

PASO 4 – CONSTRUCCION DEL PROCESO ETL

JHON SEBASTIAN ZUÑIGA LOPEZ

TUTOR PEDRO ANTONIO SAMPER IBAÑEZ

GRUPO 30

UNIVERSIDAD NACIONAL ABIERTA Y A DISTANCIA  
ESCUELA DE CIENCIAS BÁSICAS, TECNOLOGÍA E INFORMACIÓN  
INGENIERIA EN SISTEMAS

Argelia – Cauca  
Abril 2022

## **Introducción**

Los metadatos permiten la extracción de información para tener mayor claridad de los datos, por lo cual es de gran importancia tener un buen manejo de estos. Para la actividad se plantea un caso de estudio sobre una aerolínea que quiere brindar un descuento a los clientes mas fieles, pero al intentar descubrir que clientes son los que mas han viajado con ellos se enfrentan con el problema de que los datos estas desestructurados y con duplicidad de datos. Por este motivo se busca generar una estrategia junto a una arquitectura de metadatos que sirva como estrategia para tener mayor claridad de los datos y poder brindar el beneficio adecuadamente.

## **Objetivos**

Generar una estrategia referente a los metadatos que permita organizar y entender los datos para que la aerolínea identifique a sus clientes más fieles para aplicarles un descuento.

Definir una arquitectura de metadatos para que la organización pueda tomar y brindar los veneficios adecuadamente.

Realizar un informe grafico que permita visualizar y entender de mejor manera los datos en una herramienta de gestión de metadatos como lo es Power Bi

**Enlace de la presentación:** [Presentación\\_Grupo30.pptx \(sharepoint.com\)](#)

**Enlace de la presentación importancia de los metadatos en las organizaciones**

[https://www.canva.com/design/DAGCci8NAIA/smVKszlDa8aaas2nymhl6Q/edit?utm\\_content=DAGCci8NAIA&utm\\_campaign=designshare&utm\\_medium=link2&utm\\_source=sharebutton](https://www.canva.com/design/DAGCci8NAIA/smVKszlDa8aaas2nymhl6Q/edit?utm_content=DAGCci8NAIA&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton)

**Enlace de la presentación de herramientas para la gestión de metadatos**

[https://www.canva.com/design/DAGFEFzdMXM/SVdfs\\_N\\_vX8VHdAoqQxL1A/edit?utm\\_content=DAGFEFzdMXM&utm\\_campaign=designshare&utm\\_medium=link2&utm\\_source=sharebutton](https://www.canva.com/design/DAGFEFzdMXM/SVdfs_N_vX8VHdAoqQxL1A/edit?utm_content=DAGFEFzdMXM&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton)

### **Caso de estudio**

Los departamentos de Marketing y ventas y la gerencia de una aerolínea están consolidando la información relacionada con los tickets comprados por parte de sus clientes con el fin de regalar un 20% de descuento para su próxima compra a los 3 clientes que más han volado con ellos. Al revisar la información de clientes, el departamento de marketing y ventas tienen para algunos clientes diferentes direcciones, correos electrónicos y teléfonos, esta información proviene del sistema de ventas y del sistema de marketing donde se almacena la información de los clientes y los beneficios que se han otorgado en las diferentes campañas publicitarias donde se generan descuentos. Para otorgar el beneficio del 20% de descuento, los clientes deben acumular millas, esta información de millas se almacenan en un Excel, por lo tanto, se debe cruzar la información de los clientes que más han comprado tickets con el archivo Excel; al generar el reporte de millas, en el sistema aparecen números de identificación duplicados para diferentes clientes.

**Tabla NUF**

Oportunidad de mejora detectada	Partes interesadas	Fuentes de metadatos	Acción para realizar	Nuevo (valor del 1 al 10)	Útil (valor del 1 al 10)	Factible (Valor del 1 al 10)	Promedio
Diferentes direcciones, correos electrónicos y teléfonos para un mismo cliente.	Departamento de ventas y marketing	Sistema de ventas y marketing	Integrar la información de los dos sistemas en solo conjunto de datos	9	9	10	9.3
La información almacenada en un Excel contiene números de identificación duplicados	Departamento IT, departamento de marketing y ventas, Gerencia	Hoja de calculo	Utilizar herramientas de limpieza de datos para identificar y eliminar registros duplicados. Establecer criterios claros para determinar que registros deben mantenerse.	8	9	9	8.6
Falta de integridad en la información de los contactos	Departamento de marketing y ventas, Departamento de IT	Sistema de ventas, sistemas de marketing	Establecer procesos de validación de datos durante la captura para garantizar la integridad de la información. Realizar auditorías periódicas de la calidad de los datos de contacto	8	8	9	8.3

## **Estrategia de metadatos**

### **Acciones más relevantes**

Integrar la información de los dos sistemas en solo conjunto de datos: permite consolidar y unificar toda la información dispersa en diferentes sistemas de almacenamiento. Esta integración facilita la gestión de los datos con el objetivo de proporcionar una visión global y unificada de los datos disponible en la aerolínea

Utilizar herramientas de limpieza de datos para identificar y eliminar registros duplicados y establecer criterios claros para determinar que registros deben mantenerse: estas estrategias permiten la mejora de calidad de datos asiendo que la información almacenada se consistente, lo que hace que la búsqueda y recuperación de datos sea mucho más fácil y precisa.

### **Responsable directo**

El departamento de IT será el responsable de gestionar y supervisar todas las actividades relacionadas con la limpieza, integridad y establecimiento de estándares y de mas actividades establecidas en las acciones de mejora encontradas.

### **Financiación de la estrategia**

El presupuesto será compartido entre el departamento de IT y la gerencia, esto con el objetivo de invertir en las mejores herramientas y métodos que lleven a una buena organización de los datos.

### **Encargado de gestionar la estrategia**

La responsabilidad de esta se llevará en conjunto entre el departamento de IT y los dueños de los datos en este caso el departamento de ventas y marketing. IT se encargará de lo

técnico como limpieza de datos, integración de datos e implementación de sistemas de gestión, mientras tanto ventas y marketing se encargarán de definir y mantener estándares de los datos.

### **Datos Maestros**

Personas	Cosas	Lugares	Conceptos
Información de los clientes: nombre, apellido, correo, teléfono, # identificación, dirección, historial de vuelos, millas acumuladas	Tiquetes comprados: numero de vuelo, fecha de compra, precio, destino, clase del vuelo, estado del tiquete	Destino del vuelo: nombre del destino, país, departamento, ciudad, aeropuerto	Descuentos otorgados: porcentaje del descuento, cantantica, fecha de aplicación
	Beneficios otorgados: tipo de beneficio, cantidad, fecha de emisión, fecha de vencimiento	Dirección del usuario: dirección de facturación, ciudad, país, código postal	Millas acumuladas: cantidad de millas, fecha de acumulación, estado de las millas

### Metadatos técnicos

Fuente de datos	Metadato	Detalles técnicos de los datos
Sistemas de ventas	Cliente	Nombre de la entidad
	Tipo_id	Campo de la tabla cliente de tipo varchar con extensión de 3 caracteres, este dato no poder nulo
	Numero_id	Campo de la tabla cliente que representa el número de identificación del usuario, este valor debe de ser de tipo entero
	Primer_nombre	Campo de a tabla cliente que representa el nombre del usuario, este campo no puede ser nulo y su extensión es de 50 caracteres
	Primer_apellido	Campo de la tabla cliente que representa el apellido del usuario, este campo puede ser nulo y su



		extensión es de 50 caracteres
	Vuelo	Nombre de la entidad
	Modelo_avion	Campo de la tabla vuelo, este campo es de tipo varchar el cual no puede ser nulo y tiene una extensión de 100 caracteres
	Origen	Nombre de la entidad
	Ciudad	Campo de la tabla origen de tipo varchar con extensión de 100 caracteres. Este campo permite valores nulos
	Aeropuerto	Campo de la tabla origen de tipo varchar con extensión de 200 caracteres. Este campo permite valores nulos
Sistemas de ventas	Destino	Nombre de la entidad
	Ciudad	Campo de la tabla destino de tipo varchar con extensión de 100

		caracteres. Este campo no permite valores nulos
	Aeropuerto	Campo de la tabla destino de tipo varchar con extensión de 200 caracteres. Este campo no permite valores nulos
Base marketing	Cliente	Nombre de la entidad
	Identificación	Campo de la tabla cliente de tipo varchar de extensión 255, permite valores null
	Nombre	Campo de la tabla cliente de tipo varchar de extensión 255, permite valores null
	Apellido	Campo de la tabla cliente de tipo varchar de extensión 255, permite valores null
Base millas	Identificación	Campo numérico que identifica a un usuario

	Nombre	Campo varchar para el nombre completo del usuario
--	--------	---

**Arquitectura de metadatos**

Fuente de metadatos	Repositorio (Centralizado, distribuido o híbrido)	Usuarios finales, beneficiados
Sistema de ventas Base marketing Base millas	<p>Híbrida</p> <pre> graph TD     SV[Sistema de ventas] -- "datos del cliente" --&gt; SM[Sistema de marketing]     SM &lt;--&gt; SMI[Sistema de millas]     SM -- "Datos de interacciones y beneficios" --&gt; BDR[Base de datos relacional]     SMI -- "Datos de millas" --&gt; BDR     BDR -- "Metadatos" --&gt; SMet[Servicio de metadatos]     SMI -- "Metadatos" --&gt; SMet           </pre> <p>El flujo de los datos inicia en el sistema de ventas recolectando la información relacionada con las ventas de los tickets incluyendo datos de los clientes y los vuelos, a continuación, el sistema de márketing registra toda la interacción relacionada con los clientes, así mismo lo hace el sistema de millas, toda esta información es almacenada en una base de datos relacional la cual suministra datos al servicio de metadatos. Estos datos pueden ser analizados por todos los sistemas con el objetivo de utilizarlos para generar reportes, ver el comportamiento de los clientes y</p>	Clientes, área de marketing, área de ventas

	por supuesto aplicar las promociones que se tienen pensadas.	
--	--	--

### **Importancia de las herramientas de gestión de metadatos en una organización.**

Según la empresa Informática, el enfoque de gestión de metadatos ofrece varias ventajas. Permite el autoservicio en el análisis de datos, mejora la calidad y el gobierno de los datos, asegura la privacidad de los datos, facilita la gestión de datos maestros, agiliza la modernización hacia la nube, acelera la integración de datos y facilita el DevOps para la gestión de datos. Esto se logra mediante funciones como búsqueda y detección de datos relevantes, documentación detallada, seguimiento de la privacidad y recomendaciones para la integración y la migración a la nube.

Algunas empresas que permiten realizar esta gestión son:

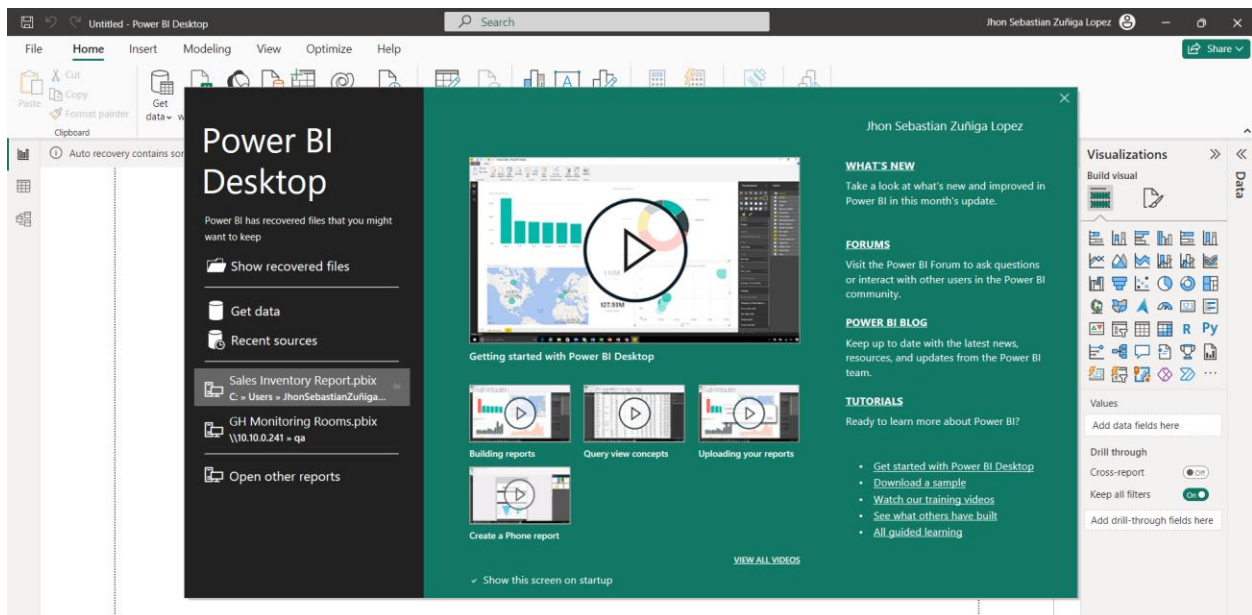
**Capgemini:** una empresa francesa, que a través de esta actividad garantiza el valor continuo de los activos de datos de una empresa. La solución de gestión de metadatos que ofrecen permite la integración con portales de metadatos, obtención y extracción de metadatos de diferentes fuentes para facilitar su actualización, mantenimiento, integración, depuración y consolidación, entre muchas otras funcionalidades.

**Talend Metadata Manager:** Conecta datos desde plataformas, bases de datos y herramientas analíticas para generar una visión integral de la cadena de suministro de información en un lenguaje comprensible para todos. Permite registrar metadatos a cada objeto de datos utilizado en una tarea de análisis de datos.

**Oracle Enterprise Metadata Management:** también es una herramienta de gestión de metadatos. Mejora la búsqueda de metadatos a través de búsquedas de texto y capacidades de exploración de metadatos proporcionando el linaje de datos e informes de análisis de impacto en todas las tecnologías.

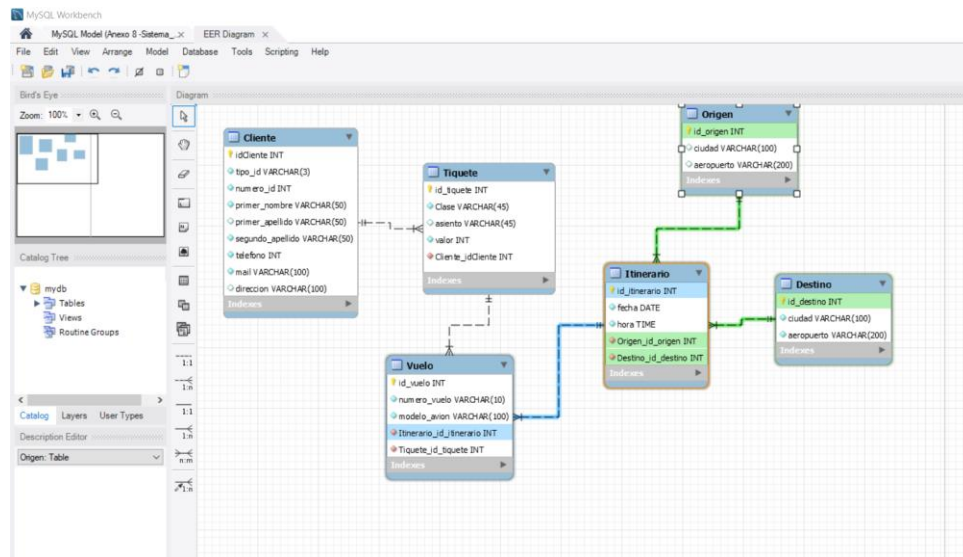
## Sinopsis de la herramienta Microsoft Power BI

Power BI es un conjunto de herramientas que pone el conocimiento al alcance de todos y nos brinda acceder a nuestros datos de forma segura y rápida, generando grandes beneficios para nosotros y para nuestra empresa. Es un sistema predictivo, inteligente y de gran apoyo, capaz de traducir los datos (simples o complejos) en gráficas, paneles o informes por sus cualidades como la capacidad gráfica de presentación de la información, o la integración de Power Query: el motor de extracción, transformación y carga (ETL) incluido en Excel.

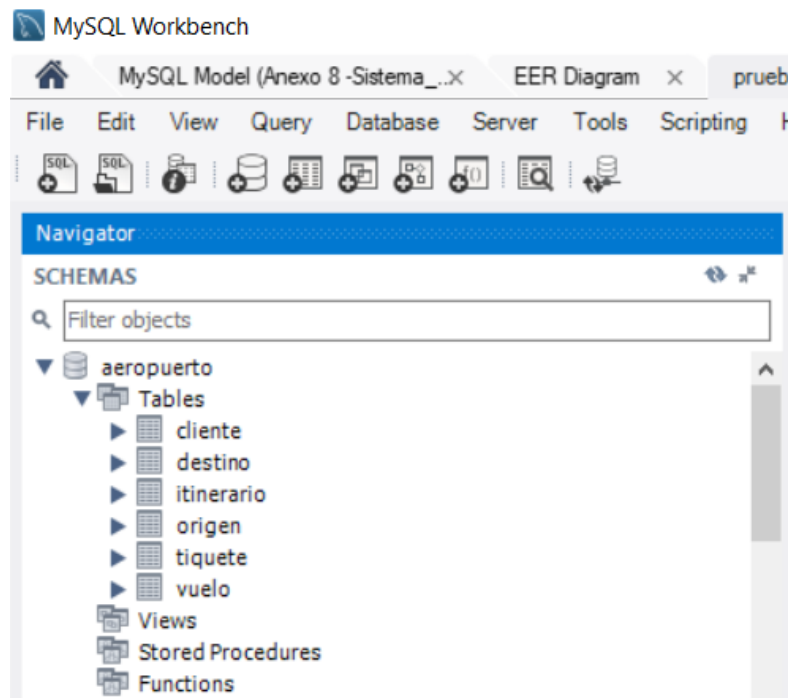


## Configuración sistemas de ventas en Workbench MySQL

Diagrama de clases.



Tablas creadas.



Insertión de datos en las diferentes tablas de la base.

cliente cliente - Table

Limit to 50000 rows

1 • `SELECT * FROM aeropuerto.cliente;`

<

Result Grid

Filter Rows:

Edit:

Export/Import:

Wrap Cell Content:

	idCliente	tipo_id	numero_id	primer_nombre	primer_apellido	segundo_apellido	telefono	mail	direccion
1	CC	11111111	Juan	Carrillo	Perez	3178342314	tuan@gmail.com	Cali calle 23 #12	
2	CC	22222222	Tatiana	Arboleda	Jaramillo	3142546781	tatiana@gmail.com	Bucaramana calle 21 #89	
3	TI	33333333	Pedro	Lopez	Villar	3154789213	pedro@gmail.com	Cali calle 78 # 2	
4	CC	44444444	Sandra	Martinez	Figueroa	3190874567	sandra@gmail.com	Cali calle 1 #12	
5	CC	55555555	Manuela	Muñoz	Núñez	3178231423	manuela@gmail.com	Booota calle 4 #89	
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	

cliente **tiquete** tiquete - Table

Limit to 50000 rows

1 • `SELECT * FROM aeropuerto.tiquete;`

Result Grid

	id_tiquete	Clase	asiento	valor	Cliente_idCliente
1	1	ejecutiva	B23	500000	2
2	2	economica	A5	150000	5
3	3	premium	F1	1000000	2
4	4	economica	A9	180000	1
5	5	economica	A10	100000	3
	NULL	NULL	NULL	NULL	NULL

cliente tiquete **origen** destino

Limit to 50000 rows

1 • `SELECT * FROM aeropuerto.origen;`

Result Grid

	id_origen	ciudad	aeropuerto
1	1	Cali	aero-cali
2	2	Bucaramana	aero-bucaramana
3	3	Booota	aero-booota
4	4	Medellin	aero-medellin
5	5	Barranquilla	aero-barranquilla
	NULL	NULL	NULL

cliente	tiquete	origen	destino
<div> <div>1</div> <div>SELECT * FROM aeropuerto.destino</div> </div>			
<div> <div>Result Grid</div> <div>Filter Rows:</div> </div>			
	id_destino	ciudad	aeropuerto
1	1	Barranquilla	aero-barranquilla
2	2	Cali	aero-cali
3	3	Boacota	aero-boacota
4	4	Pasto	aero-pasto
5	5	manizales	aero-manuzales
	NULL	NULL	NULL

cliente	tiquete	origen	destino	itinerario	vuelo	vuelo - Tabl
<div> <div>1</div> <div>SELECT * FROM aeropuerto.itinerario;</div> </div>						
<div> <div>Result Grid</div> <div>Filter Rows:</div> <div>Edit:</div> <div>Export/Import:</div> </div>						
	id_itinerario	fecha	hora	Origen_id_origen	Destino_id_destino	
1	1	2024-06-17	11:17:00	1	4	
2	2	2024-06-20	01:00:00	2	5	
3	3	2024-07-02	13:40:00	5	3	
4	4	2024-07-17	12:10:00	1	5	
5	5	2024-08-23	09:00:00	3	1	
	NULL	NULL	NULL	NULL	NULL	

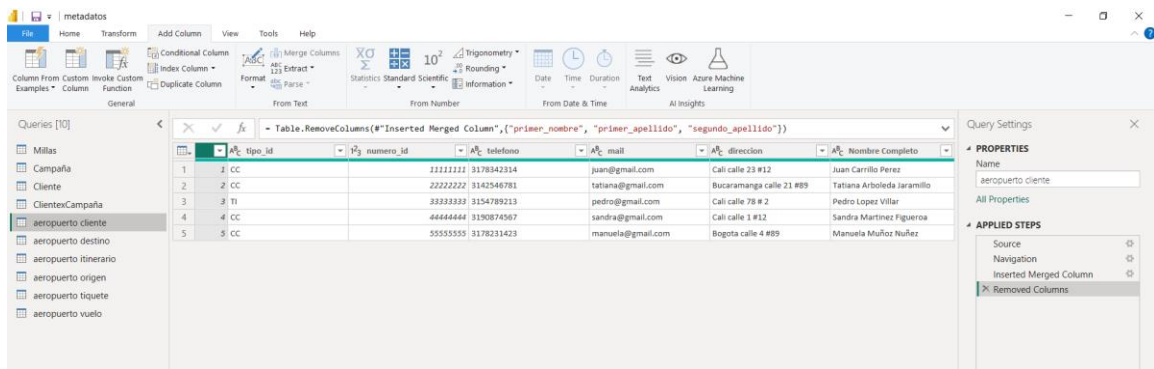


cliente	tiquete	origen	destino	itinerario	vuelo	vuelo - T
<div> <div>1</div> <div>SELECT * FROM aeropuerto.vuelo;</div> </div>						
Result Grid						
	id_vuelo	numero_vuelo	modelo_avion	Itinerario_id_itinerario	Tiquete_id_tiquete	
	1	01	01AB	2	1	
	2	02	02CD	4	3	
	3	03	03EF	5	2	
	4	04	04GH	1	4	
	5	04	05IJ	3	1	
	NULL	NULL	NULL	NULL	NULL	

## Transformación de datos en Power BI.

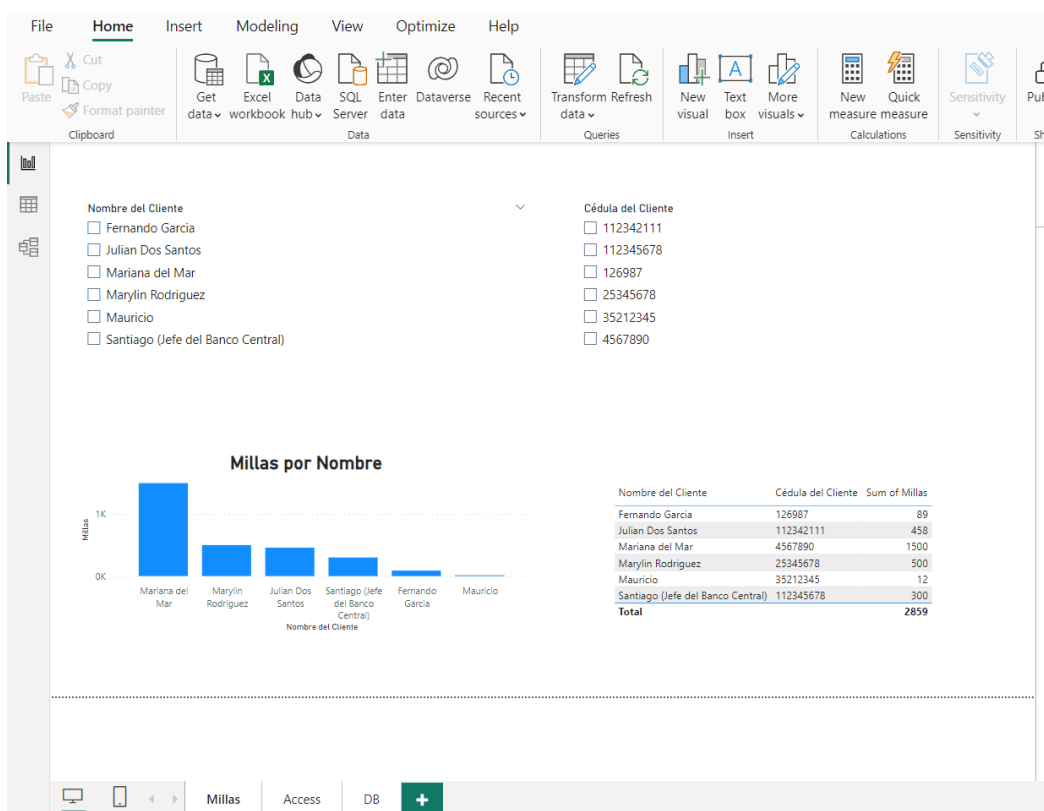
Para que se pueda dar un buen análisis gráficos de los datos es importante realizar una limpieza de estos, esto incluye quitar datos duplicados, borrar datos nulos, dar el formato correcto a cada columna o, separar o unir datos que lo amerites.

metadatos		
File	Home	Transform
Close & Apply	New Source	Recent
Close	Source	Enter Data
		Data source settings
		Manage Parameters
		Refresh Preview
		Advanced Editor
		Manage Columns
		Remove Columns
		Keep Rows
		Remove Rows
		Sort
		Split Column
		Group By
		Transform
		Data Type: Whole Number
		Use First Row as Headers
		Replace Values
		Merge Queries
		Append Queries
		Combine Files
		AI Insights
		Text Analytics
		Vision
		Azure Machine Learning
		Combine
		AI Insights
Queries [10]		
Hoja1		
Campaña		
Cliente		
ClientexCampaña		
aeropuerto cliente		
aeropuerto destino		
aeropuerto itinerario		
aeropuerto origen		
aeropuerto tiquete		
aeropuerto vuelo		
= Source([Schema="aeropuerto",Item="destino"])[Data]		
id_destino	ciudad	aeropuerto
1	1 Barranquilla	aero-barranquilla
2	2 Cali	aero-cali
3	3 Bogota	aero-bogota
4	4 Pasto	aero-pasto
5	5 manizales	aero-manizales
Query Settings		
PROPERTIES		
Name		
aeropuerto destino		
All Properties		
APPLIED STEPS		
Source		
Navigation		



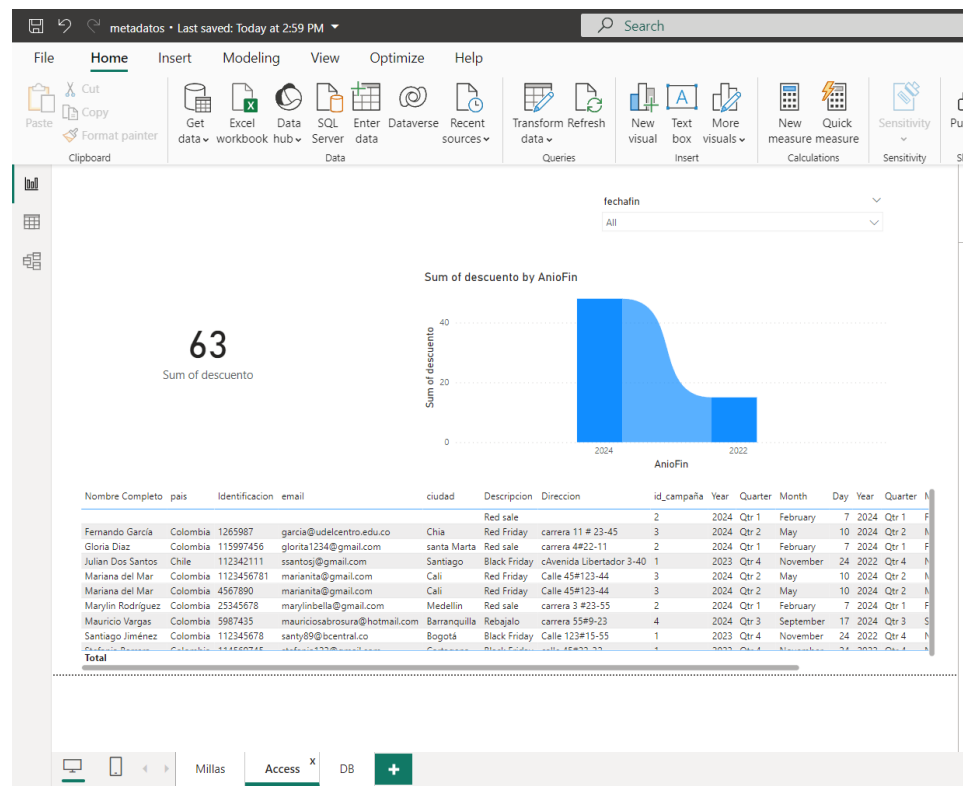
## Grafico para la base de Millas

En grafico esta diseñado con el objetivo de identificar rápidamente el o los usuarios que ms millas tengan acumulas además de que permite realizar filtros ya sea por medio del nombre o número de identificación.



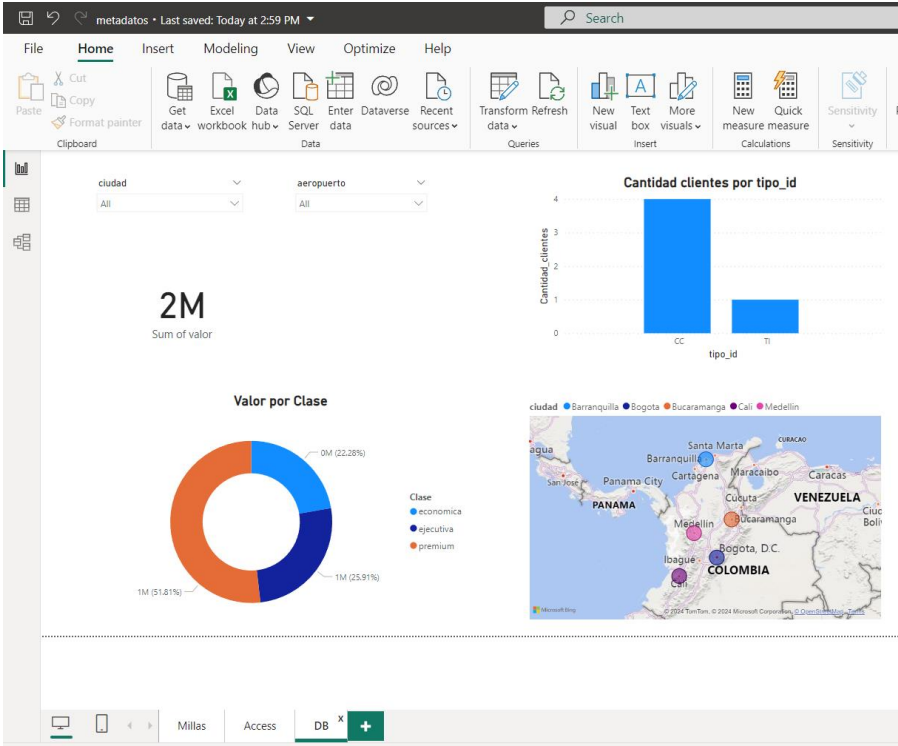
## Grafico para la base de marketing

En el siguiente reporte se da a conocer información relevante de los descuentos realizados. Se establece una gráfica en la cual se identifica la cantidad descuentos realizados durante cada año. Al igual que una tabla en la que se puede observar la información de completa de cada descuento realizado.



## Reporte para la base de datos de sistemas de ventas

Este grafico permite identificar varios aspectos de este sistema, tales como el valor total de las ventas, como también la clasificación de ventas por cada clase identificada. Otro aspecto importante que se puede observar es el tipo de documento que utilizan los usuarios.



## **Conclusiones**

Se puede decir que implementar una estrategia de metadatos es de gran beneficio para garantizar la calidad, consistencia y utilidad de los datos en una organización. Esto a través de acciones como la integración de datos, limpieza y establecimiento de criterios claros para la gestión de metadatos.

La integración de datos provenientes de múltiples fuentes, como sistemas de ventas y marketing, y la limpieza de registros duplicados, son pasos críticos en la estrategia de metadatos. Estas acciones no solo contribuyen a la mejora de la calidad de los datos, sino que también facilitan la generación de informes precisos y análisis significativos que respaldan la toma de decisiones estratégicas.

Definir una arquitectura de metadatos es crucial para cualquier organización, ya que proporciona una estructura ordenada y coherente para la gestión efectiva de los datos. Esta arquitectura establece estándares y prácticas que mejoran la comprensión de los datos, garantizando su calidad y facilitando su búsqueda y recuperación. Al organizar y etiquetar los metadatos de manera adecuada, se promueve la interoperabilidad entre sistemas y se facilita el intercambio de información, lo que aumenta la eficiencia operativa y la colaboración entre equipos. Además, una arquitectura de metadatos bien definida permite a las organizaciones adaptarse ágilmente a los cambios tecnológicos y empresariales, manteniéndolas competitivas en un entorno empresarial dinámico. En resumen, establecer una arquitectura de metadatos sólida es esencial para maximizar el valor de los datos de una organización y fomentar una toma de decisiones informada y eficiente.

Las herramientas de gestión de metadatos son fundamentales para garantizar la precisión y confiabilidad de los informes y análisis. Al utilizar estas herramientas como Power BI, las organizaciones pueden definir y mantener consistentemente la estructura y semántica de los datos utilizados. Esto asegura que los usuarios tengan acceso a información coherente y precisa, lo que facilita la toma de decisiones informadas. Además, las herramientas de gestión de metadatos como Power BI permiten la documentación completa de conjuntos de datos, informes y visualizaciones, lo que facilita la colaboración y el entendimiento compartido entre los equipos.

## Referencias bibliográficas

Amórtégui, M. P. (2020). [Introducción a los metadatos](#). [Objeto\_virtual\_de\_Informacion\_OVI].

Repositorio Institucional UNAD. <https://repository.unad.edu.co/handle/10596/36140>

Capgemini, «Capgemini,» Capgemini, [En línea]. Available: <https://www.capgemini.com/>

Díaz, J. C. (n.d.). Gestión de datos maestros. Uoc.edu. Retrieved April 13, 2024, from

[https://openaccess.uoc.edu/bitstream/10609/140887/1/Gobierno%20del%20dato\\_M%C3%B3dulo%202\\_Gesti%C3%B3n%20de%20datos%20maestros.pdf](https://openaccess.uoc.edu/bitstream/10609/140887/1/Gobierno%20del%20dato_M%C3%B3dulo%202_Gesti%C3%B3n%20de%20datos%20maestros.pdf)

Feria Basurto, L. (2007). [Metadatos para la sociedad de la información: encontrando la aguja en el pajar](#). Páginas 3-6, Revista Bibliotecas. Vol. 25, No. 1, 2007. Red Universidad

Nacional de Costa Rica. <https://elibro->

[net.bibliotecavirtual.unad.edu.co/es/ereader/unad/23376?page=3](https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/23376?page=3)

Gestión de metadatos. (s/f). Informatica.com. Recuperado el 5 de mayo de 2024, de

<https://www.informatica.com/co/products/informatica-platform/metadata-management.html>

Jiménez, J. H. (2023). Datos Maestros. [Objeto\_virtual\_de\_Informacion\_OVI]. Repositorio

Institucional UNAD. <https://repository.unad.edu.co/handle/10596/58642>

Muñoz Reja, I. C. Gómez Carretero, A. I. y Gualo Cejudo, F. (2018). [Calidad de datos](#). Páginas 25-39, 67-124, RA-MA

Editorial. <https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=2498292&lang=es&site=eds-live&scope=site>

Muñoz Reja, I. C. Gómez Carretero, A. I. y Gualo Cejudo, F. (2018). Calidad de datos. Páginas 41-56 RA-MA Editorial.

<https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=2498292&lang=es&site=eds-live&scope=site>

Oracle, «Oracle Enterprise Metadata Management,» Oracle, [En línea]. Available:

<https://www.oracle.com/middleware/technologies/enterprise-metadata-management.html>

Talend, «es.talend,» Talend, 2006. [En línea]. Available: <https://es.talend.com/>