que permitan leer las tablas de los sistemas transaccionales (normalizadas).***

Para esta actividad realice unos store proceder, que con ayuda de librerías de SQL Sever que se conectan a los archivos físicos, en este caso, Data.xlsx y lee hoja por hoja el libero y los convierte en tablas y de estas se toma la información base para llenar o poblar los distintas DW que se definieron para el ejercicio.



Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

* ***Realice las transformaciones necesarias para que puedan ser cargadas en las tablas dimensiones con su detalle incluido, los cuales se integrarán en un Data Warehouse centralizado.***

Para este caso específico se dio l información bastante limpia y clara excepto algunos datos nulos a los cuales se le dio manejo con el ETL, el store procedure que se encarga de esta tarea primero borra todos los datos que tienen las tablas de Excel asignadas (apropiado para al ejemplo del ejercicio, dado que no tiene muchos datos) luego constriye cada una de las talas dimensionales y las de movimiento.

* ***Indique qué tipo de transformaciones realizó en el proceso.***

Específicamente, solo transformé los valores de tipos de datos nchar a date que contenía la tabla de pedidos de Excel y en la tabla de dimensiones de tiempo de pedidos.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

**Visualización**

**Presente una visualización mediante un dashboard en Power BI con el top de los 3 vendedores más destacados de los últimos 2 meses de despacho hacia Germany.**

**Caso de negocio:**

**Informe técnico: Solución Estratégica para el Gobierno de Datos**

**Descripción del problema:** La entidad ofrece servicios en línea a nivel nacional, permitiendo a los clientes crear cuentas y realizar compras. Actualmente, los procesos no están integrados, no existe un flujo de trabajo definido entre las áreas, y no hay una gobernanza de datos adecuada.

La situación actual es la siguiente:

•***Procesos Desintegrados:*** Diferentes áreas manejan sus propios procesos, resultando en datos aislados.

•***Falta de Gobernanza de Datos:*** No hay una estructura ni diseño de procesos claros para la gestión y almacenamiento de datos.

•***Necesidad de Centralización:*** Las directivas requieren centralizar procesos e información para mejorar la eficiencia y la toma de decisiones.

•***Gestión de información:*** El área TI tiene bajo su control un inventario digital de software y bases de datos; sin embargo, cada dependencia maneja sus propios procesos que encadenan en datos aislados al inventario anterior.

***Objetivo del Informe:***

Proponer una solución estratégica para la gestión y gobernanza de datos que centralice los procesos y la información resultante de las diferentes áreas.

***Resultado esperado:***

Diseñe y presente una solución estratégica mediante un informe técnico para un gobierno de datos que dé solución a la problemática planteada.

***Tener en cuenta:***

•Arquitectura de datos (Modelo de datos relacional)

•Herramientas y tecnologías (ETL y visualización)

•Seguridad y cumplimiento (Políticas de acceso, auditorias y monitoreo)

**Informe Técnico: Solución Estratégica para la Gestión y Gobernanza de Datos.**

**1. Introducción**

La entidad enfrenta desafíos críticos en la gestión de sus datos debido a la desintegración de procesos, la falta de una gobernanza de datos efectiva, y la necesidad urgente de centralizar la información para mejorar la eficiencia operativa y la toma de decisiones. Este informe presenta una solución estratégica integral para abordar estos desafíos, enfocándose en la creación de una arquitectura de datos centralizada, la implementación de herramientas y tecnologías avanzadas, y el establecimiento de políticas robustas de seguridad y cumplimiento.

**2. Objetivo del Informe**

El objetivo de este informe es proponer una solución estratégica integral para la gestión y gobernanza de datos, diseñada para centralizar los procesos y la información resultante de las diferentes áreas de la entidad. La solución propuesta se basa en tres pilares fundamentales: Arquitectura de Datos (Modelo de Datos Relacional), Herramientas y Tecnologías (ETL y Visualización), y Seguridad y Cumplimiento (Políticas de Acceso, Auditorías y Monitoreo).

**3. Arquitectura de Datos**

***3.1. Modelo de Datos Relacional Centralizado***

La solución propone la creación de un Data Warehouse (DWH) centralizado que unifique los datos actualmente dispersos en diferentes áreas de la entidad. Este modelo relacional permitirá:

*Unificación y Centralización de Datos:* Integrar toda la información en una única fuente de verdad, eliminando silos y garantizando la consistencia y calidad de los datos.

*Estandarización de Flujos de Trabajo:* Definir y automatizar flujos de trabajo entre las áreas, asegurando que los datos fluyan de manera controlada y eficiente a través de la organización.

*Diseño de Esquema Relacional:* Implementar un esquema en estrella o copo de nieve según la complejidad, que incluye dimensiones como Clientes, Productos, Áreas de la entidad, Tiempos, y Canales de Venta, y tablas de hechos como Transacciones de Ventas.

***3.2. Beneficios***

*Centralización y Consistencia:* Mejora la visibilidad y accesibilidad de los datos para todas las áreas de la entidad.

Optimización de Recursos: Reduce la duplicación de esfuerzos y costos asociados con la gestión de datos dispersos.

**4. Herramientas y Tecnologías**

***4.1. Procesos ETL (Extract, Transform, Load)***

La implementación de herramientas de ETL como Microsoft SQL Server Integration Services (SSIS) o Apache NiFi es fundamental para:

*Integración y Limpieza de Datos:* Extraer datos de diferentes sistemas, transformarlos para cumplir con las reglas de negocio y cargarlos en el Data Warehouse.

*Automatización:* Reducir la intervención manual y minimizar errores a través de la automatización de flujos de datos.

***4.2. Visualización de Datos***

El uso de herramientas como Power BI o Tableau permitirá:

*Dashboards y Reportes Interactivos:* Crear visualizaciones dinámicas que faciliten el análisis de datos y la toma de decisiones en tiempo real.

*Visibilidad en Tiempo Real:* Monitorear operaciones y KPIs críticos, proporcionando transparencia a todos los niveles de la organización.

***4.3. Beneficios***

*Eficiencia Operativa:* Mejora la toma de decisiones mediante datos precisos y actualizados.

Escalabilidad: La solución puede crecer con la organización, manejando grandes volúmenes de datos de manera eficiente.

**5. Seguridad y Cumplimiento**

***5.1. Políticas de Acceso y Seguridad***

Se establecerán políticas estrictas de acceso y seguridad basadas en:

*Roles y Perfiles de Acceso:* Implementar una política de mínimo privilegio, asegurando que solo el personal autorizado tenga acceso a los datos necesarios.

*Autenticación Multifactor (MFA):* Proteger el acceso a datos sensibles con MFA, reduciendo el riesgo de brechas de seguridad.

***5.2. Auditorías y Monitoreo Continuo***

*Auditorías Regulares:* Monitorear y registrar todas las actividades de acceso a los datos para detectar comportamientos sospechosos.

*Herramientas de Monitoreo:* Utilizar Azure Monitor o Splunk para analizar eventos en tiempo real y garantizar la integridad de los datos.

***5.3. Beneficios***

Cumplimiento Normativo: Asegura que la entidad cumpla con las regulaciones de protección de datos, minimizando riesgos legales.

*Proactividad en Seguridad:* Permite anticipar y mitigar riesgos antes de que se conviertan en problemas.

**6. Implementación y Roadmap**

***6.1. Fases de Implementación***

Evaluación y Planificación: Definir metas claras y establecer un equipo de gobernanza de datos.

*Diseño del Modelo de Datos:* Crear y validar el modelo de datos relacional.

*Desarrollo de ETL y Data Warehouse:* Implementar los procesos ETL y construir el Data Warehouse.

Implementación de Herramientas de Visualización: Configurar y personalizar Power BI o Tableau.

*Capacitación y Adopción:* Capacitar al personal en el uso de las nuevas herramientas.

*Monitoreo y Mejora Continua:* Establecer un ciclo de monitoreo y optimización.

***6.2. Gobernanza de Datos***

*Comité de Gobernanza de Datos:* Crear un comité responsable de supervisar la gestión de datos y garantizar el cumplimiento de las políticas establecidas.

*Ciclo de Vida de los Datos:* Definir políticas claras para la gestión del ciclo de vida de los datos, desde su creación hasta su eliminación.

**7. Conclusión**

La implementación de esta solución estratégica permitirá a la entidad superar sus desafíos actuales relacionados con la gestión de datos. Con una arquitectura de datos centralizada, herramientas avanzadas de ETL y visualización, y un enfoque sólido en seguridad y cumplimiento, la entidad estará mejor posicionada para operar de manera eficiente, segura y con un enfoque basado en datos para la toma de decisiones. Esta solución no solo aborda las necesidades actuales, sino que también sienta las bases para un crecimiento sostenible y exitoso en el futuro.