

UNIVERSIDAD DON BOSCO

Dirección de Educación a Distancia



TRABAJO DE GRADUACIÓN PARA OBTAR AL GRADO DE

Maestro en Arquitectura de Software

PROYECTO DE GRADUACIÓN:

Implementación de Data Warehouse en Dirección General de Migración y Extranjería en las gerencias de Control Migratorio y Restricciones

Autores:

Ing. Francisco José Huevo Alas

Ing. Héctor Augusto Hernández Argueta

Lic. Ernesto Enrique Menjívar Colorado

Asesor:

Mg. Herson Miguel Serrano Chacón

Antiguo Cuscatlán, La Libertad, El Salvador, Centroamérica.

11 de Julio de 2023

Contenido

1. Introducción.....	6
2. Identificación de las necesidades.....	8
5. Contexto empresarial.....	11
6. Identificación de los orígenes de datos	12
6.1. Sistema Integrado.....	12
6.2. Sistema de Pasaportes.....	12
7. Justificación de la solución.....	15
7.1. La estructura del Data Warehouse.....	21
7.2. Tipos de modelos en el esquema multidimensional	22
7.2.1. ¿Qué es el esquema estrella?	23
7.2.2. ¿Qué es el esquema copo de nieve?.....	24
7.2.3. ¿Cómo funciona el esquema estrella?.....	25
7.2.4. ¿Cómo funciona el esquema copo de nieve?	26
7.2.5. Ventajas y limitaciones del esquema tipo estrella	28
7.2.6. Ventajas y limitaciones del esquema copo de nieve	29
8. Construcción del nuevo diagrama de datos	31
9. Indicadores de rendimiento.....	33
9.1. Ventajas de utilizar KPI en el control migratorio	34
9.2. Identificación de los KPI relevantes	35
10. Definición de los procesos de ETL.....	38
11. Sistemas de información orientados a toma de decisiones.	42
11.1. Sistema de soporte a la decisión	43
11.2. Tipos de Sistemas de Soporte a la Decisión.....	46
11.3. Diferencias con otras herramientas de Business Intelligence.....	47
12. Análisis y definición de las herramientas a utilizar	48
12.1. Herramientas ETL.....	48

12.1.1.	Microsoft SSIS.....	50
12.1.2.	Pentaho Kettle	51
12.1.3.	Informática Powercenter.....	54
12.1.4.	Oracle Data Integrator	55
12.1.5.	Talend	56
12.2.	Herramientas de análisis de datos.....	58
12.2.1.	Power BI.....	59
12.2.3.	QLik Sense.....	63
12.3.	Comparación de Herramientas	65
12.3.1.	Herramienta de migración de datos	65
12.3.2.	Herramienta para visualización de datos	66
13.	Análisis y definición de la infraestructura de la solución.....	68
14.	Plan de implementación.....	70
14.1.	Fases de Implementación.....	70
14.1.1.	Fase de Planificación y Preparación	70
14.1.2.	Fase de Diseño y Modelado:	70
14.1.3.	Fase de Extracción y Transformación de Datos.....	70
14.1.4.	Fase de Carga de Datos y Configuración de la Capa de Almacenamiento: 71	
14.1.5.	Fase de Implementación de la Capa de Presentación y Visualización.....	71
14.1.6.	Fase de Ajustes Finales y Lanzamiento:.....	71
14.2.	Plan de ejecución	71
14.3.	Presupuesto	72
14.4.	Roles y responsables	75
15.	Conclusiones	77
16.	Bibliografía.....	79

Índice de Ilustraciones

Ilustración 1. Base de datos actual de la Dirección General de Migración y Extranjería. Fuente: Gerencia de informática y desarrollo tecnológico de la DGME	14
Ilustración 2. Ejemplo de la orientación temática de los datos. Fuente: Inmon (2005).....	17
Ilustración 3. La cuestión de la integración. Fuente: Inmon (2005).....	18
Ilustración 4. La cuestión de la no volatilidad. Fuente: Inmon (2005)	19
Ilustración 5. La cuestión de la variación temporal. Fuente: Elaboración propia a partir de Inmon (2005)	20
Ilustración 6. La estructura del Data Warehouse. Fuente: Inmon (2005)	21
Ilustración 7. Esquema tipo estrella. Fuente: www.sqlybi.com	23
Ilustración 8. Esquema copo de nieve. Fuente: www.sqlybi.com	25
Ilustración 9. Diagrama del Data Warehouse. Fuente: Elaboración propia.....	31
Ilustración 10. Componentes de un sistema de soporte a la toma de decisiones fuente (Bonilla Botia & Briceño Díaz, 2006)	45
Ilustración 11. Diagrama de la infraestructura de la solución. Fuente: Elaboración propia.	68
Ilustración 12. Plan de ejecución del proyecto. Fuente: Elaboración propia	71

Índice de Tablas

Tabla 1. Comparativa de ETLs en el mercado actual. Fuente: Elaboración propia.....	65
Tabla 2. Comparativa de herramientas para la visualización y análisis de datos. Fuente: Elaboración propia.....	67
Tabla 3. Presupuesto del proyecto.....	74
Tabla 4 Roles para la ejecución del proyecto.....	76

1. Introducción

En las últimas décadas, la migración ha sido un fenómeno de gran relevancia en el contexto socioeconómico global. El Salvador, como muchos otros países, ha experimentado flujos significativos de migrantes, tanto dentro como fuera de sus fronteras. La gestión eficiente y efectiva de los datos relacionados con la migración se ha convertido en un desafío para las instituciones gubernamentales encargadas de formular políticas y programas que aborden esta problemática de manera integral.

El acceso a información precisa y actualizada es fundamental para tomar decisiones informadas y diseñar estrategias efectivas en el ámbito de la migración. En este contexto, la implementación de un Data Warehouse (almacén de datos) surge como una solución tecnológica prometedora para consolidar y gestionar la gran cantidad de datos dispersos que se generan en diferentes entidades gubernamentales relacionadas con la migración.

El objetivo de este documento de tesis es analizar la implementación de un Data Warehouse en la Dirección General de Migración y Extranjería (DGME) entidad de gobierno dedicada a la gestión de datos sobre movimientos migratorios, pasaportes, registro de extranjeros que residen legalmente en El Salvador. Se busca examinar cómo esta solución tecnológica puede mejorar la recolección, almacenamiento, integración y análisis de la información migratoria, brindando una visión más completa y precisa de los flujos migratorios y sus implicaciones.

Para lograr este objetivo, se llevará a cabo una revisión exhaustiva de la literatura existente sobre la implementación de Data Warehouses en el ámbito gubernamental y su aplicación en el campo de la migración. Además, se realizará un estudio de caso en DGME, con el fin de evaluar los beneficios y desafíos asociados con la implementación de esta tecnología.

Se espera que los resultados de esta investigación proporcionen una base sólida para comprender el potencial impacto de un Data Warehouse en la gestión de datos que maneja la institución, y contribuyan a la toma de decisiones informadas en políticas y programas relacionados con la migración.

La implementación de un Data Warehouse DGME representa una oportunidad para mejorar la gestión de datos sobre migración y promover un enfoque más integral y basado en evidencias en la formulación de políticas migratorias.

A través de este estudio, se espera contribuir al fortalecimiento de la capacidad institucional y al desarrollo de estrategias efectivas para abordar los desafíos asociados con la migración en el país.

En esta investigación, se analizarán los beneficios potenciales de la implementación de un Data Warehouse, como la mejora en la eficiencia en la gestión de solicitudes y trámites migratorios, la detección temprana de casos de migración ilegal, la generación de informes más ágiles y la toma de decisiones más informadas. Además, se explorarán los desafíos y consideraciones clave que deben abordarse en la implementación exitosa de un Data Warehouse en la DGME.

En resumen, la aplicación de un Data Warehouse en la DGME representa una oportunidad para mejorar la gestión y el control migratorio en El Salvador. La centralización de datos, el análisis avanzado y la generación de informes eficientes permitirán una toma de decisiones más informada y una mayor eficacia en los procesos migratorios. A lo largo de este estudio, se explorarán los beneficios y desafíos asociados con la implementación de un Data Warehouse en esta institución, sentando las bases para una gestión migratoria más sólida y eficiente.

2. Identificación de las necesidades

La Dirección General de Migración y Extranjería (DGME) es el ente encargado de controlar la entrada y salida de personas al territorio salvadoreño, la ubicación de El Salvador en Centro América hace vital tener herramientas de gestión de datos con respuestas rápidas a los diferentes escenarios que puedan surgir. Dentro de la institución existen diferentes gerencias o áreas donde cada una cumplen una función esencial para el correcto funcionamiento del andamiaje gubernamental. Del control migratorio, puede surgir muchas necesidades de solicitud de información vital para el país, por ejemplo, cuantas personas salen y no regresan al país, cuantas personas entran ilegalmente al país, puntos de inflación de movimientos migratorios en el año, nacionalidades predominantes que entran al país etc. La información crece a diario y por este motivo la institución necesita de una solución que reduzca los tiempos de respuesta y el esfuerzo en obtención de información; manteniendo siempre la integridad y calidad de los datos.

Dentro de DGME existe también otra división importante llamada Gerencia de Atención al Migrante, cuyas necesidades también deberán ser contempladas dentro del proyecto, ya que son los encargados de recibir las donaciones de otros países, además de gestionar el tema de refugiados o atención a retornados; para esta gerencia es vital llevar un registro de cuantas personas son retornadas al país y que nacionalidad es la predominante, ya que con esta información se puede proceder a crear planes de contingencia.

El proceso de obtención de datos actualmente inicia con una solicitud de información por parte de la Gerencia de Control migratorio, realizada por medio de su gerente o la persona que el designe. Luego, la solicitud llega al jefe de la Unidad de informática y Desarrollo Tecnológico (UIDT), este último por medio del encargado de base de datos de la institución, realiza una serie de consultas a la base de datos. Al obtener la información, esta es llevada a un documento de Excel en donde se ordena y se hace la entrega por medio de correo electrónico a la gerencia solicitante. El problema central, radica en que este proceso es tardado, empezando por la gestión que debe hacerse para poder solicitar la información entre gerencias y, además, por los tiempos que tardan los técnicos en obtener la información de la base de datos en ciertos casos, por ejemplo, cuando se solicita datos de hace tres años a la fecha actual de entradas y salidas de personas al territorio nacional, teniendo en cuenta que cada año son miles de personas que cruzan la frontera.

Estos procesos pueden tardar hasta días, con la desventaja que depende bastante de la habilidad del encargado de la base de datos en turno.

La UIDT cuenta con reportes ya establecidos de información recurrente, la mayoría de ellos los parámetros son fechas, mientras más amplio sea el rango entre una fecha y otra, mayor será el tiempo en que el reporte se tarde en ejecutar la consulta, ya que básicamente son consultas a la base de datos, esto generando una carga extra al servidor de producción lo cual hace que sea lenta algunas transacciones.

Para poner un poco en contexto, actualmente la DGME cuenta con:

- Sucursales
 1. Centro de gobierno San Salvador
 2. Sucursal Masferrer
 3. Sucursal Cascadas
 4. Sucursal plaza mundo (Soyapango)
 5. Sucursal Pasatiempo
 6. Sucursal Santa Ana
 7. Sucursal San Miguel
 8. Sucursal Usulután
- Fronteras terrestres con Guatemala
 1. Chinamas
 2. La Hachadura
 3. Anguiatú
 4. San Cristóbal
- Fronteras terrestres con Honduras
 1. El Amatillo
 2. El Poy
 3. Perquín
 4. Puente Integración
- Aeropuertos
 3. Aeropuerto Internacional de Ilopango
 4. Aeropuerto San Oscar A. Romero
- Puertos marítimos
 1. Puerto de Acajutla
 2. Puerto Bahía del sol

3. Puerto CORSAIN
4. Puerto La Unión
5. Puerto Unión Centroamericana
6. Puerto Meanguera del golfo

Actualmente la institución cuenta con 24 puntos donde cada día se realizan miles de peticiones a la base de datos, dependiendo de la afluencia de personas, la época del año, entre otras. También es importante saber que la DGME forma parte del ministerio de justicia y seguridad pública, por lo cual, las decisiones que toman sus gerentes y director general deben ser siempre pensados en la seguridad de la población.

Este contexto es para dar a conocer lo importante y vital que la institución cuente con una herramienta que de información de forma rápida y fácil a las gerencias que lo requieran.

5. Contexto empresarial

Para conocer el contexto en el que se va a desarrollar el proyecto, debemos conocer la institución, a continuación, se describe un poco sobre ella:

La Dirección General de Migración y Extranjería forma parte de las diferentes instituciones que trabajan bajo el enfoque de respeto y garantía de Derechos Humanos y migratorios y su actuar está regido bajo un marco legal amplio tanto del ordenamiento jurídico nacional como de instrumentos internacionales, sin embargo se tenía la limitante en cuanto a que la institución trabajaba con una Ley obsoleta y por ende desactualizada, por lo que se recurría a diversos decretos y normas jurídicas que regulaban su accionar no solo en materia migratoria sino de extranjería.

En virtud de lo anterior la DGME crea un grupo multidisciplinario constituido con personal con una amplia experiencia operativa y profesional para crear y redactar la elaboración de una nueva Ley que permitiera unificar y garantizar en un solo cuerpo legal el eficaz ordenamiento migratorio que regulara la entrada y salida de nacionales y de extranjeros del territorio nacional, así como el tránsito y la permanencia de estos últimos dentro del mismo, la nacionalización y naturalización, la expedición del documentos de viaje en un marco de respeto a los derechos humanos, conforme la Constitución, leyes e instrumentos internacionales, así como la organización y funciones de la Dirección General, entrando en vigencia la nueva Ley Especial de Migración y de Extranjería el día cuatro de Julio del año dos mil diecinueve y su reglamento operativo el nueve de julio del referido año. La referida ley se regirá por los principios de:

- Dignidad humana
- Movilidad humana,
- Interés superior de la niña, niño y adolescente
- Debido proceso
- Igualdad.
- Unidad familiar y reunificación
- Integración
- No devolución
- No sanción por ingreso irregular de las personas refugiadas o apátridas.

6. Identificación de los orígenes de datos

La institución cuenta con dos sistemas informáticos para el funcionamiento de casi todas las actividades que realizan.

6.1. Sistema Integrado

Es un sistema web que se utiliza para realizar movimientos migratorios en todos los puntos fronterizos, terrestres, aéreo y marítimo. Así como también se utiliza para generar carnés de extranjeros, registro de nuevos empleados, consulta de DUI y pasaportes entre otras actividades.

Este sistema es el más utilizado en la institución, cada empleado tiene acceso a él con diferentes perfiles de usuario, por ejemplo, para pasar permisos personales por falta a laborar, ver horas de tiempo compensatorio etc.

Esta plataforma también tiene un perfil para la generación de reportes predefinidos.

6.2. Sistema de Pasaportes.

Es un sistema de escritorio, con la finalidad de captar la información necesaria de las personas que saquen su pasaporte. Este sistema está desplegado en todas las sucursales en el país, así como también en todos los consulados alrededor del mundo donde tengan la potestad de emitir pasaportes. En captación de datos es de los más importantes, ya que este alimenta a una base de datos muy utilizada en la institución, la cual es la de pasaportes y todo su correcto flujo depende de la buena recepción de datos por medio de los empleados que manipulan el sistema.

En general estos dos sistemas antes expuestos son los pilares para que se genere la información, el flujo de la información depende del buen uso y las validaciones que tiene el sistema.

Es importante mencionar que ambos sistemas fueron desarrollados en la institución, ya que tiene su área de desarrollo informático, el cual se encarga de velar por el buen funcionamiento y futuras actualizaciones y mantenimiento de los sistemas que se usan dentro de la institución

La DGME nos ha brindado ciertos diagramas de sus bases de datos de producción, cabe aclarar que por motivos de seguridad no se puede exponer toda su base de datos, así que se mostrara una versión con nombres de tablas y campos alterados, también no mostrando todos los campos de las tablas respetando las medidas que la institución vea conveniente, ya que nuestro deber como profesionales es ser garantes de la seguridad de la información.

En el siguiente diagrama se muestra el flujo de datos central para el proyecto, ya que se tiene como objetivo elaborar una herramienta para que esta información esté disponible y ordenada con menor tiempo de ejecución. Tenemos la tabla central del flujo, la cual es “table_mov_migratorio” la cual se guardan más que todo llaves relacionales de otras tablas, las cuales podemos resaltar el código persona, el cual es la llave que identifica a la persona a la cual se está relacionando el movimiento migratorio, tenemos otros que son vitales para su registro, el idFrontera, el cual identifica desde que frontera (aérea, terrestre o marítimo) se está haciendo el movimiento migratorio, el tipo de viajero, motivo de viaje entre otra información relevante. Como dato importante, cabe resaltar que el movimiento de un menor tiene un flujo diferente, ya que se guarda la información de los responsables, ya sea los padres o encargados del menor.

Se tienen tablas catálogos, tales como motivos de viajes, tipos de documento de viaje el cual se distingue mediante el CódigoDocumentoPorPersona dentro de la tabla de table_mov_migratorio.

Existen tablas que actualmente se utilizan de bitácoras y a la vez de consulta, tal como “table_consulta_bitacora” la cual se tiene como respaldo de cierta información relevante para la identificación de las personas.

Como se puede ver existe una relación con la tabla llamada “carné” ya que, el nombre de la institución hace alusión a Extranjería, la cual es un área especial que se encarga de otorgar permisos a extranjeros para estar en el país, y este tipo de viajero, para que su transición sea rápida, se toma en cuenta los datos de su carné para realizar su movimiento migratorio.

Para la generación de movimiento aéreo, no está plasmado en este diagrama, ya que es uno de los flujos con mayor riesgo, pero en resumen se tiene información como aerolíneas, vuelos, aeropuertos alrededor del mundo, un motor dedicado a la captación de vuelos. Esto último es muy interesante, ya que este servicio o software captura de anticipado la información de viajeros que están saliendo de otros países y llegan a El Salvador, esta información siendo muy importante para la inteligencia de avanzada dentro de los aeropuertos.

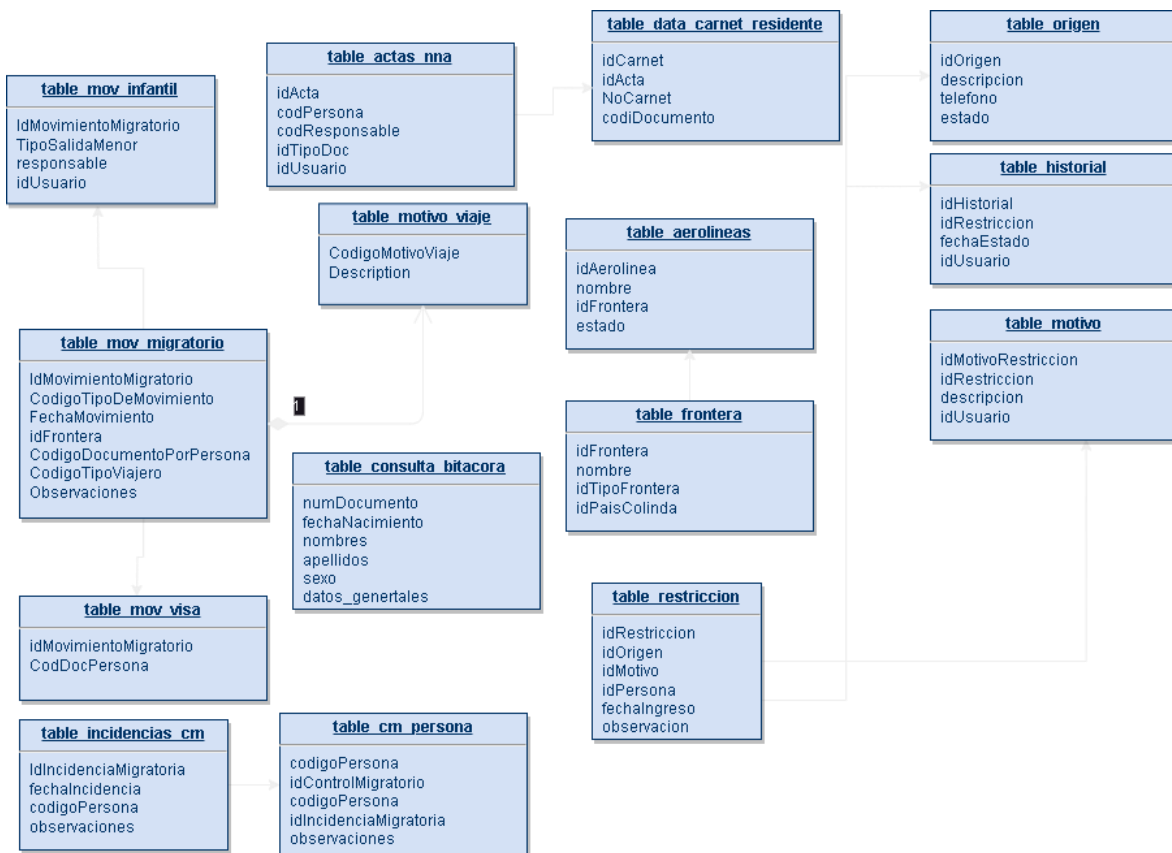


Ilustración 1. Base de datos actual de la Dirección General de Migración y Extranjería. Fuente: Gerencia de informática y desarrollo tecnológico de la DGME.

7. Justificación de la solución

Ante la problemática planteada anteriormente se ha construido una solución que permita a los diferentes interesados tener una mayor accesibilidad a los datos, esta solución está basada en construir un Data Warehouse que almacene toda la información importante para la construcción de reportes gerenciales.

El concepto de Data Warehouse se originó en 1988 con el trabajo de los investigadores de IBM, Barry Devlin y Paul Murphy, aunque el término Data Warehouse fue acuñado por William H. Inmon, el cual es conocido como el padre de Data Warehousing. Inmon describió un Data Warehouse como “un sistema de almacenamiento de datos diseñado para facilitar la toma de decisiones empresariales mediante la integración, organización y análisis de grandes volúmenes de datos provenientes de diversas fuentes. Proporciona una visión global y consolidada de los datos, permitiendo a los usuarios realizar consultas y generar informes para respaldar la toma de decisiones estratégicas.” (Inmon 2005).

Algunas de las razones que se tomaron en cuenta para el planteamiento de la solución e implementación de un Data Warehouse son las siguientes:

Integración de datos: un Data Warehouse permite integrar datos de diversas fuentes, como sistemas transaccionales, archivos, bases de datos externas, entre otros. Esto facilita la consolidación de datos dispersos y heterogéneos en una única fuente de información coherente y estructurada. Este punto es importante ya que la institución como vimos anteriormente posee 2 bases de datos necesarias para la generación de los reportes gerenciales.

Análisis y toma de decisiones: un Data Warehouse proporciona una base sólida para el análisis de datos y la generación de informes. Al contar con datos históricos y actuales almacenados de manera centralizada, los usuarios pueden realizar consultas complejas y obtener información estratégica para la toma de decisiones informadas.

Rendimiento optimizado: los Data Warehouse están diseñados para ofrecer un rendimiento óptimo en consultas y análisis complejos.

Utilizan técnicas como la optimización de consultas, el uso de índices y la agregación de datos pre calculados para acelerar los tiempos de respuesta y mejorar la experiencia del usuario.

Consistencia y calidad de los datos: al centralizar los datos en un Data Warehouse, se pueden aplicar reglas de calidad y consistencia de datos. Esto implica la limpieza, transformación y validación de los datos antes de su carga en el Data Warehouse, lo que mejora la confiabilidad y la integridad de la información. Estos procesos son llamados ETL y serán definidos más adelante en este documento.

Apoyo a la toma de decisiones estratégicas: el acceso a datos históricos y a largo plazo en un Data Warehouse permite el análisis de tendencias y patrones a largo plazo. Esto ayuda a las organizaciones a identificar oportunidades de crecimiento, detectar áreas de mejora y tomar decisiones estratégicas basadas en información confiable y precisa.

Escalabilidad y adaptabilidad: los Data Warehouse están diseñados para manejar grandes volúmenes de datos y soportar cargas de trabajo analíticas intensivas. Además, son flexibles y pueden adaptarse a medida que los requisitos de análisis y las necesidades de la organización evolucionan.

Aptitudes de la institución: además de las razones y ventajas antes descritas, es un punto importante para destacar que antes de decantarse por una solución se verifico que la institución cuente con los insumos para llevar a cabo esta solución, en este caso la DGME cuenta con la infraestructura y equipo adecuado para poderlo llevar a cabo. Más adelante se detallarán las herramientas y el personal que será necesario para realizar cada una de las fases de implementación de la solución.

A continuación, se describen algunos de los aspectos más importantes del Data Warehouse. Un Data Warehouse es una colección de datos orientada a temas, integrada, no volátil y variable en el tiempo en apoyo de las decisiones de gestión. El Data Warehouse contiene datos corporativos granulares. Los datos en el Data Warehouse se pueden utilizar para muchos propósitos diferentes, incluido sentarse y esperar requisitos futuros que se desconocen en la actualidad.

La orientación temática del Data Warehouse se muestra en la Figura 2. Los sistemas de operaciones clásicos se organizan en torno a las aplicaciones funcionales de la empresa. Para una compañía de seguros, las aplicaciones pueden ser para el procesamiento de automóviles, vida, salud y accidentes. Las principales áreas temáticas de la corporación de seguros pueden ser el cliente, la póliza, la prima y la reclamación. Para un fabricante, las principales áreas temáticas pueden ser producto, pedido, proveedor, lista de materiales y materias primas. Para un minorista, las principales áreas temáticas pueden ser producto, SKU, venta, proveedor, etc. Cada tipo de empresa tiene su propio conjunto único de temas.

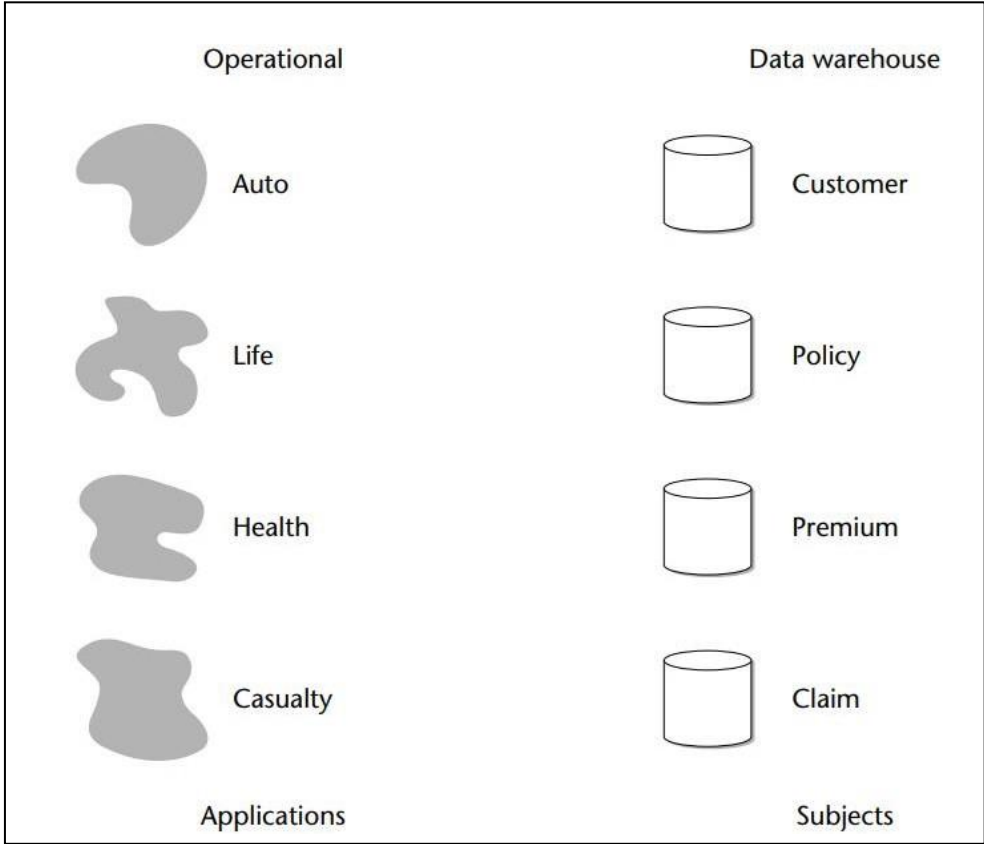


Ilustración 2. Ejemplo de la orientación temática de los datos. Fuente: Inmon (2005)

La segunda característica destacada del Data Warehouse reside en su nivel de integración. Entre todos los aspectos que conforman un Data Warehouse, la integración se posiciona como el más crucial. Los datos son extraídos de diversas fuentes heterogéneas y consolidados dentro del Data Warehouse. Durante este proceso, los datos se transforman, reformatean, reseleccionan, resumen, entre otros. Como resultado, una vez que los datos residen en el Data Warehouse, se obtiene una única imagen corporativa física.

La Ilustración 3 ejemplifica la integración que ocurre al trasladar los datos desde el entorno operativo, orientado a aplicaciones, hacia el Data Warehouse.

Las decisiones de diseño adoptadas por los desarrolladores de aplicaciones a lo largo de los años se manifiestan de diversas formas. En el pasado, los diseñadores de aplicaciones no contemplaban la posibilidad de integrar los datos con otros conjuntos de datos al crear una aplicación. Tal consideración era considerada una mera teoría descabellada.

Como resultado, existe una falta de consistencia entre las aplicaciones en términos de codificación, convenciones de nomenclatura, atributos físicos, medición de atributos, entre otros aspectos. Cada diseñador de aplicaciones ha tenido total libertad para tomar sus propias decisiones de diseño. Como resultado, cada aplicación difiere significativamente de las demás.

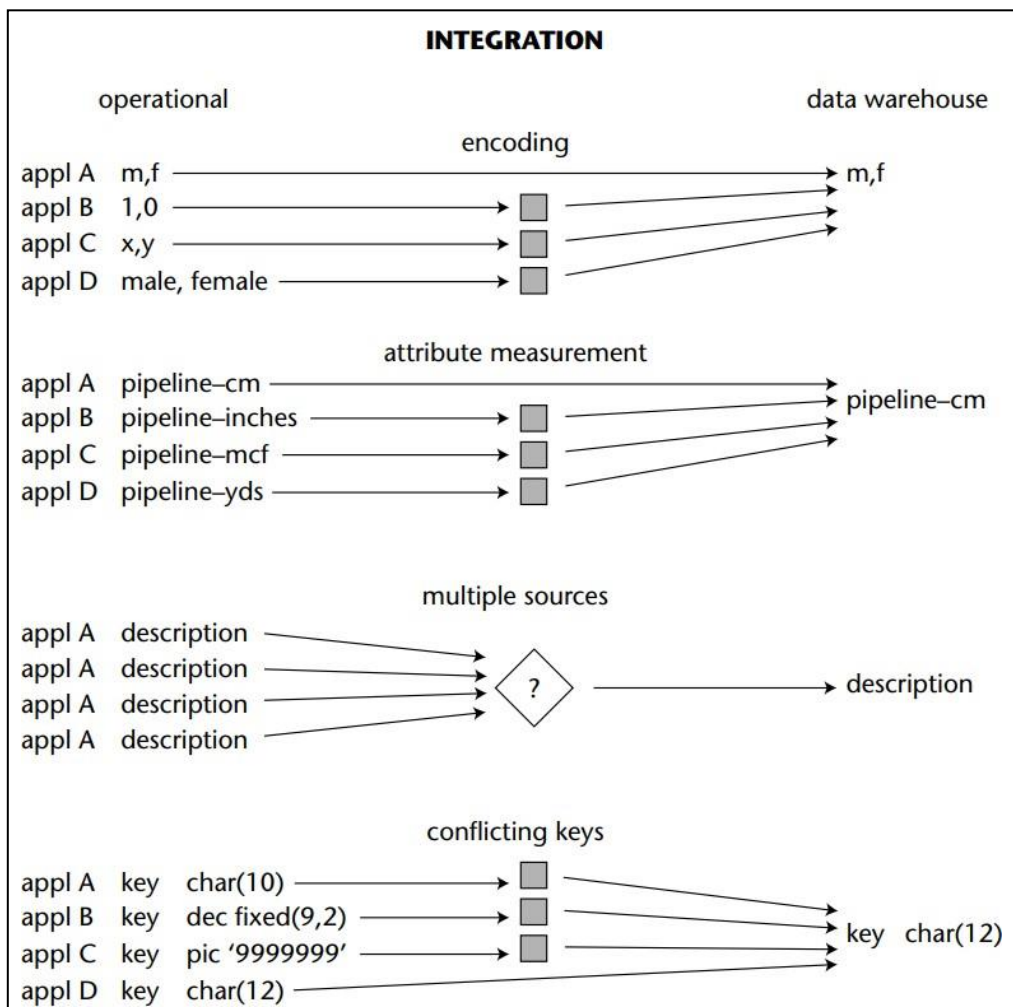


Ilustración 3. La cuestión de la integración. Fuente: Inmon (2005)

Los datos se ingresan en el Data Warehouse de tal manera que se deshacen las muchas inconsistencias en el nivel de la aplicación. Por ejemplo, como se mostró anteriormente en la ilustración 3, en lo que se refiere a la codificación de género, importa poco si los datos en el almacén están codificados como m/f o 1/0. Lo que importa es que, independientemente del método o la aplicación de origen, la codificación del almacén se realiza de forma coherente. Si los datos de la aplicación están codificados como X/Y para el sexo, se convierten a medida que se mueven al almacén. La misma consideración de coherencia se aplica a todos los problemas de diseño de aplicaciones, como las convenciones de nomenclatura, la estructura clave, la medición de atributos y las características físicas de los datos.

La tercera característica importante de un Data Warehouse es que no es volátil. La ilustración 4 muestra la no volatilidad de los datos y muestra que los datos operativos se acceden y manipulan regularmente, un registro a la vez. Los datos se actualizan en el entorno operativo de forma habitual, pero los datos del Data Warehouse exhiben un conjunto de características muy diferente. Los datos del Data Warehouse se cargan (generalmente, pero no siempre, en masa) y se accede a ellos, pero no se actualizan (en el sentido general). En cambio, cuando se cargan los datos en el Data Warehouse, se cargan en una instantánea, en formato estático. Cuando se producen cambios posteriores, se escribe un nuevo registro de instantánea. Al hacerlo, se mantiene un registro histórico de datos en el Data Warehouse.

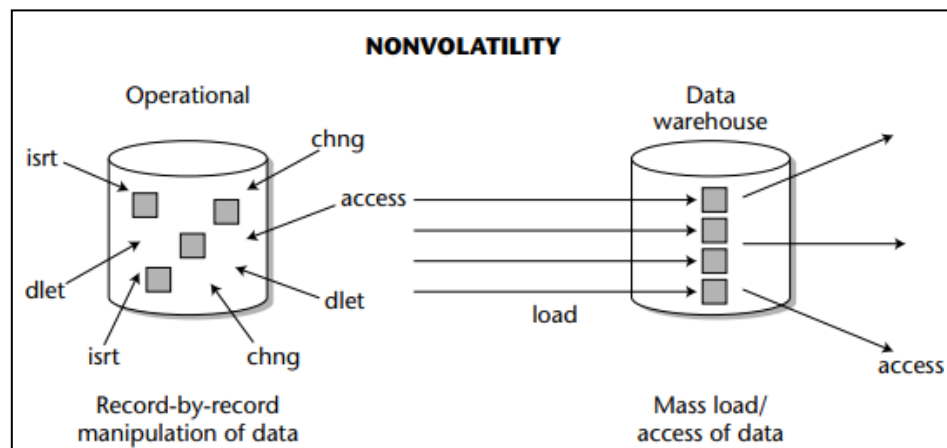


Ilustración 4. La cuestión de la no volatilidad. Fuente: Inmon (2005)

La última característica destacada del Data Warehouse es que es variable en el tiempo. La variación de tiempo implica que cada unidad de datos en el Data Warehouse es precisa en algún momento en el tiempo. En algunos casos, un registro tiene una marca de tiempo.

En otros casos, un registro tiene una fecha de transacción. Pero en todos los casos, hay algún tipo de marca de tiempo para mostrar el momento en el tiempo durante el cual el registro es preciso. La ilustración 5 muestra cómo la variación temporal de los datos del Data Warehouse puede mostrarse de varias maneras.

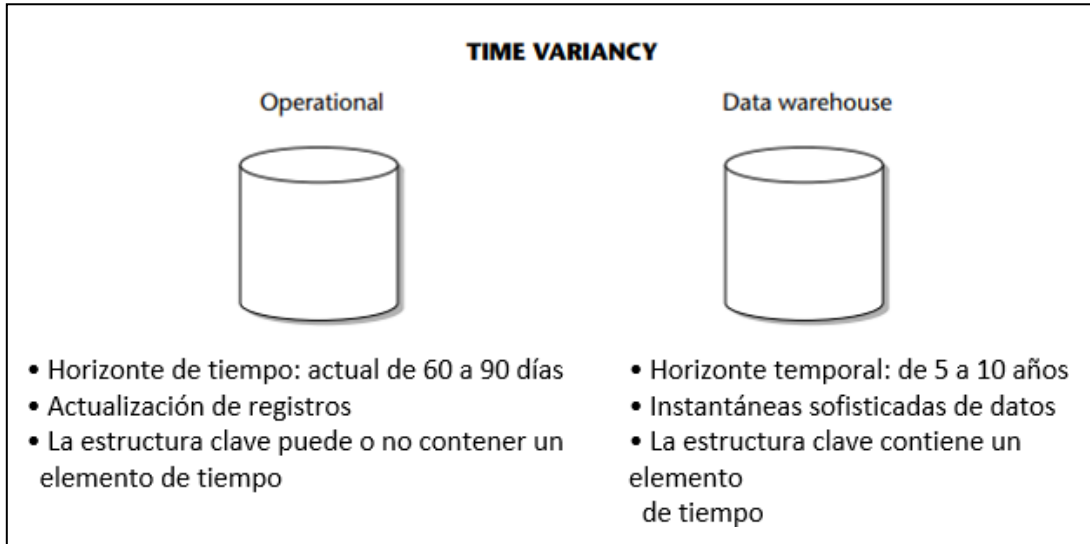


Ilustración 5. La cuestión de la variación temporal. Fuente: Elaboración propia a partir de Inmon (2005)

Diferentes entornos tienen diferentes horizontes de tiempo asociados con ellos. Un horizonte temporal es el período de tiempo durante el cual se representan los datos en un entorno. El horizonte de tiempo colectivo para los datos que se encuentran dentro de un Data Warehouse es significativamente más largo que el de los sistemas operativos. Un horizonte temporal de 60 a 90 días es normal para los sistemas operativos; un horizonte temporal de 5 a 10 años es normal para el Data Warehouse. Como resultado de esta diferencia en los horizontes temporales, el Data Warehouse contiene mucho más historial que cualquier otro entorno.

La estructura clave de los datos operativos puede o no contener algún elemento de tiempo, como año, mes, día, etc. La estructura clave del Data Warehouse siempre contiene algún elemento de tiempo. La incrustación del elemento de tiempo en el registro del Data Warehouse puede adoptar muchas formas, como una marca de tiempo en cada registro, una marca de tiempo para toda la base de datos, etc.

7.1. La estructura del Data Warehouse

La ilustración 6 muestra que existen diferentes niveles de detalle en el entorno del Data Warehouse. Hay un nivel de detalle más antiguo (generalmente en almacenamiento masivo alternativo), un nivel de detalle actual, un nivel de datos ligeramente resumidos (el nivel de Datamart) y un nivel de datos altamente resumidos. Los datos fluyen hacia el Data Warehouse desde el entorno operativo. Por lo general, se produce una transformación significativa de los datos al pasar del nivel operativo al nivel del Data Warehouse.

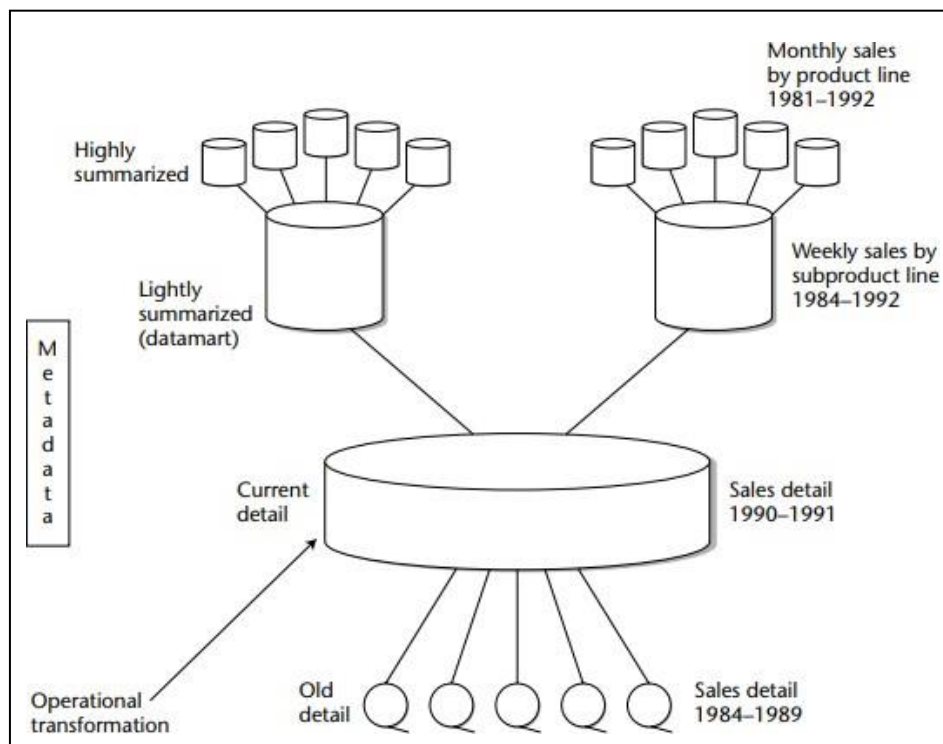


Ilustración 6. La estructura del Data Warehouse. Fuente: Inmon (2005)

Una vez que los datos en el Data Warehouse envejecen, pasan del detalle actual al detalle más antiguo. A medida que los datos se resumen, pasan de los detalles actuales a los datos ligeramente resumidos, y luego de los datos ligeramente resumidos a los datos altamente resumidos.

7.2. Tipos de modelos en el esquema multidimensional

Un esquema multidimensional es una estructura de datos utilizada en el diseño de un Data Warehouse que organiza y representa la información de manera eficiente para facilitar el análisis multidimensional. Se basa en el concepto de modelos dimensionales y se utiliza para representar datos en forma de cubos de datos, donde cada dimensión del cubo representa una característica o atributo relevante para el análisis.

En un esquema multidimensional, los datos se organizan en dimensiones y medidas. Las dimensiones representan las características o atributos que describen los datos y se utilizan para filtrar, agrupar y visualizar los datos. Por ejemplo, en un Data Warehouse de ventas, las dimensiones podrían ser el tiempo, el producto, el cliente y la ubicación geográfica. Las medidas, por otro lado, son los valores numéricos que se analizan, como el total de ventas, el número de unidades vendidas, etc.

El esquema multidimensional se basa en la idea de que los datos se pueden representar en forma de una matriz o tabla cruzada, donde cada celda contiene un valor numérico que representa una medida específica en la intersección de las dimensiones correspondientes. Esta representación facilita el análisis y permite a los usuarios explorar y visualizar los datos desde diferentes perspectivas.

Al construir un esquema multidimensional, se puede hacer uso de uno de los siguientes 3 modelos:

- Modelo estrella.
- Modelo copo de nieve.
- Modelo galaxia (o constelación).

El ultimo se comprende como la unión de 2 o más modelos de tipo estrella. A continuación, se detallarán las diferencias que existen entre un modelo estrella y un modelo copo de nieve, que más adelante en este documento se utiliza como modelo de datos para la solución propuesta a la DGME.

7.2.1. ¿Qué es el esquema estrella?

El modelo tipo estrella, es uno de los modelos más populares y ampliamente adoptados debido a su simplicidad y eficiencia en el análisis de datos.

En el modelo tipo estrella, los datos se organizan alrededor de una tabla central llamada tabla de hechos que contiene las medidas numéricas o cuantitativas que se desean analizar, como las ventas, los ingresos o las unidades vendidas. La tabla de hechos está rodeada por tablas de dimensiones, que representan los atributos o características relacionados con las medidas.

Cada tabla de dimensiones representa una dimensión específica, como el tiempo, el producto, el cliente o la ubicación geográfica. Estas tablas contienen los atributos descriptivos de la dimensión, como el nombre del producto, la fecha, la dirección del cliente, etc. Estas dimensiones están vinculadas a la tabla de hechos a través de claves externas.

La tabla de hechos contiene claves externas que se conectan a las claves primarias en las tablas de dimensiones. Esta estructura permite realizar análisis y consultas eficientes, ya que las dimensiones se pueden utilizar para filtrar y agrupar los datos, mientras que la tabla de hechos contiene las medidas para el análisis numérico.

El modelo tipo estrella ofrece ventajas en términos de rendimiento y facilidad de uso, ya que las consultas suelen ser rápidas y sencillas de construir. Además, la estructura desnormalizada de este modelo facilita la comprensión y la navegación de los datos.

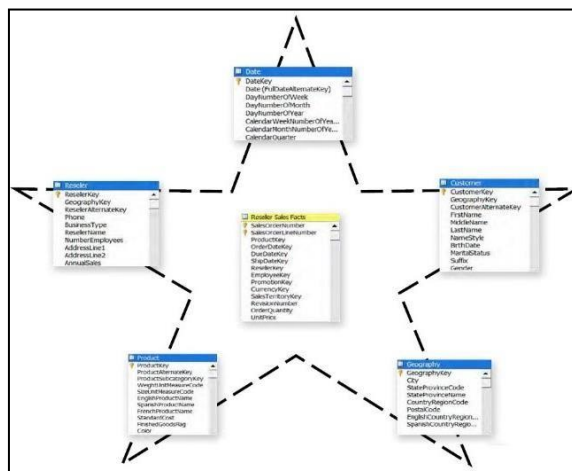


Ilustración 7. Esquema tipo estrella. Fuente: www.sqlybi.com

7.2.2. ¿Qué es el esquema copo de nieve?

El esquema copo de nieve, también conocido como modelo copo de nieve, es otro tipo de modelo dimensional utilizado en el diseño de un Data Warehouse. Es una variante del modelo tipo estrella y se caracteriza por una mayor normalización de las dimensiones.

En el esquema copo de nieve, las dimensiones se descomponen en subdimensiones y se normalizan en múltiples tablas. Esto significa que las tablas de dimensiones se dividen en tablas más pequeñas, lo que resulta en una estructura más compleja y en una mayor cantidad de tablas en comparación con el modelo tipo estrella.

La normalización en el esquema copo de nieve se logra mediante la separación de atributos de la dimensión en tablas independientes. Por ejemplo, en lugar de tener todos los atributos de producto en una sola tabla de dimensión, se pueden tener tablas separadas para la información del producto, la categoría del producto, el proveedor, etc. Cada tabla de subdimensión contiene un conjunto específico de atributos relacionados con ese aspecto de la dimensión.

La estructura del esquema copo de nieve puede parecer similar a un copo de nieve, ya que los subdimensiones se asemejan a los copos de nieve que se forman a partir de una estructura central. De ahí proviene su nombre.

El esquema copo de nieve tiene algunas ventajas, como una mayor eficiencia de almacenamiento debido a la normalización, lo que puede resultar en un menor consumo de espacio. Además, puede ser útil cuando se tienen dimensiones con muchos atributos y se necesita una mayor flexibilidad para agregar nuevos atributos en el futuro.

Sin embargo, el esquema copo de nieve también puede tener desventajas, como una mayor complejidad de diseño y consultas más complejas que pueden afectar el rendimiento.

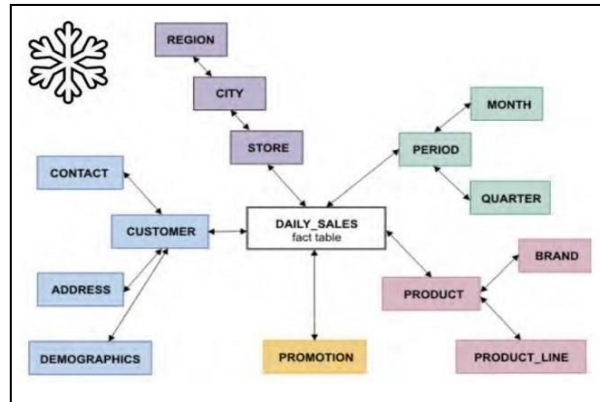


Ilustración 8. Esquema copo de nieve. Fuente: www.sqlybi.com

7.2.3. ¿Cómo funciona el esquema estrella?

La tabla de hechos, ubicada en el centro del modelo estrella, almacena dos tipos de información: valores de atributos numéricos y valores de atributos de dimensión. Para comprender mejor, consideremos un ejemplo de una base de datos de ventas.

Los valores numéricos son únicos para cada fila y punto de datos, sin correlación ni relación con los datos almacenados en otras filas. Estos valores representan datos específicos de una transacción, como el monto total de la venta, la cantidad de productos pedidos, la hora exacta de la transacción, la ganancia neta obtenida, el ID del pedido, entre otros.

Por otro lado, los valores de atributos dimensionales no almacenan datos directamente, sino que contienen valores de claves externas que hacen referencia a filas en tablas dimensionales. Estos valores dimensionales proporcionan información adicional relacionada con la fila en la tabla central, como el valor de los datos, la identificación del empleado de ventas, la identificación de la sucursal o tienda, la identificación del producto, entre otros.

Las tablas de dimensiones siempre almacenan información de soporte para la tabla de hechos. Cada tabla dimensional se relaciona con una columna en la tabla de hechos utilizando un valor dimensional y almacena datos adicionales relacionados con ese valor.

Este esquema además posee las siguientes características:

- El esquema en estrella permite filtrar datos normalizados para satisfacer las necesidades de almacenamiento de datos, generando una clave única a partir de la información asociada a cada tabla de hechos para identificar cada fila.
- Este esquema ofrece cálculos y agregaciones rápidos, como los ingresos obtenidos y el total de artículos vendidos al final de cada mes. Estos detalles pueden ser filtrados según sea necesario mediante la formulación de consultas apropiadas.
- La tabla de hechos representa eventos que contienen valores numéricos finitos, los cuales están compuestos por claves foráneas relacionadas con las tablas de dimensiones. Existen diversos tipos de tablas de hechos que se estructuran con valores a nivel atómico.
- La tabla de hechos de transacciones contiene datos sobre eventos específicos, como ventas y días festivos. Por otro lado, los hechos de registro incluyen información de cuenta al final del año o de cada trimestre.
- La tabla dimensional proporciona datos detallados sobre los atributos o registros presentes en la tabla central. Esto permite que el usuario diseñe una estructura acorde a sus necesidades.
- Además, el esquema en estrella se puede utilizar para acumular tablas de instantáneas en el Data Warehouse.

7.2.4. ¿Cómo funciona el esquema copo de nieve?

Este modelo es similar al esquema estrella, pero con cambios mínimos. A diferencia del esquema estrella, el esquema copo de nieve expande sus tablas de subdimensiones, las cuales se encuentran vinculadas a las tablas de dimensiones.

El propósito principal de este modelo es normalizar la información que se encuentra desnormalizada en el modelo estrella. De esta manera, se pueden solucionar problemas comunes asociados con el esquema estrella.

En el núcleo del esquema, se encuentra una tabla de hechos que se conecta con la información contenida en las tablas de dimensiones. A su vez, estas tablas se extienden hacia las tablas de subdimensiones que contienen información detallada que describe la información presente en las tablas de dimensiones.

Este esquema además posee las siguientes características:

- El esquema copo de nieve requiere una menor cantidad de espacio en disco debido a su estructura.
- Este modelo se destaca por su facilidad de implementación gracias a la utilización de tablas de dimensiones separadas y principales.
- Las tablas de dimensiones contienen al menos dos atributos que permiten definir información en múltiples niveles de detalle o granularidad.
- Es importante tener en cuenta que, debido al mayor número de tablas presentes en el esquema copo de nieve en comparación con el esquema en estrella, el rendimiento puede ser más bajo.
- No obstante, el esquema copo de nieve ofrece un nivel de integridad de datos más alto y presenta menos redundancias debido a la normalización de las tablas.

7.2.5. Ventajas y limitaciones del esquema tipo estrella

Algunas de las ventajas que podemos encontrar en utilizar un esquema de tipo estrella son las siguientes:

- El esquema en estrella representa la forma más sencilla entre los esquemas de Datamart disponibles.
- Este modelo se caracteriza por su lógica de informes simple, la cual está implícita de manera dinámica.
- El diseño del esquema en estrella se basa en la alimentación de cubos a través del proceso de transacción en línea, lo que permite que los cubos funcionen de manera eficiente y efectiva.
- El esquema en estrella se construye utilizando una lógica y consultas simples que son fáciles de extraer del proceso transaccional.
- Este enfoque ofrece un rendimiento mejorado para las aplicaciones de generación de informes y se implementa para garantizar una recuperación rápida de datos.
- La información filtrada y seleccionada se puede aplicar fácilmente en diferentes casos, lo que brinda flexibilidad en la obtención de datos pertinentes.

En relación con las limitaciones, se puede destacar que el esquema en estrella presenta un alto nivel de desnormalización e integridad. Si el usuario no puede actualizar los datos, el proceso en su totalidad puede verse afectado. Asimismo, las medidas de seguridad y protección son limitadas en este esquema.

Además, en comparación con el modelo analítico, el esquema en estrella no ofrece la misma flexibilidad. No brinda un soporte eficiente para múltiples relaciones, lo que puede limitar su capacidad para manejar y representar adecuadamente conexiones complejas entre los datos.

7.2.6. Ventajas y limitaciones del esquema copo de nieve

Algunas de las ventajas que podemos encontrar en utilizar un esquema de tipo copo de nieve son las siguientes:

- **Ahorro de espacio en disco:** Debido a su estructura normalizada, el modelo copo de nieve tiende a requerir menos espacio en disco en comparación con otros modelos, como el modelo estrella. La separación de atributos en tablas de subdimensiones permite eliminar redundancias y optimizar el almacenamiento de datos.
- **Mayor flexibilidad y escalabilidad:** El modelo copo de nieve proporciona mayor flexibilidad para agregar nuevas dimensiones o atributos a los subdimensiones existentes sin afectar directamente la tabla central de hechos. Esto facilita la adaptación del modelo a medida que evolucionan las necesidades del negocio y se agregan más datos.
- **Mejor rendimiento en consultas específicas:** En algunas situaciones, el modelo copo de nieve puede ofrecer un mejor rendimiento en consultas específicas en comparación con el modelo estrella. Esto se debe a la capacidad de dividir las dimensiones en subdimensiones y evitar la duplicación de datos en la tabla central de hechos.
- **Mayor nivel de normalización:** La normalización en el modelo copo de nieve ayuda a mantener un alto nivel de integridad de datos y reduce la redundancia. Cada tabla de subdimensión almacena información específica y se relaciona con otras tablas a través de claves externas, lo que permite un mejor control de la calidad de los datos.
- **Manejo eficiente de dimensiones con muchos atributos:** Cuando se tienen dimensiones con una gran cantidad de atributos, el modelo copo de nieve puede resultar más eficiente y manejable que el modelo estrella. La estructura descompuesta en subdimensiones facilita la gestión y el mantenimiento de atributos complejos.

La principal limitación que se presenta con el modelo copo de nieve (Snowflake) reside en los esfuerzos adicionales de mantenimiento que requiere debido al aumento en el número de tablas de dimensiones más pequeñas. A medida que se agregan más tablas al esquema, el mantenimiento y la gestión se vuelven más complejos.

Además, la ejecución de consultas complejas puede dificultar la búsqueda y extracción de los datos necesarios. La complejidad del modelo puede afectar el rendimiento y la eficiencia en la recuperación de datos.

Otra limitación del modelo copo de nieve es que el tiempo de implementación de las consultas puede ser más largo debido al mayor número de tablas involucradas en comparación con otros modelos. Esto puede afectar el tiempo de respuesta y la capacidad de generar resultados rápidamente.

También se debe tener en cuenta que el modelo copo de nieve tiende a ser más rígido en términos de estructura y requiere mayores costos de mantenimiento en comparación con otros enfoques de modelado. Es importante considerar los recursos necesarios para mantener y actualizar las tablas y relaciones del modelo.

8. Construcción del nuevo diagrama de datos

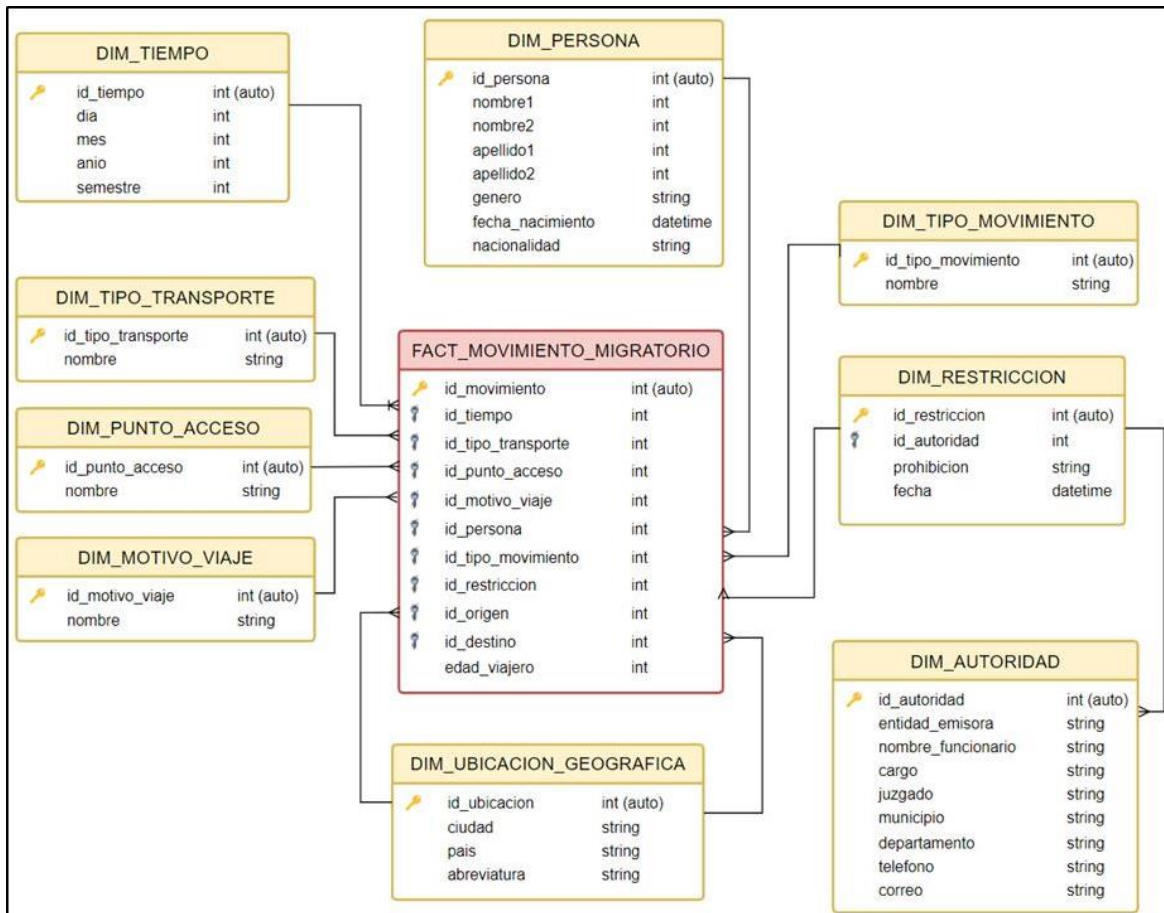


Ilustración 9. Diagrama del Data Warehouse. Fuente: Elaboración propia

La Ilustración 2, muestra el diagrama de Data Warehouse que se ha construido tomando como base los orígenes de datos descritos. Este Data Warehouse pretende facilitar la obtención de reportes por edades, genero, fecha, etc.

Es importante destacar que este es un modelo funcional que se adapta al requerimiento planteado por la DGME, pero que, sin embargo, puede crecer en el tiempo en caso de que en algún futuro sea necesario generar reportes con información que aún no este contemplado en ese modelo, es por ello por lo que a continuación se describirá el proceso de construcción de este modelo, a modo que sea más intuitivo el proceso de agregar más datos.

Este es un Data Warehouse de tipo copo de nieve, ya que cuenta con una sola tabla de hechos (se muestra en rojo) y muchas tablas de dimensiones (en amarillo), de la cual una está segregada para aportar mayor granularidad a los datos (la dimensión de restricción posee una llave foránea hacia la dimensión de autoridad), convirtiendo en modelo de tipo estrella a copo de nieve.

La tabla de hechos (FACT_MOVIMIENTO_MIGRATORIO) contiene los aspectos de negocio, en este caso particular, los movimientos migratorios y también valores pre calculados. Es importante tener presente que esta tabla de hechos debe contener solo datos cuantificables, por ejemplo, para un Data Warehouse de una tienda, tendría sentido agregar precios o cantidades en stock de un producto; pero para los movimientos migratorios, la mayoría de los reportes se basan en cantidad de movimientos dadas ciertas características, es por ello por lo que solo se cuenta con un campo de edad, ya que el valor cuantificable que se estará utilizando será la cantidad de registros como tal. En cuanto al campo de edad, es un valor que se utiliza en muchos de los reportes de la DGME por lo tanto se almacenará en un campo específico al momento de registrar el movimiento migratorio, este es un ejemplo de porque las consultas utilizando este tipo de modelos es más rápida, pues en el actual modelo normalizado, obtener un reporte por edades requiere de realizar la operación matemática utilizando la fecha de nacimiento de la persona, sin mencionar la obtención de grandes cantidades de datos que no se utilizan.

Por otro lado, las dimensiones servirán para filtrar las características que determinan un movimiento migratorio como lo son la ubicación de origen y destino, la persona que viaja, el tipo de movimiento, el punto de acceso, entre otras cosas. Estas son tablas de catálogo que deberán ser pobladas por medio de procesos ETL que se describirán más adelante, por ello es importante entender que datos debe almacenar cada tabla de dimensiones. Para las entidades mostradas en la imagen 1, cada una ha sido elaborada a partir de información que existe en el modelo relacional.

9. Indicadores de rendimiento

El propósito de la implementación del Data Warehouse en la Dirección General de Migración y Extranjería es lograr una mejora notoria en la eficiencia y precisión en el manejo de datos relacionados con los movimientos migratorios. Para evaluar el impacto y éxito de esta implementación, se aplicarán los Key Performance Indicators (KPI) como métricas clave para medir el rendimiento y el logro de los objetivos establecidos.

Los Key Performance Indicators (KPI), también conocidos como Indicadores Clave de Rendimiento, son métricas cuantitativas utilizadas para evaluar el rendimiento y el logro de los objetivos de una organización o proyecto. Los KPI proporcionan una medida objetiva y concreta del desempeño en áreas específicas, permitiendo el monitoreo, la medición y el análisis de los resultados.

La selección adecuada de los KPI es fundamental para garantizar que se estén midiendo los aspectos clave del rendimiento que impactan directamente en los resultados deseados. Los KPI deben ser específicos, medibles, alcanzables, relevantes y estar limitados en tiempo, lo que se conoce como criterios SMART.

Al implementar los KPI, se establecen metas y se realizan mediciones periódicas para evaluar el desempeño actual en relación con esas metas. Esto permite identificar áreas de mejora, detectar desviaciones y tomar acciones correctivas para mantener o mejorar el rendimiento.

Los KPI pueden aplicarse en diversas áreas de una organización, como ventas, marketing, recursos humanos, operaciones, servicio al cliente, entre otras. También son ampliamente utilizados en proyectos para evaluar la eficacia en la ejecución y el logro de los objetivos establecidos.

Los KPI se seleccionarán en función de las prioridades y metas de la institución, definiendo indicadores de rendimiento específicos para cada uno de ellos. Asimismo, se recopilarán y almacenarán los datos necesarios en el DW, permitiendo su posterior análisis y seguimiento. Los resultados obtenidos mediante los KPI serán interpretados para identificar áreas de mejora y tomar acciones correctivas, con el fin de asegurar el cumplimiento efectivo de los objetivos establecidos en materia migratoria.

En la era de la información, las instituciones gubernamentales de todo el mundo enfrentan diversos desafíos en la gestión y análisis de datos.

Durante los últimos años el tema migratorio en Latinoamérica y principalmente en El Salvador ha sido objeto de diferentes estudios y cada institución tiene la tarea de modernizar sus procesos y saber interpretar la información que tienen en sus registros. La correcta administración de la información resulta fundamental para garantizar la seguridad, el control y la eficacia de los procesos migratorios. En este contexto, la implementación de un Data Warehouse (DW) emerge como una solución tecnológica que permite centralizar, integrar y analizar datos provenientes de diversas fuentes, facilitando la toma de decisiones y mejorando la eficiencia operativa.

Como parte de la investigación se prioriza que los KPI puedan ser utilizados de manera efectiva en la institución, siendo utilizados para medir y evaluar el rendimiento del DW en relación con los objetivos específicos establecidos por la institución, proporcionando una visión clara del impacto y la efectividad de la implementación en los procesos de control migratorio.

9.1. Ventajas de utilizar KPI en el control migratorio

La implementación de KPI en un Data Warehouse para el control migratorio brinda una serie de ventajas significativas en comparación con la ausencia de estos indicadores:

- **Mejora en la toma de decisiones:** Los KPI proporcionan información objetiva y cuantitativa sobre el rendimiento y los resultados de los procesos migratorios. Esto permite a los responsables de la toma de decisiones tener una visión clara de las áreas que requieren mejoras, así como identificar y priorizar acciones correctivas.
- **Monitoreo y seguimiento continuo:** Los KPI permiten establecer un sistema de monitoreo y seguimiento constante del desempeño del control migratorio. Esto posibilita la detección temprana de desviaciones o problemas, lo que facilita la adopción de medidas correctivas oportunas para mantener el cumplimiento de los objetivos establecidos.
- **Identificación de tendencias y patrones:** Los KPI proporcionan una visión amplia del comportamiento de los datos migratorios a lo largo del tiempo. Esto permite identificar tendencias, patrones y anomalías que podrían no ser evidentes de manera intuitiva.

Al comprender y anticipar estos factores, la institución puede implementar estrategias **preventivas y adaptar sus políticas de control migratorio en consecuencia.**

- **Transparencia y rendición de cuentas:** Al tener KPI establecidos, la institución puede comunicar de manera clara y transparente su desempeño en el control migratorio. Estos indicadores proporcionan una base objetiva para evaluar el cumplimiento de los objetivos y brindar rendición de cuentas a los ciudadanos y otras partes interesadas. La transparencia fortalece la confianza en la institución y mejora su reputación en el ámbito migratorio.

9.2. Identificación de los KPI relevantes

- **Tiempo de respuesta para la emisión de documentos de migración:** Este KPI se enfoca en medir la eficiencia de la institución en la emisión oportuna de documentos migratorios. Un tiempo de respuesta rápido agiliza los trámites y mejora la experiencia del solicitante, lo que contribuye a un control migratorio más eficiente y una mayor satisfacción de los usuarios.

Acciones:

1. Establecer un proceso de seguimiento y registro de las solicitudes de documentos migratorios desde su recepción hasta la emisión final.
 2. Implementar un sistema de notificaciones automatizadas para mantener informados a los solicitantes sobre el estado de su trámite.
 3. Identificar y eliminar posibles cuellos de botella en el proceso de emisión de documentos migratorios, optimizando los recursos y la asignación de personal.
 4. Establecer metas y plazos claros para la emisión de cada tipo de documento migratorio, basados en la complejidad del trámite y los requisitos legales.
 5. Realizar un seguimiento regular del tiempo promedio de respuesta y compararlo con los objetivos establecidos, identificando oportunidades de mejora.
- **Precisión en la detección de casos de migración ilegal:** Este KPI evalúa la capacidad del Data Warehouse para identificar de manera precisa los casos de migración ilegal.

Al contar con un sistema centralizado y análisis de datos, la institución podrá detectar patrones y anomalías que indiquen actividades migratorias irregulares. Esto permitirá tomar medidas oportunas para prevenir y controlar la migración ilegal, fortaleciendo la seguridad y el cumplimiento de las regulaciones migratorias.

Acciones:

1. Implementar algoritmos y técnicas de análisis de datos en el Data Warehouse para identificar patrones y anomalías que puedan indicar posibles casos de migración ilegal.
 2. Integrar bases de datos y fuentes de información relevantes para tener una visión completa y actualizada de la situación migratoria.
 3. Establecer alertas y notificaciones automáticas para informar sobre posibles casos de migración ilegal y activar medidas de control adicionales.
 4. Capacitar al personal encargado del análisis de datos para interpretar correctamente los resultados y actuar de manera adecuada ante las alertas generadas.
 5. Realizar revisiones periódicas de los resultados obtenidos y ajustar los algoritmos y técnicas utilizados según sea necesario.
- **Eficiencia en la gestión de solicitudes y trámites migratorios:** Este KPI mide la eficiencia en el manejo de las solicitudes y trámites migratorios, desde su recepción hasta su resolución.

Con un Data Warehouse, la institución puede automatizar y agilizar los procesos, reduciendo los tiempos de espera, minimizando errores y optimizando los recursos. Una gestión eficiente contribuye a un control migratorio más efectivo y a una mayor transparencia en los procedimientos.

Acciones:

1. Mapear y optimizar los procesos de gestión de solicitudes y trámites migratorios, identificando posibles ineficiencias y puntos de mejora.
2. Automatizar aquellos procesos que sean susceptibles de ser ejecutados de manera más eficiente mediante el uso de herramientas tecnológicas, como formularios en línea y sistemas de gestión documental.
3. Implementar un sistema de seguimiento de solicitudes y trámites migratorios, que permita tener visibilidad en tiempo real del estado de cada uno de ellos.

4. Capacitar al personal involucrado en la gestión de solicitudes y trámites migratorios para asegurar un manejo eficiente de los mismos y garantizar la aplicación de mejores prácticas.
 5. Establecer indicadores de tiempo de respuesta y calidad en cada etapa del proceso y realizar seguimientos periódicos para medir el cumplimiento y tomar acciones correctivas si es necesario.
- **Mejora en la generación de informes y análisis de datos migratorios:** Este KPI se centra en la capacidad del Data Warehouse para generar informes completos y realizar análisis de datos migratorios de manera rápida y precisa. Al contar con una estructura de datos consolidada y herramientas de análisis avanzadas, la institución podrá obtener perspectivas y tomar decisiones informadas basadas en datos actualizados y confiables. Esto brinda una ventaja estratégica al mejorar la capacidad predictiva, la identificación de tendencias y la planificación de políticas migratorias.

Acciones:

1. Definir los requisitos y objetivos específicos de los informes migratorios, considerando las necesidades de los diferentes usuarios internos y externos.
2. Diseñar y desarrollar modelos y visualizaciones de datos relevantes para la toma de decisiones y el análisis de tendencias migratorias.
3. Garantizar la calidad y la integridad de los datos almacenados en el Data Warehouse, estableciendo mecanismos de validación y limpieza de datos.

10. Definición de los procesos de ETL

El sistema de extracción, transformación y carga (ETL por sus siglas en inglés) consta de un área de trabajo, estructuras de datos instanciadas y un conjunto de procesos. El sistema ETL es todo entre los orígenes de datos y el Data Warehouse. (Kimball y Ross 2013).

Hasta ahora hemos definido los orígenes de datos y hemos realizado el diagrama del Data Warehouse, que será nuestro destino de información y la fuente de la reportaría. Sin embargo, es importante que los datos sean trasladados de su origen a su destino de una manera óptima, y manteniendo la confiabilidad de los datos. Existen diversas herramientas para lograr este objetivo, más adelante en este documento se encuentra una comparativa de algunas de las alternativas por las que se puede optar, sin embargo, por ahora nos centraremos en la lógica del proceso como tal, por ese motivo ahora definiremos los procesos ETL que serán necesarios para poblar el Data Warehouse. Para empezar, se definirá las 3 diferentes fases del proceso ETL.

El primer paso del proceso ETL es la extracción de datos, en el cual se lee y entiende la información de los orígenes de datos, además se obtiene la información que será necesaria para el Data Warehouse y se prepara para la fase de transformación y carga.

En este caso el proceso de carga deberá realizar las consultas correspondientes a la base de datos de restricciones y movimientos migratorios, incluyendo solo los campos que aportan valor para poblar nuestro Data Warehouse.

Una vez se haya extraído los datos se pasa al proceso de transformación. En este proceso se realizan varias operaciones como totalizaciones, limpia de datos, corrección de datos, etc. Un ejemplo de transformación de datos muy común es el de las fechas, en nuestro Data Warehouse contamos con una tabla de hechos de movimientos migratorios, pero esta tabla tendrá registros provenientes de 2 fuentes; la base de datos de movimientos migratorios, y la base de datos de restricciones. Supongamos que ambas bases de datos almacenan las fechas en formato diferente, una parte del proceso de transformación es convertir esas fechas a un solo formato para mantener el orden de los datos. También en este proceso se calculará la edad del viajero en base a su fecha de nacimiento, etc.

Por último, se encuentra el proceso de carga de los datos, en el cual ya con los datos procesados procederemos a insertarlos al Data Warehouse dejándolo listo para reportes.

Estos procesos ETL deben ser automatizados y mantener los datos actualizados lo más rápido posible con respecto a los orígenes de datos, pero para ello es muy importante traducir entonces que significa cada uno de los campos en el destino, con base a los orígenes. A continuación, se describe la información que cada tabla del Data Warehouse deberá contener.

DIM_TIEMPO: esta dimensión será utilizada para almacenar la fecha de registro del movimiento migratorio, se obtendrá a partir del campo FechaMovimiento de la tabla movimiento_migratorio, es importante destacar como en esta dimensión se ha agregado un campo de semestre el cual deberá contener los valores de 1 o 2, esto facilita la obtención de reportes semestrales sin necesidad de operar fechas en la consulta, de esta misma manera se podrían agregar campos para cuartos de año o trimestres.

DIM_TIPO_TRANSPORTE: esta dimensión es un catálogo pre poblado a partir de la tabla tipodetransporte, Que servirá para identificar movimientos terrestres, marítimos o aéreos.

DIM_PUNTO_ACCESO: esta dimensión es un catálogo pre poblado a partir de la tabla frontera y la tabla puertos, que servirá para identificar el punto de acceso o salida en el cual se produjo el movimiento migratorio.

DIM_MOTIVOS_VIAJE: esta dimensión es un catálogo pre poblado a partir de la tabla motivosdeviaje, Que servirá para identificar la razón del registro de movimientos migratorios. Esta dimensión es importante para generar el reporte de retornados, que es uno de los motivos dentro del catálogo.

DIM_UBICACION_GEOGRAFICA: esta dimensión es un catálogo pre poblado a partir de la tabla ubicaciones geográficas, que servirá para almacenar los orígenes y destinos dentro del movimiento migratorio.

DIM_PERSONA: esta dimensión deberá contener la información de la persona motivo del movimiento migratorio, deberá ser poblada a partir de los valores de la tabla personas del modelo normalizado.

DIM_TIPO_MOVIMIENTO: esta dimensión es un catálogo pre poblado a partir de la tabla tipodemovimiento, que servirá para determinar si es un movimiento de entrada o salida del país entre otras cosas.

DIM_RESTRICCION: esta dimensión será útil para registrar movimientos migratorios que fueron restringidos. Proviene de la base de datos de restricciones de la tabla prohibiciones, además incluye a la entidad que realiza la restricción.

DIM_AUTORIDAD: esta dimensión especifica el ente que realiza una prohibición, tales como a la procuraduría, el ministerio de hacienda, entre otras. Sus valores provendrán de la tabla autoridad del modelo normalizado.

FACT_MOVIMIENTO_MIGRATORIO: esta tabla se mapeara con la tabla movimiento_migratorio del modelo normalizado, sin embargo, en el modelo normalizado solo se encuentran los registros de movimientos migratorios que se llevaron a cabo, no obstante, ya que se plantea la necesidad de incluir reportes de restricciones migratorias, también se usara esta tabla para almacenar aquellos movimientos que no se llevaron a cabo por restricciones migratorias, estos registros provendrán de la base de datos restricciones y de la tabla restricciones. En términos de reportaría se podrán diferenciar estos movimientos por medio de la llave foránea a la dimensión DIM_RESTICCION. Aquellos registros que posean un valor en esa llave pertenecen a movimientos que no se llevaron a cabo.

Una vez que hemos comprendido los orígenes y destinos de los datos, podemos definir cuáles y cuantos procesos de ETL será necesario realizar.

- El primer proceso será el de **ETL_MOVIMIENTOS_MIGRATORIOS** en este proceso se extraerá la data necesaria de la tabla de movimiento_migratorio y se puede realizar 1 vez por cada hora, a modo de tener la información lo más actualizada posible, este proceso puede tomar tiempo debido al volumen de datos que se pueden registrar en una hora. Es importante mencionar que, por cada registro, se creara uno nuevo en la tabla de hechos, sin embargo, las tablas de dimensiones solo registraran datos cuando estos no existan previamente, por ejemplo, en la dimensión de tiempo, solo deberá ingresar se 1 registro al día, ya que es el nivel de granularidad más bajo en esa tabla.
- Un segundo proceso ETL será el de **ETL_RESTRICCIONES** en el cual se cargará la data de la base de datos de restricciones y se registrará un movimiento migratorio en la tabla hechos.

A diferencia del proceso anterior se espera que la carga de datos de estas consultas sea mucho menor, pues no es el caso más común que se realice una restricción si lo comparamos con los movimientos migratorios que si se llevan a cabo.

Es recomendable que este proceso no se programe al mismo tiempo que el proceso anterior, para evitar sobrecargar el procesamiento del servidor en el cual se ejecutarán.

Es importante destacar que los procesos ETL son ejecutados a intervalos regulares para garantizar que la información se mantenga actualizada. Esto implica la repetición de las etapas de extracción, transformación y carga para capturar nuevos datos y reflejar los cambios en la situación migratoria y de los extranjeros.

En resumen, los procesos ETL en la Dirección General de Migración y Extranjería de El Salvador son fundamentales para gestionar y analizar datos relacionados con la migración y los extranjeros. Estos procesos permiten la extracción de datos de diversas fuentes, su transformación para mejorar la calidad y consistencia, y su carga en un repositorio centralizado para su posterior análisis y toma de decisiones informada.

11. Sistemas de información orientados a toma de decisiones.

Si bien tener datos es de carácter importante saber tomar las decisiones es crucial, de esta manera se puede asegurar que la información que se obtenga ayude a crecer a la organización o a la empresa. Para ello, hay ciertas herramientas que ofrecen la ayuda necesaria, una estructura para poder identificar el valor de la información.

Los sistemas de información (SI) están formados por elementos que se comunican e interactúan entre sí para apoyar las estrategias de un negocio y la toma de decisiones dentro de las organizaciones. Estos elementos están dados por datos, procesos, sistemas transaccionales, Data Warehouse o base de datos, interfases, información interna y externa.

Es importante que se cuenta de una estructura e infraestructura organizacional en la cual la información fluye en todos los sentidos (áreas de la empresa) teniendo en cuenta su entorno.

Todo sistema tiene cuatro actividades fundamentales: entrada, almacenamiento, procesamiento y salida de información. Que, dependiendo de la complejidad, involucra interfases automáticas de entrada y de salida.

Este tipo de sistema que apoyan el proceso de toma de decisiones están orientados a los altos ejecutivos y usuarios que tienen esta responsabilidad dentro de la organización. Estos son conocidos como DSS (Decisión Support System), EIS (Executive Information Systems) y numerosos sistemas expertos que cada día proveen información en tiempo real y de manera gráfica y resumida.

Los Sistemas de Información que se orientan a la toma de decisiones, están dirigidos a apoyar a los altos ejecutivos de una organización, presentando información relevante y haciendo uso de recursos visuales y de fácil interpretación, sacándole el mayor partido a la tecnología de los Sistemas de Información.

Las principales características son las siguientes:

- Se enfocan en cálculos más que sobre la entrada y salida de información. Así, por ejemplo, un modelo de indicadores de gestión necesita poca información de entrada, proporciona poca información como resultado, pero puede realizar muchos cálculos durante el proceso y accesos a la base de datos o repositorio de datos.

- Están diseñados a la medida de cada organización.
- Logra que el usuario desarrolle de manera directa los modelos sin la intervención de profesionales de informática, lo que ayuda para que no sea dependiente del área de Tecnología.
- Se les considera como soluciones que hacen parte del plan de mejoramiento organizacional y como un plan para lograr una ventaja competitiva.
- Se desarrollan con altos estándares en sus interfases, caracterizado por gráficas de alta calidad, información tabular y en forma de texto.
- Posee un protocolo de comunicación entre el ejecutivo y el sistema permite interactuar sin un entrenamiento previo.

11.1. Sistema de soporte a la decisión.

Los sistemas de soporte a la decisión (DSS) son herramientas tecnológicas diseñadas para ayudar a las personas o las organizaciones a tomar decisiones más informadas y efectivas. Estos sistemas utilizan una combinación de datos, modelos analíticos y software especializado para proporcionar información relevante y análisis detallados que respalden el proceso de toma de decisiones.

Como plantea la definición anteriormente, los DSS resuelven problemas con diferentes grados de estructura (Turban, Aronson, & Liang, 2005)

- Decisión estructurada: se realiza de acuerdo con procesos específicos.
- Decisión no estructurada: posee un alto grado de libertad e incertidumbre. La información que se obtiene de un sistema es solo una porción del conocimiento total que se necesita para tomar una decisión.
- Decisión semiestructurada: existen componentes estructurados como, por ejemplo: datos fijos, modelos o reglas de decisión, pero la decisión pertenece al usuario.

Adicionalmente se conoce que un sistema DSS es una de las herramientas más emblemáticas del Business Intelligence, debido a que ofrece entre otras propiedades, poder resolver gran parte de las limitaciones de los programas de gestión. Estas son algunas de sus características principales (Bonilla Botia & Briceño Díaz, 2006):

- Informes dinámicos, flexibles e interactivos, de manera que el usuario no tenga que limitarse a los listados predefinidos que se configuraron en el momento de la implantación, y que, probablemente, no siempre responden a sus dudas reales.
- No requiere conocimientos técnicos. Como se mencionó en un apartado anterior, no es necesario estar anclado al grupo de tecnología de la empresa y organización, un usuario no sin previo conocimiento técnico de la herramienta puede crear nuevos gráficos e informes y navegar entre ellos. Por tanto, para analizar la información disponible o hasta llegar a crear nuevas métricas.
- Rapidez en el tiempo de respuesta, debido a que la base de datos suele ser un Data Warehouse corporativo o un Datamart, con modelos de datos en estrella o copo de nieve. Este tipo de bases de datos se caracterizan por estar optimizadas para el análisis de grandes volúmenes de datos.
- Integridad entre todos los sistemas o departamentos de la compañía. El proceso de ETL previo a la implantación de un Sistema de Soporte a la Decisión asegura la calidad y la integración de los datos entre las diferentes unidades de la empresa. Llegando a lo que se llama: integridad referencial absoluta.
- Cada usuario tiene información adecuada a su perfil. Está segmentado, no es conveniente que toda la organización tenga acceso a toda la información, sino de que el usuario tenga acceso a la información que necesita para que su trabajo sea lo más eficiente posible.
- Información histórica. En estos sistemas está a la orden del día comparar los datos actuales con información de otros períodos históricos de la compañía, con el fin de analizar tendencias, fijar la evolución de parámetros de negocio, entre otros.

A continuación, podemos una imagen de carácter ilustrativo de los componentes que conforman a un Sistema de Soporte de Decisiones (Bonilla Botia & Briceño Díaz, 2006). (Figura 10).

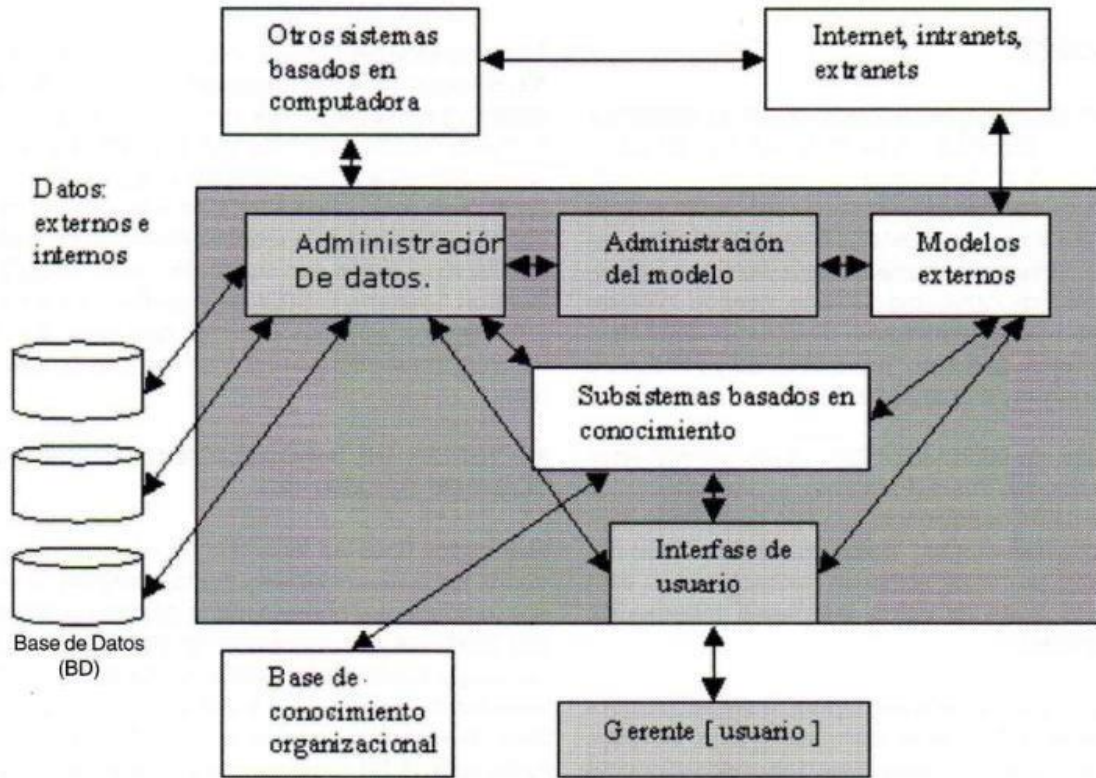


Ilustración 10. Componentes de un sistema de soporte a la toma de decisiones fuente (Bonilla Botia & Briceño Díaz, 2006)

Entre sus componentes tenemos:

- Las bases de datos (BD) es la estructura de datos de la organización, debido a esto es uno de los aportes más positivos para los Sistemas de Información, ya que provee a la organización de los datos de la empresa.
- El subsistema de datos del DSS está compuesto de la base de datos del DSS, del sistema de administración de la base de datos, del directorio de datos y de la facilidad para hacer consultas.
- El subsistema de administración del modelo del DSS, este comprende la base de modelo, el sistema de administración de la base de modelo, el lenguaje de modelación, el directorio del modelo, y el procesador de comandos, integración y ejecución del modelo.
- El subsistema de interfase de usuario incluye no sólo el hardware y el software, sino también factores involucrados con la facilidad de uso, accesibilidad, e interacciones entre el usuario y la máquina.

- El usuario es la persona que debe tomar la decisión que pretende ser soportada por el DSS. Un DSS tiene dos clases de usuarios: los gerentes y los especialistas de staff. Generalmente, los gerentes esperan una interfase más amigable que aquella esperada por los especialistas de staff ya que estos últimos son más detallistas y están dispuestos a utilizar sistemas más complejos.

Sistemas más complejos adaptan otros componentes como el subsistema de administración del conocimiento, así como también módulos hechos a la medida para la resolución de problemas específicos.

11.2. Tipos de Sistemas de Soporte a la Decisión

Actualmente existen tecnologías que forman parte de los sistemas de soporte administrativo, las cuales solo se listarán, debido a que cada tema puede ser el inicio de un nuevo artículo (Ramos, 2006)

- Management Science (MS)
- Enterprise Resource Planning (ERP)
- Business Analytics
- Customer Resource Management (CRM)
- Data Mining Supply Chain Management (SCM)
- Data warehousing
- Knowledge Management System (KMS)
- Business Intelligence
- Expert Systems (ES)
- Online Analytical Processing
- Artificial Neural Networks (ANN)
- Computer Assisted Engineering
- Intelligent Agents
- Group Support Systems
- Electronic Commerce DSS
- Enterprise Information Systems
- Enterprise Resource Management (ERM)

11.3. Diferencias con otras herramientas de Business Intelligence.

Podemos destacar como principal objetivo de los Sistemas de Soporte a Decisiones, explotar al máximo la información residente en una base de datos corporativa, a diferencia de otras herramientas como los Cuadros de Mando (CMI) o los Sistemas de Información Ejecutiva (EIS), ya que proporcionan informes muy dinámicos y con gran potencial a nivel de información de datos, pero siempre con una interfaz gráfica amigable y sencilla.

Otra diferencia importante radica en los usuarios a los que están destinadas las plataformas DSS: cualquier nivel gerencial dentro de una organización, tanto para situaciones estructuradas como no estructuradas.

Por último, destacar que los DSS suelen requerir de un motor OLAP como base, que facilite el análisis casi ilimitado de los datos para hallar las causas raíz de los problemas de la organización (Sinergia e Inteligencia de Negocio S.L., 2023).

12. Análisis y definición de las herramientas a utilizar

En este estudio, se ha realizado un análisis profundo sobre las herramientas existentes en el mercado. Dado que es una institución gubernamental, los datos deben estar a disposición de manera inmediata, esto se traduce a alta disponibilidad, del mismo modo, debe ser escalable en el tiempo. Se debe tener en cuenta, como esta institución puede crecer en gran manera, los datos a ingresar, ya que se trata de movimientos de personas que se realizan todos los días, sobre cada punto fronterizo, por lo que se busca que pueda efectuar análisis de datos en poco tiempo y ofreciendo los resultados más versátiles posibles para posteriormente, se efectúen las tomas de decisiones.

Hoy en día, los datos están en el centro de cualquier negocio. Las empresas precisan de datos confiables para poder tomar decisiones correctas en el menor tiempo posible y así lograr mantenerse a la vanguardia y poder ser competitivas en el mercado.

Sin embargo, poder obtener valor o información de los datos que sea precisa para luego tomar decisiones, es una tarea cada vez más compleja. La proliferación de fuentes de datos distintas y heterogéneas, el crecimiento de su volumen, los tipos de datos complejos, la rápida adopción de entornos de nube pública, hacen que la integración sea un proceso importante.

La integración de datos hace posible gestionar la complejidad en un único dominio, el de la tecnología de integración, dejando los datos libres de ser utilizados por las aplicaciones de negocio.

12.1. Herramientas ETL

Las empresas que manejan grandes volúmenes de datos con el fin de convertirlos en información significativa para reutilizarla en operaciones o simplemente para toma de decisiones, exigencias operacionales, análisis, grandes extracciones, así como transformaciones y carga de datos, requieren elegir la herramienta correcta.

A medida que las empresas se acercan y son participes de la transformación digital, deben consolidar datos provenientes de múltiples fuentes para realizar un repositorio de datos centralizado y poder así, garantizar una única fuente de información.

Tener los datos, veraces, recopilarlos y almacenarlos de manera segura y organizada es sumamente importante para obtener información oportuna basada en ellos. Es debido a esto que las herramientas ETL se han vuelto indispensables para las organizaciones con visión de futuro, con la meta y objetivo de crecer. Un ETL permite a las organizaciones basadas en datos, extraer datos de fuentes dispares, transformarlos y cargarlos en sistemas de destino, como un lago de datos o un almacén de datos, para informes y análisis.

Se describen sus siglas y contexto de la siguiente forma (Díaz, 2016):

- Extracción

El primer paso del proceso ETL es la extracción de datos.

Las empresas reciben datos de varias fuentes, sistemas de archivos, bases de datos y aplicaciones en la nube. Estos datos están disponibles en diferentes tipos de formatos, como hojas de cálculo, archivos de texto, XML/JSON, PDF y más.

Esto representa un desafío dentro de las empresas con aplicativos legados, ya que los datos no estructurados son difíciles de procesar y analizar. Afortunadamente, las soluciones ETL actuales permiten a las organizaciones extraer sin problemas datos estructurados, semiestructurados y no estructurados de múltiples fuentes de datos.

- Transformación

El segundo paso del proceso ETL es transformación de datos.

Los datos extraídos de diferentes tipos de fuentes no siempre cumplen con un estándar y, a menudo, están plagados de calidad de los datos asuntos. Además, está en un formato normalizado con uniones que difícilmente se pueden analizar, lo que puede obstaculizar la eficiencia de la infraestructura de la base de datos.

El proceso de transformación conlleva limpiar, estandarizar y validar los datos, mejorando así su calidad. Este paso garantiza que los datos consolidados sean precisos, completos y valiosos para informes y análisis antes de que lleguen a su destino final.

- Carga

El tercer paso del proceso ETL es carga de datos.

Este es el último paso, donde los datos que ya han sido transformados se cargan en un almacén de datos, una base de datos o a su destino. Dependiendo del volumen, puede cargar datos en un almacén de datos en dos maneras:

- Carga de tipo completa: también conocida como carga destructiva, implica cargar todos los datos de origen en el destino de una sola vez. En este enfoque, una tabla de destino se trunca antes de cargar los datos y se reemplaza por completo con conjuntos de datos actualizados.
- Carga incremental: una carga incremental implica solo actualizar de manera paulatina los nuevos conjuntos de datos. Este enfoque es más eficiente ya que reducen los recursos informáticos y el tiempo que se necesita para cargar datos, lo que ayuda a procesar la información en tiempo real.

12.1.1. Microsoft SSIS

SQL Server Integration Services (SSIS) es una herramienta cuyo dueño es la empresa Microsoft, utilizada para realizar tareas de integración de datos. SSIS permite el manejo de datos, de la siguiente manera, ofrece desarrollar y administrar paquetes que extraen, transforman y cargan datos de diversas fuentes a destinos específicos.

SSIS permite crear flujos de trabajo visuales llamados paquetes, en los cuales se especifican las tareas de extracción, transformación y carga de datos. Los anteriormente denominados paquetes, se pueden ejecutar de manera programada o manual, y ofrecen una amplia variedad de opciones y componentes para manipular datos de forma eficiente (Cote, Lah, & Sarka, 2017).

Algunas de las características principales de SSIS incluyen:

1. Conectividad: SSIS ofrece conectividad con una gran fuente de fuentes de datos, como bases de datos SQL Server, Oracle, Excel, archivos planos, servicios web y muchas más.

2. Transformaciones de datos: SSIS brinda una cantidad superior de transformaciones que permiten limpiar, filtrar, combinar, agregar y enriquecer datos durante el proceso de ETL.
3. Flujo de control: SSIS ofrece componentes para manejar el proceso de ejecución de los paquetes, como bucles, tomas de decisiones y tareas condicionales.
4. Programabilidad: SSIS permite personalizar el código, pudiendo escribir ya sea, en lenguaje SQL, C# o Visual Basic para realizar operaciones avanzadas de manipulación de datos o proporcionando una capa personalizada de la lógica del paquete.
5. Programación y automatización: SSIS proporciona la versatilidad de poder programarse para ejecutarse de forma programada o se puede integrar con otras aplicaciones utilizando API y servicios web.
6. Monitoreo y administración: SSIS proporciona herramientas para monitorear y administrar los paquetes en ejecución, así como para realizar tareas de depuración y registro de eventos.

Se puede concluir que SSIS es una herramienta versátil para la integración de datos en el entorno de Microsoft SQL Server. Permite realizar actividades complejas de ETL de manera eficiente y brinda un entorno visual para diseñar, desarrollar y administrar los flujos de trabajo de datos.

12.1.2. Pentaho Kettle

Pentaho es una plataforma de Business Intelligence (BI) cuyo enfoque es la solución y de forma centrada en procesos que incluye los componentes requeridos para implementar soluciones basadas en procesos como minería de datos, ETL o informes.

En la actualidad, muchas empresas han decidido utilizar Pentaho como herramienta BI para la mejora de la capacidad de análisis y de toma de decisiones, debido a que ofrece soluciones que se componen fundamentalmente de una infraestructura de herramientas de análisis e informes integrado con un motor de workflow de procesos de negocio (Pentaho Corporation, 2008).

Algunas características clave de Pentaho incluyen:

1. Integración de datos: Pentaho brinda herramientas de ETL que facilitan extraer, transformar y cargar datos de múltiples fuentes hacia destinos específicos. Estas herramientas ayudan a limpiar, filtrar y transformar datos para su posterior análisis.
2. Análisis de datos: Pentaho contiene herramientas para el análisis y la exploración de datos, permitiendo a los usuarios descubrir patrones, tendencias y relaciones dentro de los conjuntos de datos. Ofrece capacidades de generación de informes y paneles interactivos para la presentación de los resultados del análisis.
3. Minería de datos: Pentaho ofrece funciones de minería de datos que permiten descubrir información oculta en grandes volúmenes de datos. Estas funciones pueden utilizarse para análisis predictivos, detección de patrones y segmentación de datos.
4. Integración con Big Data: Pentaho es compatible con el procesamiento y análisis de datos en entornos de Big Data, como Hadoop. Permite la integración y el análisis de datos estructurados y no estructurados provenientes de diferentes fuentes.
5. Visualización de datos: Pentaho ofrece capacidades de visualización de datos que permiten crear gráficos, tablas y paneles interactivos para presentar datos de manera comprensible y atractiva.

Pentaho es utilizado en el ámbito empresarial y a nivel corporativo, debido a que brinda ayuda para la toma de decisiones basada en datos y la generación de informes.

Al ser una plataforma de código abierto, ofrece flexibilidad y personalización, lo que la hace accesible para organizaciones de diferentes tamaños y necesidades (Pentaho Corporation, 2006).

Dentro de la herramienta Kettle Pentaho se pueden diferenciar cuatro grandes grupos de procesos de trabajo:

- **Reporting:** este es el módulo de Kettle Pentaho está destinado a realizar informes, estos se adaptan a las distintas necesidades del usuario final. Los informes, pueden ser exportados fácilmente a pdf, xls, HTML o texto. Además, pueden automatizarse procesos de generación de informes según la ejecución de diferentes acciones o mediante una periodicidad estipulada.

- **Análisis:** por medio de tablas dinámicas, el módulo de analítica de Kettle Pentaho permite al usuario navegar por los datos extraídos, ajustando los filtros, campo, entre otros. Estos datos puros, pueden ser extraídos a su vez en formato SVG, Flash e incluso Excel para su tratamiento fuera de la herramienta.
- **Dashboards:** ofrece la posibilidad de crear dashboards personalizados en base a las necesidades de la empresa u organización, ya que todos los módulos de la herramienta pueden integrarse dentro de estos dashboards. De esta manera se pueden visualizar gráficos, tablas o cualquier otro dato que se desee.
- **Integración de datos:** este módulo trabaja mediante ETL y juega un papel crucial dentro del entramado sistema de Business Intelligence de la empresa. Este proceso es el responsable de aprovisionar los datos sobre los que posteriormente se trabajarán.

12.1.3. Informática Powercenter.

Informática Power Center es una herramienta basada en una arquitectura ETL. La parte que lo componen ayudan a extraer los datos de distintas fuentes, transformarlos de una forma unificada y coherente según se estipulen en los requisitos del negocio y finalmente cargarlos en un almacén de datos de destino, típicamente un Data Warehouse o Hadoop. Además, Power Center brinda una solución completa de integración de datos y un sistema de gestión de datos.

Power Center facilita llevar a cabo diversas tareas, como la limpieza de datos, creación de perfiles de datos, y transformación y programación de flujos de trabajo de origen a destino. Informática Power Center proporciona una de las mejores soluciones de integración de datos disponible actualmente, ya que dentro de sus beneficios es que puede comunicarse con cualquier base de datos y ofrece una poderosa herramienta de transformación de datos.

En este sentido, Informática PowerCenter ofrece una plataforma única de integración de datos de extremo a extremo, con un amplio conjunto de capacidades para integrar datos brutos y fragmentados de diferentes fuentes y transformarlos en información completa, de alta calidad y preparada para el negocio (Belhaj, 2016).

Entre los beneficios más destacados de Informática Powercenter podemos destacar (Informatica LLC., 2018):

1. Proporciona a los perfiles de negocio la información correcta en el momento adecuado.
2. Ayuda a encontrar soluciones a problemas de negocios tales como el abandono de clientes, la efectividad de las campañas, la detección de fraude y la mejora continua de ventas.
3. Agiliza los proyectos gracias a una mejor productividad y colaboración del personal.
4. Proporciona la integración de datos que vienen de distintos sistemas en una base de datos coherente.
5. Permite escalar del mismo modo en que crecen las necesidades del negocio y se requiere más visibilidad sobre los datos.
6. Ayuda a que los equipos de TI y de negocio trabajen juntos, mejor y más rápidamente.

7. Garantiza el éxito continuo de los procesos críticos de negocio.
8. Minimiza los costes gracias a la reutilización de habilidades y recursos en todos los proyectos.

12.1.4. Oracle Data Integrator

Oracle Data Integrator (ODI) es una plataforma integral de integración de datos desarrollada por Oracle Corporation. Proporciona un conjunto de herramientas y tecnologías para extraer, transformar y cargar (ETL) datos en diferentes sistemas y plataformas.

ODI ofrece una solución unificada para la construcción, el despliegue y la administración de almacenes de datos enfocado para entornos de Business Intelligence. Además, se caracteriza por combinar todos los elementos de la integración de datos (movimiento, sincronización, calidad, administración y servicios de datos) para asegurar que la data esté disponible de forma precisa y consistente en sistemas complejos (Rios, 2009).

La principal característica de la arquitectura ODI es el repositorio. El repositorio de ODI puede almacenar información de configuración relacionada con metadatos de la aplicación, proyectos, escenarios y logs de ejecución. Ofrece la posibilidad de contar con varias instancias del repositorio, esto permite tener entornos separados (Hotman, 2010). Además, el repositorio actúa como un sistema de control de versiones a través de un número de versión para los objetos creados.

ODI proporciona una interfaz gráfica que permite a los usuarios poder diseñar y administrar los procesos de integración de datos. Ofrece a su vez, que diversas fuentes y destinos de datos interactúen, por ejemplo, que bases de datos, sistemas de archivos, servicios web y más puedan comunicarse fácilmente (González, 2013). Las características clave de Oracle Data Integrator incluyen:

1. Integración de datos: ODI ofrece a los usuarios definir procesos de integración de datos mediante una sintaxis declarativa.
2. Diseño visual de flujo de datos: ODI permite un entorno gráfico intuitivo, en otras palabras, fácil de usar, para crear flujos de datos que incluyen transformaciones, filtrado y mapeo de datos.

3. Conectividad flexible: ODI acepta una amplia gama de fuentes y destinos de datos, por lo que es posible interactuar con bases de datos Oracle y no Oracle, sistemas de archivos, aplicaciones empresariales, servicios web y más.
4. Transformaciones y limpieza de datos: ODI brinda una variedad de transformaciones y de limpieza de datos, lo que garantiza la calidad y consistencia de los datos integrados.
5. Planificación y programación de trabajos: ODI permite programar y ejecutar procesos de integración de datos de manera automatizada, según los requerimientos definidos.
6. Supervisión y administración: ODI proporciona herramientas para la supervisión y administración lo que facilita monitorear y gestionar los procesos de integración de datos, así como para realizar ajustes y mejoras.

Oracle Data Integrator proporciona una solución completa de integración de datos que brinda a las organizaciones extraer, transformar y cargar datos de diversas fuentes a diferentes destinos, permitiendo crear un entorno visual y herramientas para el diseño, programación y administración de procesos de integración de datos.

12.1.5. Talend

Talend es una suite que brinda un grupo muy completo de herramientas para llevar a cabo la integración de datos que se ofrece en una versión de código libre o mejor conocido como open source. Es por ello, que esta es una de las herramientas de integración ETL más utilizadas dentro del mundo Big Data; de hecho, es la cuarta en la lista después de Informática Powercenter, IBM InfoSphere Datastage y Oracle Data Integrator (ODI) (Barton, 2013).

Por otra parte, esta suite cuenta con un Community Edition (CE) totalmente funcional. Además, es posible utilizar una gran cantidad de componentes para llevar a cabo una administración de datos personalizada. De hecho, TOS (Talend Open Studio) permite tareas complicadas de manera sencilla gracias a esta variedad de servicios.

Talend es Open Source, como se ha mencionado anteriormente, lo que significa que se puede conseguir esta herramienta de forma rápida y sencilla sin necesidad de registrarse a través de la web de Talend Open Studio o destinar un presupuesto exclusivo para la compra de licencias. Talend es más que un entorno de desarrollo de aplicaciones informáticas, es una plataforma de Data Integration (ETL), que administra e implementa procesos en las empresas, lo que te proporcionará una ventaja competitiva (Talend, 2020). Procesos como Análisis Avanzado y toma de decisiones, permite a su vez la implementación de Inteligencia Artificial a los procesos existentes o a nuevos, ofrece la cualidad de mejora de procesos de ecommerce, procesos de marketing digital enfocado al público objetivo, etc.

Talend permite implementar todos los procesos de una manera sencilla, convirtiendo tu empresa en una empresa Data Driven, siendo impulsada por la importancia de los datos internos y externos que se manejan y utilizan. Entre sus características principales tenemos:

1. Interfaz gráfica de diseño: Talend ofrece una interfaz visual que basada en componentes que permite a los usuarios crear flujos de trabajo de integración de datos arrastrando y soltando componentes predefinidos lo que facilita en gran manera la interacción con el usuario ya que permite la creación y configuración de procesos de ETL sin necesidad de escribir código manualmente.
2. Conectividad y compatibilidad: brinda una compatibilidad con una amplia gama de fuentes y destinos de datos, incluyendo bases de datos relacionales, sistemas CRM, aplicaciones ERP, servicios web, archivos planos, servicios de nube, entre otros. Además, admite una variedad de estándares y protocolos de conectividad como JDBC, ODBC, FTP, HTTP, SOAP, entre otros (Barton, 2013).
3. Transformaciones de datos: brinda una amplia gama de transformaciones y operaciones para la manipulación y limpieza de datos. Se pueden realizar transformaciones como filtrado, mapeo, clasificación, duplicidad, enriquecimiento de datos, validación, entre otros. También ofrece herramientas para la gestión de metadatos y la definición de reglas de negocio.
4. Orquestación y programación de tareas: su plataforma permite programar de ETL de manera eficiente. Se pueden establecer cronogramas para ejecutar flujos de trabajo de forma automática, así como gestionar la dependencia y el orden de ejecución de las tareas.

5. Control de calidad de datos: Talend ofrece funcionalidades que garantizan la calidad de los datos, por ejemplo, realizar validaciones, correcciones y enriquecimiento de datos, así como implementar reglas de control de calidad y limpieza de datos ofreciendo integridad y la consistencia de los datos.
6. Monitorización y gestión: proporciona herramientas para monitorizar y gestionar los flujos de trabajo de integración de datos. Se puede obtener información en tiempo real sobre el rendimiento o el estado de la ejecución y los errores, así como también, generar informes y registros detallados para su posterior análisis y auditoría.
7. Escalabilidad y rendimiento: este aspecto es una de las mayores ventajas de Talend, ya que está diseñado para manejar volúmenes de datos grandes y poder trabajar con operaciones complejas de manera eficiente. La plataforma proporciona opciones de escalabilidad horizontal y vertical, así como formas de optimización de rendimiento para garantizar un procesamiento rápido y eficiente de los datos.

Estas son solo algunas de las características principales de Talend. La plataforma ofrece una amplia gama de funcionalidades adicionales y capacidades personalizables para satisfacer las necesidades específicas de integración de datos de cada organización.

12.2. Herramientas de análisis de datos.

Las herramientas de Business Intelligence ayudan a tomar decisiones basadas en el análisis de datos.

Cuando se enfrenta a una solución BI, una de las primeras cuestiones que se deben plantear es la herramienta que se ajusta a las necesidades y la cual se debe utilizar. Esto puede ser determinante en materia de éxito o fracaso en un proyecto.

Es importante intentar adaptar la necesidad a la herramienta y no al revés, ya que se puede llevar a complicaciones futuras que alguien tiene que asumir, problemas que se pretenden minimizar o eliminar por completo. Por lo que se han seleccionado herramientas disponibles en el mercado que ofrecen una gran variedad de beneficios a las organizaciones que los han implementado y a partir de este análisis, elegir la herramienta que se acopla a las necesidades de la Dirección General de Migración y Extranjería.

En las necesidades de la DGME, está, realizar análisis en tiempo real sobre los datos que son proporcionados desde diferentes puntos, como fronteras, diferentes agencias de migración en el país y entes externos también. Estos datos deberán ser agrupados para la generación de reportes.

Se busca que diferentes usuarios puedan acceder a la herramienta, no obstante, no se pretende que todos los usuarios que ingresen vean el mismo reporte, ya que, en base a diferentes reglas de la organización, los grupos de trabajo se dividen tareas, diferentes reportes serán consultados de forma simultánea, es por ello, que los datos que se almacenen deben ser veraces y funcionales, para una mejor toma de decisiones de negocios.

12.2.1. Power BI

Microsoft Power BI es una herramienta de Business Intelligence y visualización de datos para empresas. Ofrece a los usuarios no especializados herramientas para agregar, analizar, visualizar y compartir datos.

La interfaz de usuario de Microsoft Power BI es bastante intuitiva para aquellos que estén familiarizados con herramientas de la empresa Microsoft, en especial, Excel. Su integración con otros productos de Microsoft la convierte en una herramienta de autoservicio muy útil que requiere poca capacitación inicial.

Microsoft Power BI dispone de dos opciones de su producto, por un lado, una versión gratuita y por el otro de pago, dependiendo de las funcionalidades que se quieran adquirir. La versión gratuita está destinada a propietarios de pequeñas y medianas empresas. Por otro lado, la versión profesional llamada Power BI Plus está disponible por una tarifa de suscripción mensual.

Una gran ventaja de Power BI, es que es posible instalarlo tanto en computadoras personales como en dispositivos móviles, Tablet, Celulares entre otros. Esto lleva a poder tener acceso a la información desde prácticamente, cualquier dispositivo con conexión a la organización, datos que pueden ser necesarios sin tener acceso a una computadora para toma de decisiones de último minuto. También existe Power BI Report Server para empresas que deben mantener sus datos e informes en las instalaciones.

La principal función de la plataforma es ayudar a las empresas u organizaciones a encontrar información útil entre los datos de sus bases de datos. En este sentido, Microsoft Power BI se concentra en conectar conjuntos de datos dispares. También se encarga de transformar y limpiar los datos creando cuadros o gráficos que ofrecen imágenes o gráficos. Todo esto se puede compartir con otros usuarios de Power BI dentro de la organización (Russo & Ferrari, 2016), parte importante para la DGME.

Los modelos de datos creados a partir de Power BI se pueden usar de distintas formas especialmente, para contar historias a través de gráficos y visualizaciones de datos y examinar escenarios hipotéticos dentro de los datos. También proporciona informes que pueden responder preguntas a tiempo real y ayudar con la previsión para asegurarse de que los departamentos cumplan con las métricas comerciales.

Power BI también proporciona paneles ejecutivos para administradores o gerentes. Esto brinda a la administración más información sobre el desempeño de los departamentos.

Por otro lado, ayuda a los usuarios a ver no solo lo que sucedió en el pasado y lo que está sucediendo en el presente, sino también lo que podría suceder en el futuro. Ha sido creado mediante Machine Learning, lo que significa que puede detectar patrones en los datos y usar esos patrones para realizar predicciones informadas y ejecutar escenarios hipotéticos. Estas estimaciones permiten a los usuarios generar pronósticos y prepararse para satisfacer la demanda futura y otras métricas clave (Polanco & Betancourt, 2022).

Entre las características principales de power BI tenemos:

1. Panel de control: es el espacio de trabajo principal donde se organizan y se muestran los informes y visualizaciones.
2. Informe: Es un documento interactivo que muestra datos y visualizaciones de una manera estructurada y significativa.
3. Conjunto de datos: es una colección de tablas, consultas y conexiones que se utilizan para crear informes en Power BI.
4. Visualización: es una representación gráfica de los datos, como gráficos, tablas, mapas y otros elementos visuales.
5. Cuadro de mando: es una página o vista en un informe que muestra una colección de visualizaciones relacionadas.
6. Filtro: es una herramienta que permite filtrar los datos en una visualización o informe según criterios específicos.

7. Gráfico de barras: es una visualización que muestra la comparación de datos a través de barras rectangulares.
8. Gráfico circular: es una visualización circular que muestra la proporción de diferentes categorías dentro de un conjunto de datos.
9. Gráfico de líneas: es una visualización que muestra la tendencia o el cambio de datos a lo largo del tiempo mediante una línea que conecta los puntos de datos.
10. Tabla dinámica: es una visualización interactiva que permite explorar y analizar datos mediante la segmentación y el filtrado.
11. Segmentación: es una función que permite filtrar visualizaciones y datos seleccionando valores específicos de una lista desplegable.
12. Medida: es una expresión o fórmula utilizada para realizar cálculos en Power BI, como sumas, promedios o conteos.
13. Panel de navegación: es un área en Power BI que muestra las páginas y las visualizaciones disponibles en un informe.
14. Conexión de datos: es la configuración que permite importar datos desde diversas fuentes, como archivos de Excel, bases de datos o servicios en la nube.

En resumen, Power BI es una herramienta que reúne los datos y los procesa, convirtiéndolos en información inteligible. A menudo, utilizando gráficos atractivos y fáciles de procesar. Esto permite a los usuarios generar y compartir instantáneas claras y útiles de lo que está sucediendo en la empresa. Además, se puede conectar a una gran variedad de fuentes de datos, desde hojas de cálculo de Excel básicas hasta bases de datos y aplicaciones en la nube.

12.2.2. Tableau.

La plataforma de Tableau es una opción de inteligencia de negocios moderna. Permite que sea más fácil explorar y administrar los datos. Asimismo, ofrece descubrir y compartir información más rápidamente a fin de generar grandes cambios en los negocios.

Su principal objetivo es de ayudar a las personas a ver y comprender los datos (Tableau Software LLC., 2023). Y esto lo consigue ofreciendo a los usuarios toda una selección de herramientas útiles e intuitivas de inteligencia de negocios. Destaca por su facilidad para integrar datos de diferentes orígenes y su sencillez de uso, que permite realizar un análisis ágil y rápido.

Tableau permite análisis avanzados en tiempos reducidos, generando visualizaciones y demostraciones impactantes sin necesidad de pasar por desarrollos complejos, contando con una configuración muy flexible, porque puede funcionar bajo un servidor, de forma local en el equipo de los usuarios o en la nube.

El acceso a los datos es instantáneo, no importando de los orígenes de información que disponga la empresa. Tableau permite integrar toda la información de la empresa en un único modelo, de forma que se puede analizar al mismo tiempo datos de sistemas ERP como JDE o SAP, base de datos Oracle, SQL Server, hojas de cálculo, etc.

Además, el propio usuario es el encargado de gestionar y distribuir sus análisis, lo que proporciona a los usuarios de negocio total independencia del departamento de IT evitando de esta forma los errores y retrasos que eso conlleva lo que permite que el proceso de toma de decisiones sea más rápido y ágil.

Para la DGME, como objetivo, al igual que otras empresas que usan una herramienta de BI es que sea escalable y Tableau lo es, tanto en hardware como en memoria. Otro elemento clave para tener en cuenta es la seguridad. Tableau permite configurar la seguridad, lo que permite adaptarse a las necesidades de cada usuario logrando así, una configuración individualizada de proyectos, informes y usuarios, protegiendo de esta forma la información confidencial existente en determinados niveles de la compañía.

A continuación, algunas de las principales características de Tableau:

1. Visualización interactiva: con Tableau es posible crear visualizaciones interactivas y atractivas, lo que facilita la exploración y comprensión de los datos. Se puede utilizar una variedad de gráficos, tablas, mapas y otros elementos visuales para mostrar datos.
2. Conexión a múltiples fuentes de datos: permite la opción de conectarse y combinar datos de diversas fuentes, como bases de datos, archivos de Excel, servicios en la nube entre otras fuentes de datos.
3. Arrastrar y soltar: la interfaz intuitiva de Tableau, no es necesario escribir código ya que permite arrastrar y soltar dimensiones y medidas para crear visualizaciones rápidamente.

4. Análisis en tiempo real: Tableau es capaz de hacer análisis en tiempo real sobre datos en constante actualización. Se puede conectar a fuentes de datos en tiempo real y ver los cambios reflejados de forma instantánea.
5. Funciones de cálculo y expresiones: Tableau proporciona una amplia gama de funciones de cálculo y expresiones que te permiten realizar operaciones avanzadas y personalizadas sobre los datos, como crear medidas personalizadas, calculadas y campos calculados para realizar análisis complejos.
6. Interacción y filtrado: los usuarios pueden interactuar con las visualizaciones y explorar los datos de forma dinámica, aplicar filtros, realizar drill-downs, resaltar elementos en específico y aplicar una segmentación para profundizar en los detalles de los datos.
7. Publicación y colaboración: brinda opciones para publicar tus visualizaciones en línea y compartirlas con diferentes usuarios. Se puede colaborar en proyectos, crear paneles de control interactivos y compartir informes y dashboards con diversos clientes.
8. Capacidades geoespaciales: cuenta con herramientas para trabajar con datos geoespaciales y mapas. Se puede crear visualizaciones basadas en ubicaciones geográficas, utilizar datos de mapas y realizar análisis espaciales.
9. Integración con otras herramientas: Tableau también se puede integra con otras herramientas y plataformas populares, como Excel, SQL Server, Python, R y muchos más. Esto permite combinar las capacidades de Tableau con otras herramientas que utilices en tu flujo de trabajo de análisis de datos.

12.2.3. Qlik Sense

Qlik Sense es una herramienta de descubrimiento de datos diseñada para visualizar diferentes fuentes de datos, cuadros de mando interactivos, gráficos personalizados e informes. Es utilizada porque los usuarios pueden visualizar datos fácilmente y obtener información valiosa para la toma de decisiones de negocio.

Qlik Sense hace uso de una tecnología asociada de datos, lo que permite a los usuarios explorar diversos datos y descubrir relaciones ocultas entre ellos. Además, ofrece también herramientas de colaboración para que los usuarios puedan compartir sus análisis y tomar así decisiones en equipo o en toda la empresa (Neeraj, Hand, & Kharpaté, 2018).

Esta herramienta de inteligencia empresarial es versátil y a la vez poderosa ya que puede aplicarse en una amplia variedad de sectores y aplicaciones como: Ventas, Marketing, RRHH, Finanzas y mucho más. A continuación, algunas características clave de Qlik Sense:

1. Visualización de datos: Qlik Sense ofrece una amplia variedad de opciones de visualización, como gráficos, tablas, mapas interactivos y cuadros de mando. Permite crear visualizaciones personalizadas y dinámicas para explorar y comprender los datos.
2. Arrastrar y soltar: la interfaz de Qlik Sense es intuitiva y basada en arrastrar y soltar. Los usuarios pueden seleccionar dimensiones y medidas, y arrastrarlas a la interfaz para crear visualizaciones y análisis de manera rápida y sencilla, sin necesidad de programación.
3. Exploración y descubrimiento de datos: Qlik Sense permite a los usuarios explorar datos de manera libre y descubrir nuevas relaciones y patrones. Pueden realizar búsquedas interactivas, hacer drill-down, filtrar y seleccionar datos para profundizar en el análisis.
4. Análisis colaborativo: Qlik Sense fomenta la colaboración entre los usuarios. Permite compartir visualizaciones, paneles de control y análisis con otros miembros del equipo. También ofrece capacidades de anotación y comentarios para facilitar la colaboración y el intercambio de ideas.
5. Integración de datos: Qlik Sense permite la conexión con diversas fuentes de datos, como bases de datos, hojas de cálculo, servicios web y más. Puede integrarse con sistemas existentes y realizar una carga de datos automatizada para mantener la información actualizada.
6. Capacidad de autoservicio: Qlik Sense está diseñado para el autoservicio de análisis, lo que significa que los usuarios pueden crear y modificar sus propias visualizaciones y análisis sin depender del equipo de TI. Esto proporciona a los usuarios más autonomía y agilidad en el análisis de datos.
7. Seguridad y gobernanza: Qlik Sense ofrece funciones de seguridad y gobernanza robustas. Los administradores pueden establecer políticas de acceso, controlar los permisos de los usuarios y administrar el acceso a los datos y las visualizaciones.

12.3. Comparación de Herramientas

Para ofrecer un producto de calidad que se acople en gran manera a las necesidades, es primordial realizar un análisis previo para seleccionar, la herramienta que brindará la oportunidad de solventar el objetivo de la investigación.

12.3.1. Herramienta de migración de datos

En esta investigación, se ha realizado un análisis de las herramientas más populares en el mercado, entre ellos tenemos:

1. Pentaho
2. Oracle Data Integrator
3. Microsoft SSIS
4. Informática Power Center.
5. Talend

Para realizar una comparación, es importante conocer las ventajas y desventajas que posee cada una de las herramientas, ya que se necesita que el producto seleccionado sea escalable, tolerante a fallos, que pueda trabajar con diversas fuentes de datos

	Talend	Pentaho	Powercenter	Oracle ODI	Server Integrator
Costo	Green		Red		Yellow
Riegos	Green		Yellow		
Facilidad de Uso	Green				
Soporte	Green				
Implementación	Yellow				
Monitoreo	Green				
Conectividad	Green		Red		Yellow

Tabla 1. Comparativa de ETLs en el mercado actual. Fuente: Elaboración propia.

Para la solución a los procesos de extracción, transformación y carga de datos, se ha seleccionado a Pentaho, como la herramienta a utilizar, debido a los beneficios que ofrece, entre ellos la facilidad de uso y la implementación, características necesarias para poder llevar a cabo la solución propuesta en la problemática de esta investigación. Cabe mencionar que Pentaho es Open Source, lo que indica que no hay licenciamiento, por lo que, dentro de la puesta en marcha, no hay que destinar recurso económico para la compra de este software.

12.3.2. Herramienta para visualización de datos

Los datos que se cargarán por medio de Pentaho deben ser presentados a los usuarios de una forma que les sea fácil comprender la importancia de la información que la data posee. Todos los datos guardados deben ser presentados como un conjunto, ayudado de herramientas como gráficos o informes por medio de los cuales la toma de decisión sea precisa.

Entre los presentados en esta investigación tenemos:

1. Power BI
2. Tableau
3. Qlik Sense

Entre estas herramientas podemos realizar una tabla comparativa para poder tomar una mejor decisión sobre que software debe ser seleccionado por la investigación para llevar a cabo la solución para la DMGE.

Se han tomado las características que la organización necesita en su día a día para poder realizar la comparativa. Costo, Facilidad de uso e inclusive la implementación, son cualidades a las que se le debe tener cuidado al momento de seleccionar el software.

	Tableau	PowerBI	QLik Sense
Costo	Yellow	Green	Yellow
Implementación	Green	Red	Yellow
Facilidad de Uso	Green		
Transformación	Yellow		
Modelado	Green		
Visualización	Yellow		
Reportería	Yellow		
Aplicación Móvil	Green		

Tabla 2. Comparativa de herramientas para la visualización y análisis de datos. Fuente:

Elaboración propia.

Se ha elegido a Tableau como el producto para implementar en este proyecto, debido a las facilidades que ofrece, documentación sólida y extensa y referencias positivas de este producto. Se pretende que los usuarios se les facilite aprovechar la herramienta y todos los beneficios que se obtiene a partir de las consultas que se hagan, y poder tomar las mejores decisiones con ayuda de reportes, diagramas, gráficos.

13. Análisis y definición de la infraestructura de la solución

En la actualidad, la Dirección General de Migración y Extranjería (DGME) dispone de un servidor destinado al almacenamiento de sus bases de datos y sistemas. Para la implementación de esta solución, se propone utilizar el mismo servidor. Es importante destacar que se recomienda incorporar las nuevas estructuras en los planes actuales de respaldo de datos. En la Figura 2 se presenta un diagrama que ilustra la arquitectura definitiva de la solución.

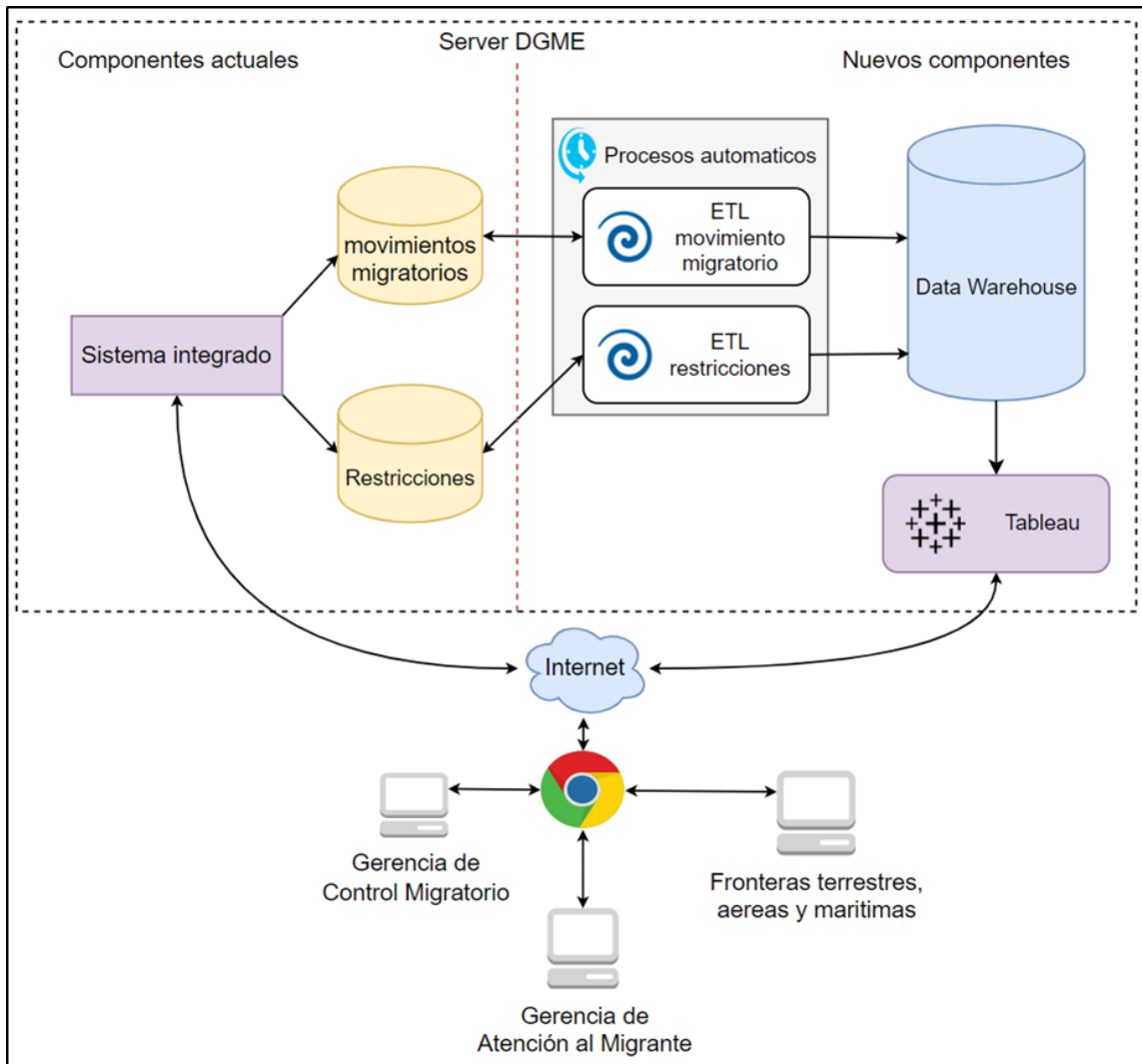


Ilustración 11. Diagrama de la infraestructura de la solución. Fuente: Elaboración propia.

La solución planteada en este documento abarca la sección identificada como "nuevos componentes" en la Imagen 2. No obstante, también se ha incluido la disposición actual de la infraestructura con el fin de comprender mejor cómo se relacionará la solución con los componentes existentes.

En la actualidad, los clientes, como la Gerencia de Control Migratorio y la Gerencia de Atención al Migrante y Fronteras, entre otros, acceden al "sistema integrado" de la DGME a través de internet. Este sistema tiene diversos propósitos, entre los cuales se encuentran el registro de movimientos migratorios y el registro de restricciones. Hasta este punto, el flujo de datos sigue su curso de manera convencional.

La solución propuesta se pone en práctica mediante los procesos ETL, los cuales se ejecutarán en momentos específicos del día que deberán especificarse en un sistema de ejecución de procesos programados, como el Administrador de Tareas de Windows o el sistema cron de Linux.

Estos procesos ETL se encargarán de mantener actualizado el Data Warehouse en la medida de lo posible con relación a las bases de datos de origen. Para determinar la frecuencia con la que se deben ejecutar los procesos ETL, es necesario realizar pruebas que nos permitan evaluar la carga del proceso en función del volumen de datos y el tiempo necesario para completar la carga en el Data Warehouse. En función de estos resultados, se podrá determinar si es posible reducir el tiempo para tener los datos lo más actualizados posible.

Por último, se proporcionará a los clientes una nueva herramienta para generar informes. La idea de utilizar una herramienta como Tableau es permitir que los usuarios no técnicos puedan acceder de manera más sencilla a los datos, es decir, no se requiere conocimientos en consultas SQL para obtener un informe. Sin embargo, será necesario capacitar a dichos usuarios para poder aprovechar al máximo la solución.

14. Plan de implementación

El presente informe proporciona una visión general de las fases clave para la implementación exitosa de un Data Warehouse en DGME. El objetivo del plan de implementación es centralizar y organizar la información para mejorar la toma de decisiones, facilitar el análisis de datos y optimizar los procesos internos.

14.1. Fases de Implementación:

A continuación, se detallan las fases de la implementación de la solución. Se incluye el proceso de la extracción, transformación y carga de datos, hasta la presentación de los datos a los usuarios utilizando herramientas disponibles en el mercado.

14.1.1. Fase de Planificación y Preparación:

En esta fase, se definirán los objetivos y requisitos del Data Warehouse, evaluando el presupuesto disponible y asignando los recursos necesarios. Además, se formará y capacitará un equipo especializado en implementación de Data Warehouses y se seleccionará la tecnología adecuada para el proyecto.

14.1.2. Fase de Diseño y Modelado:

Durante esta etapa, se diseñará la arquitectura del Data Warehouse, definiendo el modelo de datos, esquemas y tablas necesarias para organizar la información de manera eficiente. En esta fase se realizará la creación del ETL de movimientos migratorios y restricciones también se configurará el entorno de desarrollo y pruebas para garantizar la calidad del sistema.

14.1.3. Fase de Extracción y Transformación de Datos:

En esta fase, se va a implementar la capa de extracción de datos, permitiendo la integración de información proveniente de diferentes fuentes existentes. Los datos se transformaron y limpiaron para garantizar su calidad antes de ser cargados en el Data Warehouse.

14.1.4. Fase de Carga de Datos y Configuración de la Capa de Almacenamiento:

Durante esta etapa, se va a configurar la capa de almacenamiento del Data Warehouse, estableciendo los procesos necesarios para la carga y gestión de los datos. Se realizarán pruebas exhaustivas de carga y rendimiento para asegurar el funcionamiento óptimo del sistema.

14.1.5. Fase de Implementación de la Capa de Presentación y Visualización:

En esta fase, se desarrollará interfaces de visualización de datos y herramientas de informes para facilitar el acceso y análisis de la información almacenada en el Data Warehouse. Se va a configurar la capa de presentación, y se capacitará a los usuarios finales en el uso de las herramientas de visualización.

14.1.6. Fase de Ajustes Finales y Lanzamiento:

Durante esta etapa final, se realizarán ajustes y mejoras en la arquitectura y rendimiento del Data Warehouse. Se preparará la documentación y los procedimientos operativos necesarios para su correcto funcionamiento. Finalmente, se llevará a cabo pruebas finales antes del lanzamiento oficial del Data Warehouse para su uso por parte de los usuarios finales.

14.2. Plan de ejecución

Fase 1: Planificación y preparación
Semana 1-2: Definición de los objetivos y requisitos del data warehouse. Evaluación del presupuesto y asignación de recursos.
Semana 3-4: Formación y capacitación del equipo de implementación. Selección de la tecnología de data warehousing.
Fase 2: Diseño y modelado
Semana 5-6: Diseño de la arquitectura del data warehouse, incluyendo el modelo de datos, esquemas y tablas necesarias.
Semana 7-8: Creación del ETL de movimientos migratorios
Semana 9-10: Creación del ETL de restricciones
Semana 11-12: Configuración del entorno de desarrollo y pruebas. Preparación de los procesos de extracción y transformación de datos.
Fase 3: Extracción y transformación de datos
Semana 13-16: Implementación de la capa de extracción de datos. Extracción de datos de las fuentes existentes y transformación para su carga en el data warehouse.
Fase 4: Carga de datos y configuración de la capa de almacenamiento
Semana 17-20: Configuración de la capa de almacenamiento del data warehouse. Diseño e implementación de los procesos de carga de datos.
Semana 21-22: Realización de pruebas de carga y rendimiento. Ajustes y optimizaciones necesarios.
Fase 5: Implementación de la capa de presentación y visualización
Semana 23-26: Desarrollo de interfaces de visualización de datos y herramientas de informes. Configuración de la capa de presentación del data warehouse.
Semana 27-28: Pruebas de integración entre el data warehouse y las herramientas de visualización. Capacitación de los usuarios finales en las herramientas de visualización.
Semana 29-30: Creación de reportes en tableau
Fase 6: Ajustes finales, capacitación y lanzamiento
Semana 31-32: Ajustes y mejoras en la arquitectura y rendimiento del data warehouse. Preparación de la documentación y procedimientos operativos.
Semana 33: Pruebas finales y corrección de posibles problemas identificados.
Semana 34-35: Capacitación de personal
Semana 36: Lanzamiento oficial del data warehouse y puesta en marcha para uso de los usuarios finales.

Ilustración 12. Plan de ejecución del proyecto. Fuente: Elaboración propia

14.3. Presupuesto

Para la elaboración del presupuesto, primero es necesario hay que aclarar que la institución proveerá ciertos insumos para el desarrollo, como lo son el equipo informático, la infraestructura de instalación (servidores propios de la institución) y además proveerá un espacio de trabajo (en el área de informática de la institución), por lo cual no se incurrirá en gastos para suplir estas necesidades.

Sin embargo, debe contemplarse los salarios del equipo de implementación y el auditor externo. Tomando en cuenta también los salarios de las personas involucradas que trabajan en la institución, se realizara un presupuesto. La duración del proyecto está proyectado a 8 meses, por lo que el estimado del presupuesto está basado en ese lapso de meses.

Concepto	Costo Unitario	Descripción	Precio total (mensual)	Total (durante 8 meses)
Salarios equipo implementador - 1 líder de proyecto - 1 líder de implementación - 2 asistentes de implementación	\$2,500.00	El salario mensual es de \$2,500.00 por 4 profesionales. Inversión mensual por los 8 meses es \$80,000.00	\$10,000.00	\$80,000.00
Auditor externo (1)	\$3,000.00	El único pago del auditor sería por trabajo realizado durante una semana.	\$3,000.00	\$3,000.00

Software Pentaho	\$0.00	El sistema escogido es la versión Open Source de Pentaho	\$0.00	\$0.00
Director de informática	\$2,300.00	Es el salario normal del director de informática, lo cual se toma en cuenta por deducciones del precio del proyecto	\$2,300.000	\$18,400.00
Coordinador desarrollo informático	\$1,500.00	Salario normal del coordinador de desarrollo informático, se toma en cuenta por deducciones del precio del proyecto	\$1,500.00	\$12,000.00
QA	\$2,000.00	Es el encargado de asegurar el correcto funcionamiento de la implementación, su contratación será de 3 meses	\$2,000.00	\$6,000.00

Depreciación equipo (HP ProDesk 400 G5)	\$7.30 (depreciación mensual)	El cálculo de depreciación es precio del equipo entre años útiles de vida y se realizó deducción mensual. Contando los 8 integrantes del proyecto exceptuando auditor, sumando \$58.40	\$58.40	\$467.20
			Total	\$119,867.20

Tabla 3. Presupuesto del proyecto

Formula depreciación:

Precio equipo informático: \$1,200.00

Vida útil: 120000 horas = 13.69 años

$$\frac{1200}{13.69} = 87.65 / 12 \text{ meses} = 7.30$$

NOTA: La depreciación del equipo y el costo de salarios de los empleados de la institución se toma en cuenta por efectos de costos netos del proyecto. El impuesto por servicios profesionales únicamente se les aplica a los salarios del equipo implementador y auditor externo. Los roles están descritos más adelante en el documento en el apartado “roles y responsables”.

14.4. Roles y responsables

Para la ejecución del proyecto serán necesarios los siguientes perfiles:

Perfil	Responsable	Descripción
Product Owner	Lic. Ismael Diaz	Es el representante de la institución, será responsable de velar porque el proyecto se lleve a cabo según lo acordado en el acta de implementación y además será un canal de comunicación entre el equipo implementador y los stakeholders
Líder de Proyecto	Por definir	Actuará como un facilitador para el equipo, su principal actividad es minimizar los obstáculos que el equipo pueda encontrar para el desarrollo de sus actividades, además, velara por que la metodología SCRUM se cumpla en cuanto a sus marcos de tiempo, reuniones y resultados.
Equipo de implementación	Por definir	Estará conformado por 3 personas que serán las encargadas de implementar la solución.
QA	Por definir	El equipo de aseguramiento de calidad estará conformado por 1 persona, su función es evaluar las tareas que el equipo de implementación vaya finalizando, además de diferentes tipos de pruebas generales en el sistema.
Auditoria	Por definir	Equipo conformado por 1 persona, que serán los encargados de realizar la auditoria según el plan expuesto una vez se finalice el proyecto.

Director informático	Lic. Ismael Diaz	Funge actualmente como director de informática de la institución, y fungirá como QA dentro del proyecto
Coordinador desarrollo informático	Lic. Linda Hernández	Funge actualmente como coordinador de desarrollo y es el encargado de programar, mantener y calcular los insumos necesarios para solventar los diferentes problemas en la institución

Tabla 4 Roles para la ejecución del proyecto

15. Conclusiones

La implementación de un Data Warehouse (DW) en la Dirección General de Migración y Extranjería para el control y restricciones migratorias ofrece numerosos beneficios y oportunidades.

En primer lugar, el Data Warehouse permite centralizar y consolidar datos de diversas fuentes en un único repositorio, lo que facilita la gestión y el análisis de información migratoria de manera más eficiente. Esta consolidación de datos elimina la duplicidad y mejora la calidad de estos, lo que a su vez fortalece la integridad y confiabilidad de la información utilizada para el control migratorio.

En segundo lugar, el Data Warehouse habilita análisis avanzados de datos migratorios. Mediante técnicas analíticas y herramientas de minería de datos, se pueden descubrir patrones, tendencias y relaciones ocultas en los flujos migratorios. Esto proporciona una visión más profunda y precisa de los perfiles migratorios, los movimientos poblacionales y los factores que influyen en la migración, lo que permite a la institución migratoria tomar decisiones más informadas y diseñar políticas y estrategias más efectivas.

En tercer lugar, el Data Warehouse agiliza la generación de informes y la obtención de datos para la toma de decisiones. Con un sistema automatizado, los informes migratorios pueden generarse rápidamente y de manera consistente, brindando a los responsables de la toma de decisiones una visión clara y actualizada de la situación migratoria. Esto facilita la identificación de áreas de mejora, la evaluación del rendimiento y la implementación de medidas correctivas de manera oportuna.

Además, el Data Warehouse mejora la eficiencia en la gestión de solicitudes y trámites migratorios. Al contar con un sistema centralizado, los procesos de solicitud, evaluación y emisión de documentos migratorios se simplifican y agilizan, lo que reduce los tiempos de respuesta y mejora la experiencia de los solicitantes. Esto contribuye a una mayor satisfacción de los usuarios y a una utilización más eficiente de los recursos de la institución migratoria.

Por último, el Data Warehouse fortalece el control y la seguridad en la gestión migratoria. Al contar con un repositorio centralizado y herramientas de análisis, se facilita la detección de casos de migración ilegal, el seguimiento de movimientos migratorios y la identificación de posibles riesgos o amenazas. Esto permite a la institución migratoria tomar medidas preventivas y correctivas oportunas, fortaleciendo la seguridad en las fronteras y garantizando el cumplimiento de las regulaciones migratorias.

En resumen, la implementación de un Data Warehouse en una institución migratoria de El Salvador ofrece una amplia gama de beneficios, desde la gestión y análisis eficientes de datos hasta la generación de informes precisos y el fortalecimiento del control migratorio. Esta estrategia permite tomar decisiones más informadas, diseñar políticas y estrategias efectivas y mantener la seguridad en las fronteras, lo que contribuye a una gestión migratoria más sólida y eficiente en beneficio de la institución y la sociedad en su conjunto.

16. Bibliografía

DGME. (2 de mayo de 2023). *Marco legal y normativo*. Obtenido de: <https://www.migracion.gob.sv/marco-institucional>

Kinball, R. y Ross, M (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. Indianapolis: John Wiley & Sons, Inc.

Barton, R. (2013). *Talend Open Studio Cookbook*. Packt Publishing.

Belhaj, M. (2016). *Herramienta de ETL Powercenter Workflow Manager y Monitor*. Madrid: Universidad Carlos III de Madrid. Departamento de Informática.

Bonilla Botia, I., & Briceño Díaz, F. A. (2006). *Sistemas de Información como apoyo a la toma de decisiones*. Bogotá: PROSPECTIVA.

Cote, C., Lah, M., & Sarka, D. (2017). *SQL Server 2017 Integration Services Cookbook*. United Kindom: Packt Publishing LTD.

Díaz, J. C. (2016). *Diseño de procesos ETL*. Cataluña: Universitat Oberta de Catalunya.

González, S. M. (2013). *UNA SOLUCIÓN PARA LA INTEGRACIÓN DE DATOS BASADA EN ORACLE DATA INTEGRATOR*. La Habana: Instituto Superior Politécnico “José Antonio Echeverría”.

Hotman, L. M. (2010). *Oracle Fusion Middleware Introducción a Oracle Data Integrator*. Oracle Corporation.

Informatica LLC. (2018). *Informatica Developer Transformation Guide*.

Neeraj, P., Hand, P., & Kharpate, N. (2018). *Qlik Sense Cookbook*. Packt.

Pentaho Corporation. (2006). *Creating Pentaho Solutions*. Pentaho Corporation.

Pentaho Corporation. (2008). *Pentaho Open Source Business Intelligence Platform Technical*. Pentaho Corporation.

Polanco, I. Y., & Betancourt, J. F. (2022). *Análisis de datos con power bi, r-studio y knime*. Ediciones de la U.

Ramos, L. A. (2006). *Sistemas Inteligentes de Soporte a la Decisión en PYMES*. CDMX: Instituto Tecnológico y de Estudios Superiores de Monterrey.

Rios, A. (2009). *Oracle Data Integrator*. Oracle Corporation.

Russo, M., & Ferrari, A. (2016). *Introducing Microsoft Power BI*. Microsoft Corp.

Sinergia e Inteligencia de Negocio S.L. (2023). *Sinnexus*. Retrieved from *Sistemas de Soporte a la Decisión (DSS)*: https://www.sinnexus.com/business_intelligence/sistemas_soporte_decisiones.aspx

Tableau Software LLC. (2023). *Tableau*. Retrieved from <https://www.tableau.com/es-es/why-tableau/what-is-tableau>

Talend. (2020). *Talend Open Studio for Data Integration User Guide*.

Turban, E., Aronson, J., & Liang, T. (2005). *Decision Support Systems and Intelligent Systems*. Editorial Prentice Hall.