



Prototipo de automatización en la clasificación de las solicitudes de reconocimiento pensional

NELSON ARTURO CASTILLO LOPEZ
LEIDY YINNETH MALDONADO MATIZ

UNIVERSIDAD CENTRAL
FACULTAD DE INGENIERIA
MAESTRIA EN ANALITICA DE DATOS
BASES DE DATOS
MAYO 2023

Contents

1	Introducción	2
2	Características del proyecto de investigación que hace uso de Base de Datos	3
2.1	Titulo del proyecto de investigación	3
2.2	Objetivo General	3
2.3	Objetivos Especificos	3
2.4	Alcance	3
2.5	Pregunta de Investigación	4
2.6	Hipotesis	4
3	Reflexiones sobre el origen de datos e información	5
3.1	¿Cuál es el origen de los datos e información?	5
3.2	¿Cuáles son las consideraciones legales o éticas del uso de la información?	5
3.3	¿Cuáles son los retos de la información y de los datos que utilizará en la base de datos en términos de calidad y la consolidación? . .	5
3.4	¿Qué espera de la utilización de un sistema de Base de Datos para su proyecto?	5
4	Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)	6
4.1	Características del SMBD (Sistema Manejador de Base de Datos) para el proyecto	6
4.2	Diagrama de modelo de Datos	6
4.3	Imágenes de la Base de Datos	7
4.4	Código SQL – lenguaje de definición de datos (DDL)	8
4.5	Código SQL – manipulación de datos (DML)	10
4.6	Código SQL + Resultados: Vistas	10
4.7	Código SQL + Resultados: Triggers	11
4.8	Código SQL + Resultados: Triggers	11
4.9	Código SQL + Resultados: Funciones	11
4.10	Código SQL + Resultados: procedimientos almacenados	12
5	Bases de Datos No-SQL	13
5.1	Diagrama Bases de Datos No-SQL	13
5.2	SMBD utilizado para la Base de Datos No-SQL	14
6	Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos	15
6.1	Ejemplo de aplicación de ETL y Bodega de Datos	15
7	Lecciones Aprendidas	16
8	Referencias	17

1 Introducción

El 1 de octubre de 2012, la entidad estatal Colpensiones entró en funcionamiento con el objetivo de administrar el Régimen de Prima Media con prestación definida. Una de las funciones desagregadas de esta entidad es determinar los derechos pensionales y las prestaciones económicas para los afiliados al Régimen de Prima Media con Prestación Definida.

Al cierre del año 2022, se evidencia que Colpensiones cuenta con 6.7 millones de afiliados, lo cual representa el 30.4% de la población ocupada del país. De estos afiliados, 1.57 millones son pensionados, y la entidad tiene una nómina de alrededor de 2.8 billones de pesos, según las cifras del Informe de Gestión 2022.

Teniendo en cuenta las cifras mencionadas y el objetivo de la entidad, se presenta un desafío en la atención de las solicitudes. Según el informe de gestión, se recibieron 342,758 solicitudes en 2022. La entidad ha implementado procesos de automatización en la toma de decisiones sobre el reconocimiento pensional, logrando atender el 48.9% de estas solicitudes en un promedio de 20.8 días. Esto indica que más de la mitad de los casos aún se procesan de manera manual o semiautomática. La información necesaria para el procesamiento se envía al área correspondiente a través de archivos txt o Excel, y la clasificación se realiza manualmente, generando demoras y errores en el proceso o en la cruzada de datos.

Con las bases de datos recibidas, se desarrollará un modelo entidad-relación que permita identificar patrones para implementar un modelo KNN. Este prototipo realizará una clasificación precisa para determinar si las solicitudes se procesarán de forma automática, manual o semiautomática.

2 Características del proyecto de investigación que hace uso de Base de Datos

2.1 Título del proyecto de investigación

Prototipo de automatización en la clasificación de las solicitudes de reconocimiento pensional

2.2 Objetivo General

Desarrollar un prototipo que utilice modelos de aprendizaje automático (machine learning) para clasificar las solicitudes de reconocimiento pensional, con el fin de mejorar la calidad de los datos y reducir el tiempo de procesamiento.

2.3 Objetivos Especificos

- * Crear el modelo entidad relación para las bases iniciales, para que el manejo de la base de datos sea adecuada en la clasificación.
- * Realizar la simulación de los datos teniendo en cuenta el comportamiento de los datos reales, para no realizar extracción de datos reales.
- * Realizar imputación de datos teniendo en cuenta los datos generados en la simulación de los mismos.
- * Realizar balanceo para que el modelo sea consistente en el aprendizaje.
- * Crear la estructura y el cargue de la data según la simulación realizada sobre los datos reales.
- * Realizar el análisis de los cruces y la clasificación de las variables a utilizar.
- * Crear el modelo mediante machine learning para la clasificación de las solicitudes de acuerdo con la data establecida con anterioridad.

2.4 Alcance

Desarrollar un prototipo que permita ingresar los números de radicado de las solicitudes nuevas y clasificar automáticamente el tipo de gestión que se asignará a cada caso (automático, manual o semiautomático). Esta clasificación se basará

en datos simulados que reflejen el comportamiento de la data real. Además, se creará un modelo entidad-relación para optimizar el tiempo de gestión de la información final.

2.5 Pregunta de Investigación

¿Cómo se puede mejorar la eficiencia en el tiempo y reducir los errores en la clasificación de las solicitudes de pensión que llegan diariamente a Colpensiones, teniendo en cuenta sus características?

2.6 Hipotesis

El proceso de clasificación de las solicitudes de pensión recibidas diariamente se llevará a cabo de manera más rápida utilizando el nuevo prototipo en comparación con el tiempo actualmente requerido.

3 Reflexiones sobre el origen de datos e información

3.1 ¿Cuál es el origen de los datos e información?

El proyecto al ser un prototipo, los datos han sido simulados, de acuerdo con el comportamiento de la data real. Esta data esta simulada teniendo en cuenta como allegan los mismos, los cuales son txt o Excel:

1. De la información entregada por los ciudadanos que solicitan la pensión.
2. De la información se extrae datos de acuerdo con la sede donde fue radicada la solicitud o donde reside el ciudadano.
3. Filtro y validación de la data del área encargada de la entrega del radicado al área de gestión de los mismos.

3.2 ¿Cuáles son las consideraciones legales o éticas del uso de la información?

Como se ha mencionado con anterioridad, y teniendo en cuenta la manipulación que se debe realizar a los datos y la sensibilidad que representa su origen, se estableció simular estos, de conformidad con el comportamiento, validando la estructura de los mismos, de esta manera, no existe afectación a la integridad de la data ni se incurre en violación a la Ley de protección de datos, y para que el prototipo clasifique de manera coecata se realiza una simulación teniendo en cuenta el comportamiento real.

3.3 ¿Cuáles son los retos de la información y de los datos que utilizará en la base de datos en términos de calidad y la consolidación?

En las bases reales se evidencia duplicidades que se deben saber manejar para tomar la información requerida para la toma de decisiones. Datos en blanco, falta de información relevante, esto en cualquiera de las datas entregadas. Limpieza de los datos, ejemplo: datos con signos de puntuación donde no deberían ir. Simulación de la data real.

3.4 ¿Qué espera de la utilización de un sistema de Base de Datos para su proyecto?

Minimizar tiempos de consulta y cruces, generando los campos requeridos de acuerdo a las solicitudes en el menor tiempo posible. Generando una copia de seguridad y traza de las modificaciones realizadas. Minimizar los errores humanos en el procesamiento de la data. Crear una base de datos entidad relación entenedible para cualquier tipo de necesidad

4 Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)

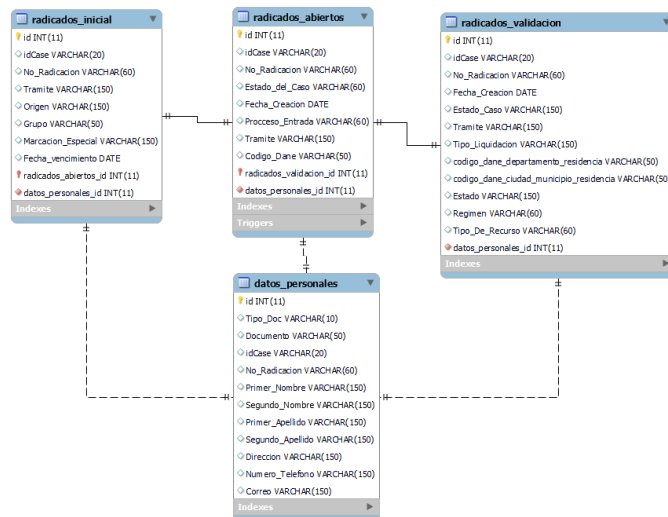
4.1 Características del SMBD (Sistema Manejador de Base de Datos) para el proyecto

En Colpensiones para manejo de ciertos programas el SMBD que se utiliza es SQL Server, el cual hasta el momento resiste las diferentes consultas y bases que se cargan como lo son las de este proyecto.

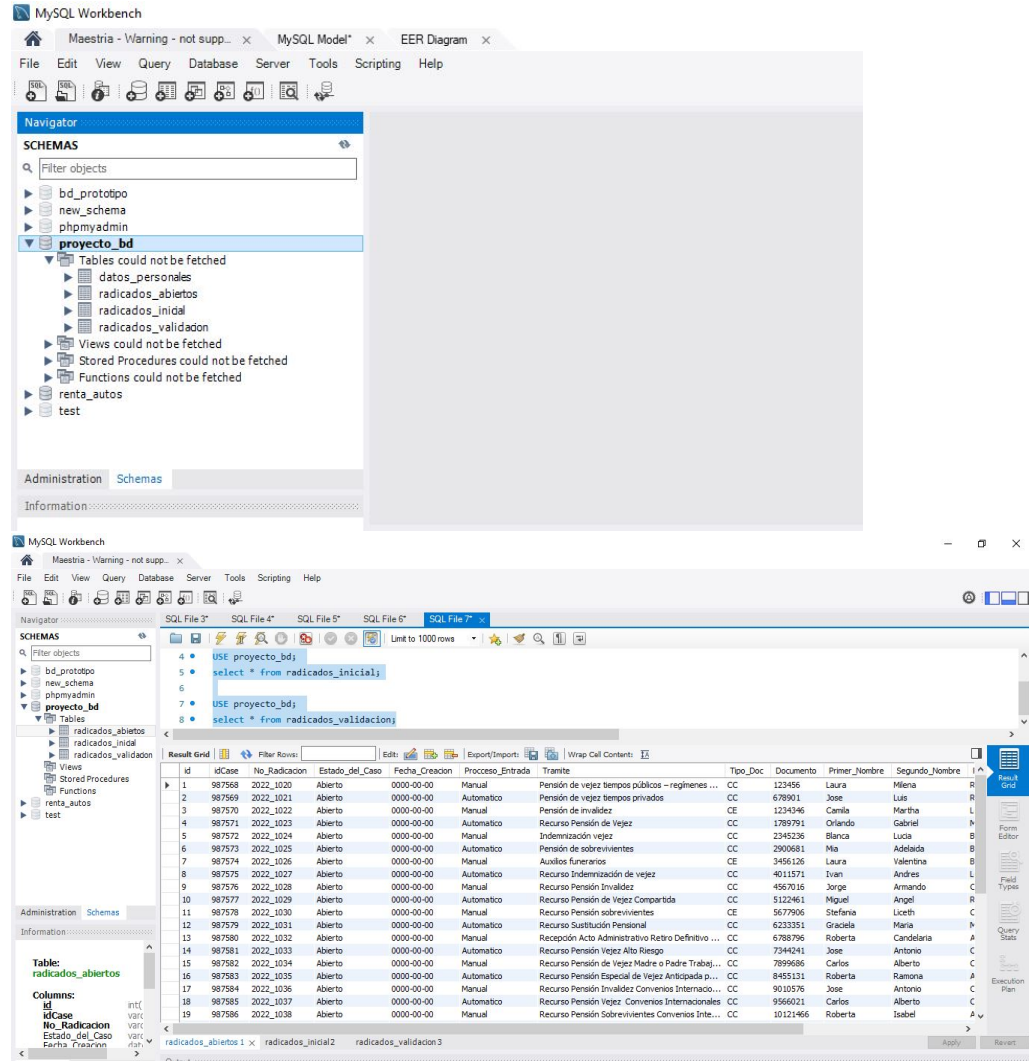
Las características de SQL Server son:

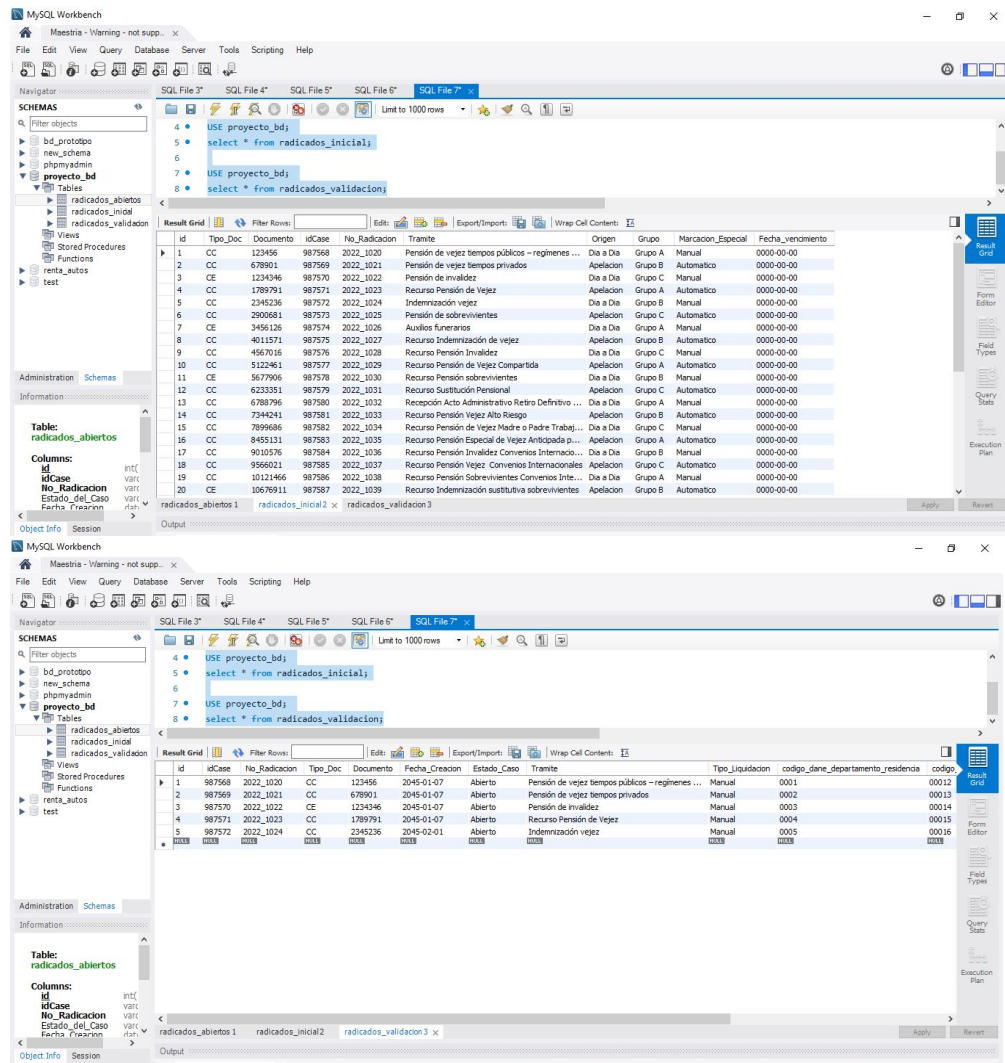
- Escalabilidad, estabilidad y seguridad
- Soporta procedimientos almacenados
- Permite trabajar en modo cliente – servidor
- Seguridad de la información, de acuerdo con los permisos generados por cada rol
- Administración de diferentes servidores de datos
- Tiene un excelente entorno grafico de administración, el cual permite el uso de comandos DDL y DML gráficamente.

4.2 Diagrama de modelo de Datos



4.3 Imágenes de la Base de Datos





4.4 Código SQL – lenguaje de definición de datos (DDL)

CREATE database Proyecto_BD

USE Proyecto_BD;

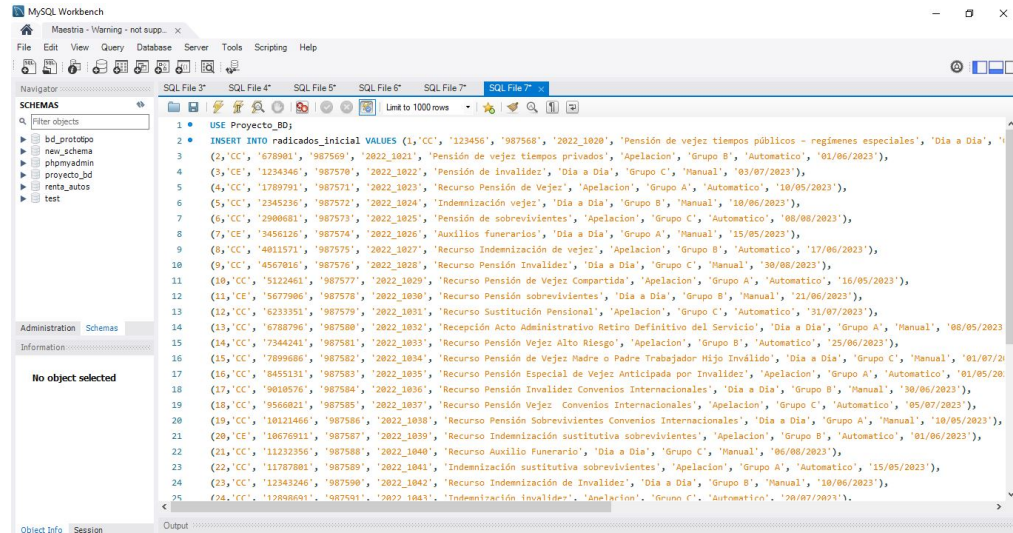
CREATE TABLE Radicados_Inicial(
id INT NOT NULL,

```
Tipo_Doc VARCHAR(10),
Documento VARCHAR(50),
idCase VARCHAR(20) UNIQUE,
No_Radicacion VARCHAR(60) UNIQUE,
Tramite VARCHAR(150),
Origen VARCHAR(150),
Grupo VARCHAR(50),
Marcacion_Especial VARCHAR(150),
Fecha_vencimiento DATE,
PRIMARY KEY (id));
```

```
USE Proyecto_BD;
CREATE TABLE Radicados_Abiertos(
id INT NOT NULL,
idCase VARCHAR(20) UNIQUE,
No_Radicacion VARCHAR(60) UNIQUE,
Estado_del_Caso VARCHAR(60),
Fecha_Creacion DATE,
Proceso_Entrada VARCHAR(60),
Tramite VARCHAR(150),
Tipo_Doc VARCHAR(10),
Documento VARCHAR(50),
Primer_Nombre VARCHAR(150),
Segundo_Nombre VARCHAR(150),
Primer_Apellido VARCHAR(150),
Segundo_Apellido VARCHAR(150),
Codigo_Dane VARCHAR(50),
PRIMARY KEY (id));
```

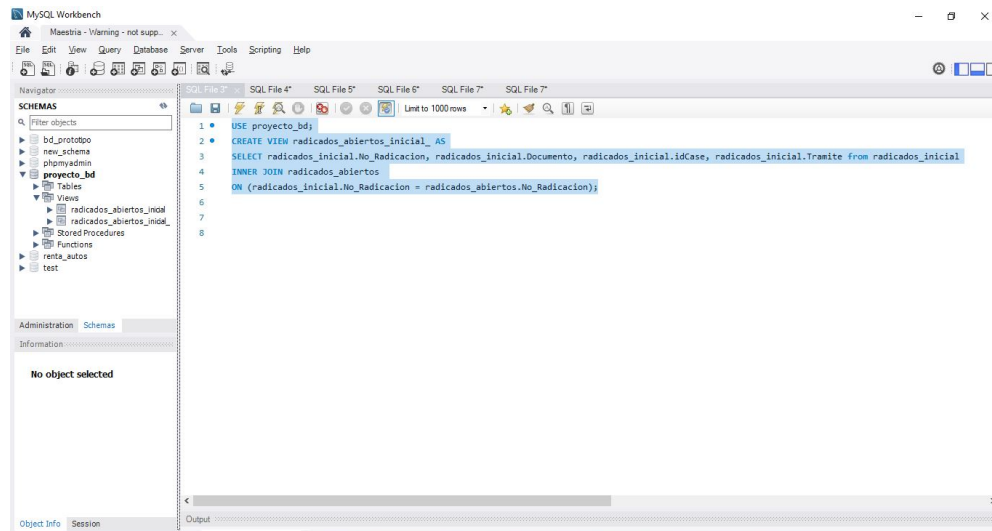
```
USE Proyecto_BD;
CREATE TABLE Radicados_Validacion(
id INT NOT NULL,
idCase VARCHAR(20) UNIQUE,
No_Radicacion VARCHAR(60) UNIQUE,
Tipo_Doc VARCHAR(10),
Documento VARCHAR(50),
Fecha_Creacion DATE,
Estado_Caso VARCHAR(150),
Tramite VARCHAR(150),
Tipo_Liquidacion VARCHAR(150),
codigo_dane_departamento_residencia VARCHAR(50),
codigo_dane_ciudad_municipio_residencia VARCHAR(50),
Estado VARCHAR(150),
Regimen VARCHAR(60),
Tipo_De_Recurso VARCHAR(60),
PRIMARY KEY (id));
```

4.5 Código SQL – manipulación de datos (DML)



4.6 Código SQL + Resultados: Vistas

```
USE proyecto_bd;
CREATE VIEW radicados_abiertos_inicial_ AS
SELECT radicados_inicial.No_Radicacion,
radicados_inicial.Documento,
radicados_inicial.idCase,
radicados_inicial.Tramite
from radicados_inicial
INNER JOIN radicados_abiertos
ON (radicados_inicial.No_Radicacion = radicados_abiertos.No_Radicacion);
```



4.7 Código SQL + Resultados: Triggers

```
USE proyecto_bd;
CREATE TRIGGER relacion_radicados AFTER UPDATE ON radcados_abiertos
FOR EACH ROW SET @No_Radicacion = @No_Radicacion + 1;
```

4.8 Código SQL + Resultados: Triggers

```
USE proyecto_bd;
CREATE TRIGGER relacion_radicados AFTER UPDATE ON radcados_abiertos
FOR EACH ROW SET @No_Radicacion = @No_Radicacion + 1;
```

4.9 Código SQL + Resultados: Funciones

```
USE proyecto_bd;

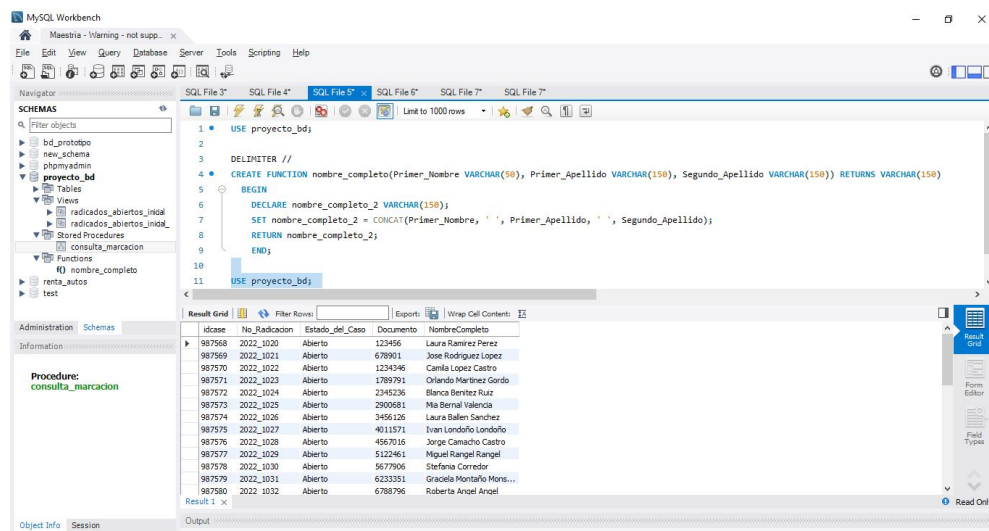
DELIMITER //
CREATE FUNCTION nombre_completo
(Primer_Nombre VARCHAR(50),
```

```

Primer_Apellido VARCHAR(150),
Segundo_Apellido VARCHAR(150))
RETURNS VARCHAR(150)
BEGIN
    DECLARE nombre_completo_2 VARCHAR(150);
    SET nombre_completo_2 =
    CONCAT(Primer_Nombre, ' ',
    Primer_Apellido, ' ', Segundo_Apellido);
RETURN nombre_completo_2;
END;

USE proyecto_bd;
SELECT idcase, No_Radicacion, Estado_del_Caso, Documento, nombre_completo
(Primer_Nombre, Primer_Apellido, Segundo_Apellido)
AS 'NombreCompleto' FROM radicados_abiertos;

```



4.10 Código SQL + Resultados: procedimientos almacenados

```

USE proyecto_bd;
delimiter //
create procedure consulta_marcacion()
begin
select * from radicados_inicial
where Marcacion_Especial = "Automatico";
end //

```

5 Bases de Datos No-SQL

5.1 Diagrama Bases de Datos No-SQL

The image displays two screenshots of the AtlasCluster web interface, showing the 'Collections' tab for a database named 'Clasificacion'. The interface includes a sidebar with navigation options like 'Database', 'Services', and 'Security'. The main content area shows the 'Collections' tab with a search bar and a list of collections. The first screenshot shows the 'Clasificacion.Datos_Personales' collection, displaying a single document with fields like '_id', 'Nombre', 'Apellidos', 'Edad', and 'Departamento'. The second screenshot shows the 'Clasificacion.Radicados_Abiertos' collection, displaying a single document with fields like '_id', 'Radicado', 'id_case', and 'Estado'.

Clasificacion.Datos_Personales

STORAGE SIZE: 20KB LOGICAL DATA SIZE: 16KB TOTAL DOCUMENTS: 1 INDEXES TOTAL SIZE: 20KB

Find Indexes Schema Anti-Patterns Aggregation Search Indexes

Filter Type a query: { field: 'value' } Reset Apply More Options

QUERY RESULTS: 1-1 OF 1

```
{
  "_id": "ObjectID('646184c2cc941887a8bc9b88')",
  "Nombre": "Dario",
  "Apellidos": "Castañeda Ruiz",
  "Edad": "67",
  "Departamento": "Cundinamarca"
}
```

Clasificacion.Radicados_Abiertos

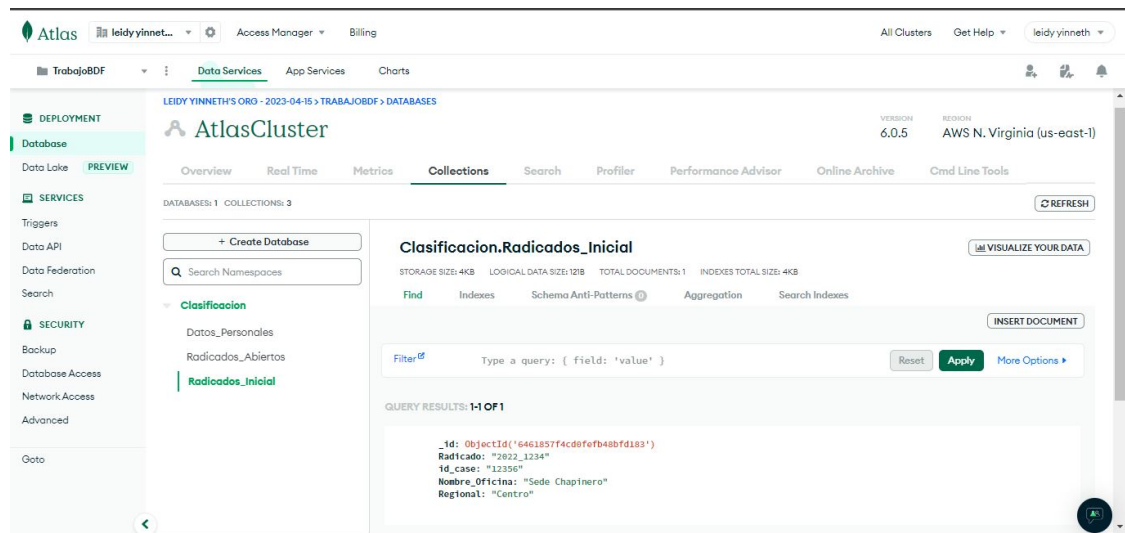
STORAGE SIZE: 20KB LOGICAL DATA SIZE: 8KB TOTAL DOCUMENTS: 1 INDEXES TOTAL SIZE: 20KB

Find Indexes Schema Anti-Patterns Aggregation Search Indexes

Filter Type a query: { field: 'value' } Reset Apply More Options

QUERY RESULTS: 1-1 OF 1

```
{
  "_id": "ObjectID('64618511cc941887a8bc9b81')",
  "Radicado": "2022_1234",
  "id_case": "23456",
  "Estado": "Abierto"
}
```



5.2 SMBD utilizado para la Base de Datos No-SQL

La base de datos No-SQL seleccionada es MongoDB, el cual en lugar de guardar los datos en tablas lo hace en estructuras de datos BSON (similar a JSON) con un esquema dinámico. Teniendo como ventajas:

- * Validación de documentos
- * Motores de almacenamiento integrado
- * Menor tiempo de recuperación ante fallos

6 Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos

6.1 Ejemplo de aplicación de ETL y Bodega de Datos

Se realiza primero la creación de la nueva tabla:

```
USE Proyecto_BD;
CREATE TABLE tabla_transformada(
id INT NOT NULL,
RadicadoA_transformada VARCHAR(60) UNIQUE,
Estado_del_Caso VARCHAR(60),
Proceso_Entrada VARCHAR(60),
Tramite VARCHAR(150),
PRIMARY KEY (id));
```

Luego el proceso para la aplicación se realiza lo siguiente: Selección

```
# Extracción
SELECT No_Radicacion, Estado_del_Caso, Proceso_Entrada, Tramite
from proyecto_bd.radicados_abiertos
where Estado_del_Caso = "Abierto"
```

Transformación

```
# Transformación
USE proyecto_bd
INSERT INTO tabla_transformada (id, RadicadoA_transformada,
Estado_del_Caso, Proceso_Entrada, Tramite)
SELECT Estado_del_Caso, Proceso_Entrada, Tramite, No_Radicacion + 1
FROM radicados_abiertos
WHERE Estado_del_Caso = "Abierto";
```

Carga

```
# Carga
USE proyecto_bd
INSERT INTO tabla_bodega (Estado_del_Caso, Proceso_Entrada,
Tramite, RadicadoA_transformada)
SELECT Estado_del_Caso, Proceso_Entrada, Tramite,
RadicadoA_transformada
FROM tabla_transformada;
```


7 Lecciones Aprendidas

Para el procesamiento de datos, primero se debe comenzar con una estructura bien organizada en la que no existan duplicados en las columnas y se divida adecuadamente cada tabla. Además, es importante que exista una correlación entre los datos. También es necesario saber elegir el motor de base de datos que mejor se adapte a los datos disponibles, teniendo en cuenta si es relacional o no, así como la cantidad de datos y transacciones que se manejarán y las consultas que se realizarán. De esta manera, se evitará perder tiempo y dinero en motores poco eficientes.

8 Referencias

References

- [1] Clasificar con K-Nearest-Neighbor ejemplo en Python. (2018, 10 de Julio). Aprende machine learning. Recuperado de <https://www.aprendemachinelearning.com/clasificar-con-k-nearest-neighbor-ejemplo-en-python/>
- [2] Una guía para el análisis de componentes principales (PCA) para el aprendizaje automático. (2022, 02 de Abril). Keboola. Recuperado de <https://www.keboola.com/blog/pca-machine-learning>
- [3] Informe Gestión 2022. (2023, Febrero). Colpensiones. Recuperado de <https://www.colpensiones.gov.co/publicaciones/4917/colpensiones-en-cifras-diciembre-2022/>
- [4] ¿Quiénes somos?. (2012, 02 de Febrero). Colpensiones. Recuperado de <https://www.colpensiones.gov.co/publicaciones/113/quienes-somos/>
- [5] ¿Qué es el RPM?. (2012, 02 de Febrero). Colpensiones. Recuperado de <https://www.colpensiones.gov.co/pensiones/publicaciones/120/que-es-el-rpm/>
- [6] Oracle PL/SQL Language Pocket Reference: A Guide to Oracle's PL/SQL Language Oracle Fundamentals. (2013). Foster Provost, Tom Fawcett. Recuperado de <https://libgen.li/edition.php?id=136796079>
- [7] Bases de datos en Mysql. (2003). Luis Alberto Casillas Santillán Marc Gibert Ginestà Óscar Pérez Mora. Recuperado de https://www.dataprix.com/files/UOC_OpenSource_bases_de_datos_en_mysql.pdf