Proyecto de Bases de datos para el análisis de la Encuesta Nacional del Uso del Tiempo

Cindy Catalina Vásquez Guerrero 1

Facultad de Ingeniería y Ciencias Básicas,
Universidad Central
Maestría en Analítica de Datos
Curso de Bases de Datos
Bogotá, Colombia

1 cvasquezg1@ucentral.edu.co,

November 12, 2022

Contents

1	Introducción	3
2	Características del proyecto de investigación 2.1 Titulo del proyecto de investigación	4 4 4 4
3	Reflexiones sobre el origen de datos e información 3.1 ¿Cuál es el origen de los datos e información ?	5 5 6 6
4	Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos) 4.1 Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto	e 7

	4.2	Diagrama modelo de datos	8
	4.3	Imágenes de la Base de Datos	9
	4.4	Código SQL - lenguaje de definición de datos (DDL)	9
	4.5	Código SQL - Manipulación de datos (DML)	11
	4.6	Código SQL + Resultados: Vistas \dots	13
	4.7	Código SQL + Resultados: Triggers $\dots \dots \dots \dots$	14
	4.8	Código SQL + Resultados: Funciones $\dots \dots \dots$	15
	4.9	Código SQL + Resultados: procedimientos almacenados $\ . \ . \ .$.	15
5	Bas	es de Datos No-SQL	16
	5.1	Diagrama Bases de Datos No-SQL	16
	5.2	SMBD utilizado para la Base de Datos No-SQL	17
6	Lec	ciones aprendidas	18
7	Bib	liografía	19

1 Introducción

La Encuesta Nacional del Uso del Tiempo ENUT es una investigación que busca generar información sobre el tiempo dedicado por las personas de 10 años o más a actividades de trabajo remunerado, no remunerado y actividades personales. En la edición 2016-2017 la encuesta se estructuró alrededor de 733 preguntas y recopiló la información de 44.999 hogares (DANE, 2018). Mientras que para la recolección 2020-2021 se definieron 767 preguntas y se recopilaron datos de 49.519 hogares (DANE, 2021). Ambas encuestas cuentan con representatividad en la mayoría del territorio nacional, pues se toman datos en los centros poblados, cabeceras municipales y rural disperso en seis regiones del país Bogotá, San Andrés, Caribe, Oriental, Central, y Pacífica. La cantidad de datos recopilados en ambas encuestas requiere de un manejo especial con el objetivo de permitir el uso eficiente de la información de los hogares durante las fases de Extracción, Transformación y Carga. Teniendo en cuenta este propósito, es necesario estructurar una base de datos que permita la organización de la información a través de tablas y de relaciones entre estas. De esta forma se puede mejorar la capacidad de modificación de los datos y su consulta, para futuros ejercicios estadísticos que permitan una mejor comprensión de los datos.

2 Características del proyecto de investigación

El tiempo dedicado al trabajo no remunerado es una de las desigualdades con mayor impacto en la vida cotidiana de las mujeres. Tradicionalmente, se piensan como tareas que deben desarrollar las mujeres de forma complementaria a su jornada laboral o incluso dedicándose por completo a ellas sin recibir ninguna compensación. La ENUT permite realizar la caracterización del tiempo dedicado por la población a actividades de trabajo no remunerado teniendo en cuenta características particulares de los hogares como quien es el jefe de hogar, sexo, estado civil, edad, número de hijos entre otras. Es por esta razón que el análisis de la encuesta permitirá describir mejor el problema de interés. Caracterizar el uso desigual del tiempo al trabajo no remunerado entre hombres y mujeres permite representar mejor las condiciones inequitativas asociadas a la dedicación a tareas del hogar relacionadas con el cuidado. La economía del cuidado propone reconocer de manera formal estas actividades. Reconociendo que "invertir en la economía del cuidado contribuye a la diversificación de la estructura productiva, sin trasgredir los límites ecológicos para la reproducción, lo que a su vez permite gestionar de forma más eficiente el patrimonio natural y garantizar la sostenibilidad de la vida" (CEPAL, 2020).

2.1 Titulo del proyecto de investigación

BASE DE DATOS PARA EL ANÁLISIS COMPARATIVO DEL USO DEL TIEMPO EN TRABAJO NO REMUNERADO A PARTIR DE LA ENCUESTA NACIONAL DEL USO DEL TIEMPO ENUT 2016-2017 Y 2020-2021.

2.2 Objetivo general

• Realizar un análisis de los datos de la ENUT 2016-2017 y 2020-2021 sobre el tiempo dedicado al trabajo no remunerado.

2.2.1 Objetivos específicos

- 1. Realizar la estructuración de los datos y definir las variables más importantes para el análisis.
- 2. Hacer una exploración inicial de los datos para comprender las características de la encuesta.
- 3. Implementar técnicas de análisis a partir de modelos no supervisados con el objetivo de reducir la dimensionalidad y mejorar la comprensión de la relación entre variables.
- 4. Implementar técnicas de análisis a partir de modelos supervisados con el objetivo de predecir una variable objetivo.

2.3 Alcance

El alcance de la propuesta es definir un sistema de bases de datos que permita el análisis de los resultados de la ENUT en las versiones 2016-2017 y 2020-2021, para la estructuración de la base de datos se utiliza la información relacionada con el trabajo no remunerado con el fin de segmentar el análisis y construir una base de datos que de soporte a un usuario que corre en un solo computador, es decir que se clasifica como una "single-user desktop database".

2.4 Pregunta de investigación

¿Al comparar las dos encuestas se pueden observar cambios en la cantidad de tiempo dedicado al trabajo no remunerado distribuido por sexo?

2.5 Hipótesis

La pandemia intensifico la carga de trabajo no remunerado en las mujeres.

3 Reflexiones sobre el origen de datos e información

La Encuesta Nacional del Uso del Tiempo es un instrumento que permite la recolección de información valiosa a cerca de la composición del hogar, el ámbito de salud y educación, el cuidado de los niños y niñas, la fuerza de trabajo y el uso del tiempo. Esta encuesta se recopila en la mayoría del territorio nacional, lo que brinda un marco de referencia importante en materia de bienestar de los hogares. Particularmente, con respecto al tiempo destinado a las labores no remuneradas como las tareas domésticas, el cuidado de los hijos o familiares, los tramites familiares como citas médicas o pago de facturas. Adicionalmente, se analiza el tiempo dedicado al descanso y a las actividades de ocio. Estas características de la encuesta la convierten en un insumo fundamental para medir el bienestar de las familias comparativamente antes y después de la pandemia.

La primera versión de la encuesta utilizada tiene un total de 733 preguntas y recopiló la información de 44.999 hogares, alcanzando una recolección de 146.190 individos. Por su parte, la segunda versión de la encuesta definió 767 preguntas y se recopilaron datos de 49.519 hogares, con un total de 126.753 registros individuales.

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

El Departamento Administrativo Nacional de Estadística DANE a través de su sitio de micro-datos ha brindado acceso a la ciudadanía a los resultados de encuestas que se desarrollan para la caracterización del país en diferentes áreas como Economía, Sociedad y Territorio. En el ámbito de Sociedad en la sección de Género y relacionado con Economía del Cuidado se publican los resultados de la ENUT, ya que previamente han sido limpiados y organizados, en la página se puede encontrar el diccionario de datos y cada uno de los capítulos de la encuesta. Los capítulos definidos son:

- 1. Capitulo b. Condiciones de la vivienda
- 2. Capitulo c. Datos del hogar
- 3. Capitulo d. Composición del hogar
- 4. Capitulo e. Salud
- 5. Capitulo f. Cuidado de los niños y niñas menores de 5
- 6. Capitulo g. Educación
- 7. Capitulo h. Fuerza de trabajo clasificación, uso del tiempo e ingreso
- 8. Capitulo i. Uso del tiempo

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

La información no presenta ninguna consideración legal o ética restrictiva, corresponde más con el compromiso del investigador con informar a los lectores de la investigación el origen de la información y el uso que se le ha dado a la información publicada.

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

Al ser datos publicados por el DANE en su portal web ya han sido revisados y avalados en cuanto a la calidad y consolidación de los datos, cada capítulo contiene un diccionario de datos, junto con los datos en formato CSV y además se encuentra publicada la documentación de la encuesta, como su metodología y resumen de los principales resultados encontrados.

Adicionalmente se han encontrado datos complementarios para la medición del Tiempo total de trabajo en el Observatorio de Igualdad de Género de América Latina y el Caribe de la CEPAl (Cepal, s.f.) que permite hacer comparaciones con otros países de latinoamérica.

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

Al implementar un sistema de bases de datos se espera que se mejore la eficiencia en la consulta de los datos contenidos en la encuesta, dado que al manejar archivos independiente para cada capítulo puede resultar poco eficiente la consulta y modificación d de los datos, temas que se podrían resolver a través de un sistema de base de datos que organice la información en tablas y permita el acceso a los datos a través de consultas personalizadas.

De acuerdo con el diseño del modelo de datos implementados se espera hacer una carga efectiva de los registros asociados a la primera versión de la encuesta (2016-2017) que contiene tres tablas:

- 1. Primera tabla Tiempo: 29 columnas con 146.190 registros
- 2. Segunda tabla Hogar: 15 columnas con 146.190 registros
- 3. Tercera tabla Vivienda: 10 columnas con 44.999 registros

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

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

Como sistema manejador de base de datos se ha seleccionado Mysql Workbench que permite recolectar datos de manera eficiente (Silberschatz et al., 2005).

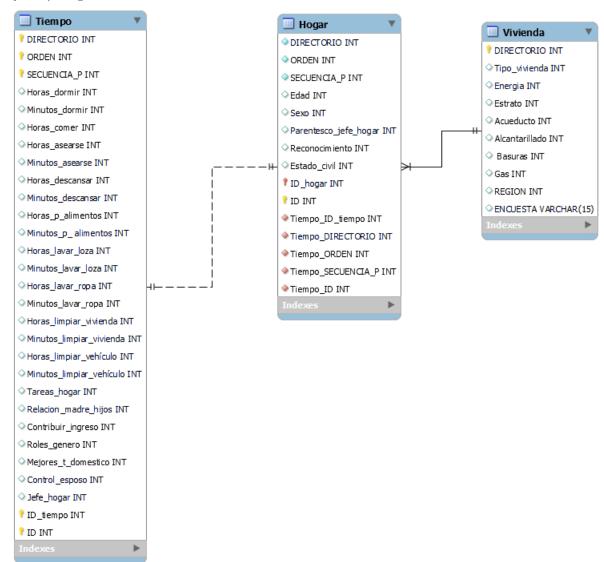
La versión utilizada de Mysql y Workbench es 8.0, esta versión ha mejorado aspectos para la interacción con la interfaz del sistema, funcionalidades específicas para Windows y sintaxis recursiva de SQL.

Algunos de los beneficios del uso de este SMBD son:

- Estructurar un esquema de datos que permite su manipulación
- Definir la estructura de almacenamiento y metodología de acceso
- Compartir la base de datos de manera organizada
- Delimitar rutinas de mantenimiento para garantizar la eficiencia de la base de datos
- Permitir el acceso y la modificación de las tablas y los registros

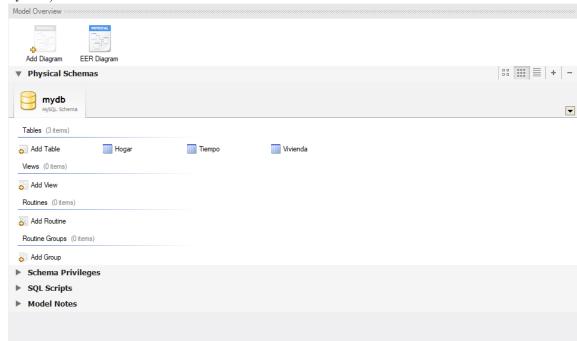
4.2 Diagrama modelo de datos

(Gráfico 1) Diagrama:



4.3 Imágenes de la Base de Datos

(Gráfico 2) Base de datos:



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

(Tabla 1) Tabla Vivienda:

```
DROP TABLE IF EXISTS 'Vivienda';

CREATE TABLE 'Vivienda' (
    'DIRECTORIO' int NOT NULL,
    'Tipo_vivienda' int NOT NULL,
    'Energia' int NOT NULL,
    'Estrato' int,
    'Acueducto' int NOT NULL,
    'Alcantarillado' int NOT NULL,
    'Basuras' int NOT NULL,
    'Gas' int NOT NULL,
    'REGION' int NOT NULL,
    'ENCUESTA' varchar(15) NOT NULL,
    PRIMARY KEY ('DIRECTORIO')
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
```

(Tabla 2) Tabla Hogar:

```
DROP TABLE IF EXISTS `Hogar`;
CREATE TABLE 'Hogar' (
      `ID` int NOT NULL,
     `DIRECTORIO` int NOT NULL,
    `ORDEN` int NOT NULL,

`SECUENCIA_P` int NOT NULL,
    `Edad` int NOT NULL,
`Sexo` int NOT NULL,
    `Parentesco_jefe_hogar` int NOT NULL,
`Reconocimiento` int NOT NULL,
   `Reconocimiento` int NOT NULL,
`Estado_civil` int,
PRIMARY KEY (`ID`),
KEY `DIRECTORIO` (`DIRECTORIO`),
KEY `ORDEN` (`ORDEN`),
KEY `SECUENCIA_P` (`SECUENCIA_P`)
ENGINE=InnoDB DEFAULT CHARSET=latin1;
```

(Tabla 3) Tabla Tiempo:

```
CREATE TABLE `Tiempo` (
'ID' int NOT NULL,

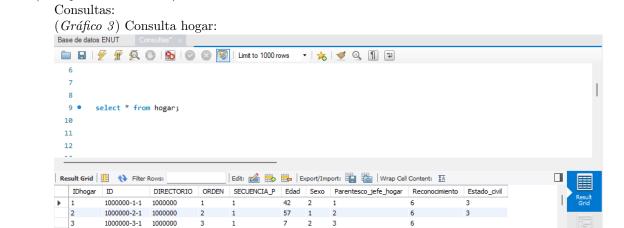
'DIRECTORIO' int NOT NULL,

'ORDEN' int NOT NULL,

'SECUENCIA_P' int NOT NULL,

'Horas_dormir' int,
 `Minutos_dormir` int,
 `Horas_comer` int,
 `Minutos_comer` int,
`Horas_asearse` int,
 `Minutos_asearse` int,
`Horas_descansar` int,
 `Minutos_descansar` int,
`Horas_p_alimentos` int,
`Minutos_p_ alimentos` int,
 `Horas_lavar_loza` int,
 `Minutos_lavar_loza` int,
 'Horas_lavar_ropa' int,
'Minutos_lavar_ropa' int,
'Horas_limpiar_vivienda' int,
'Minutos_limpiar_vivienda' int,
'Minutos_limpiar_vivienda' int,
 `Horas_limpiar_vehículo` int,
`Minutos_limpiar_vehículo` int,
 `Tareas_hogar` int,
`Relacion_madre_hijos` int,
 `Contribuir_ingreso` int,
 `Roles_genero` int,
 `Mejores_t_domestico` int,
  Control_esposo` int,
  Jefe_hogar` int,
PRIMARY KEY (`ID`),
KEY `DIRECTORIO` (`DIRECTORIO`),
   KEY `ORDEN` (`ORDEN`),
```

4.5 Código SQL - Manipulación de datos (DML)



67

57



1000004-1-1 1000004 1000004-2-1 1000004

1000000

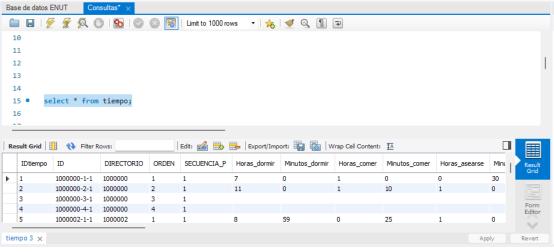
1000002

1000000-4-1

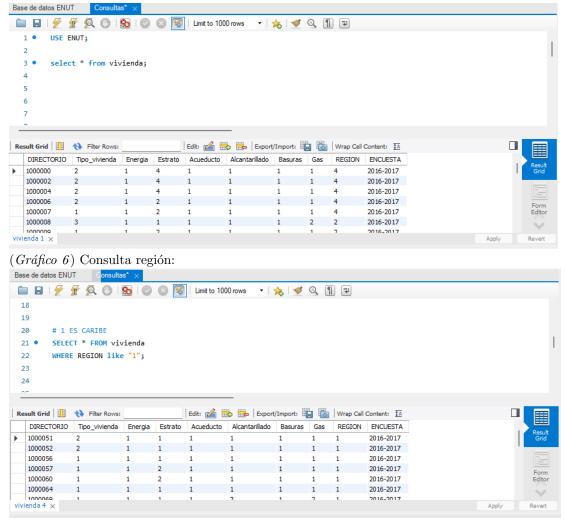
1000002-1-1

hogar 2 ×

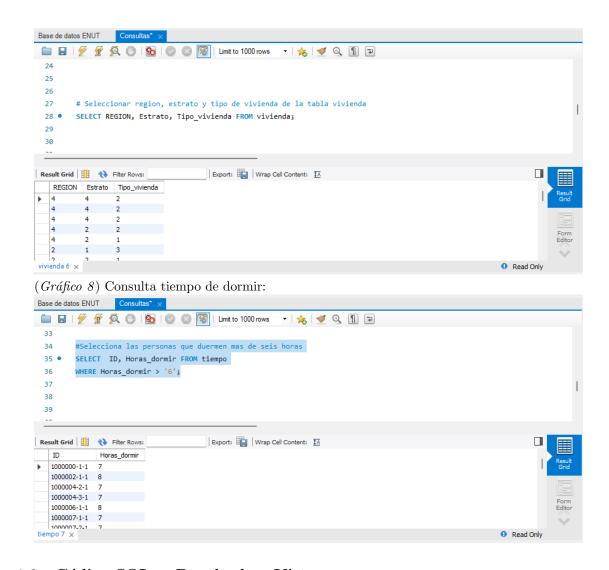
(Manipulación de datos)



(Gráfico 5) Consulta vivienda:

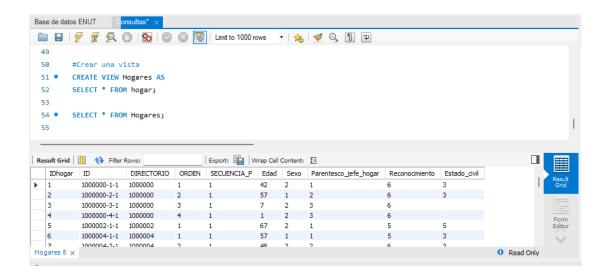


(Gráfico 7) Consulta características de la vivienda:



4.6 Código SQL + Resultados: Vistas

(Vistas)



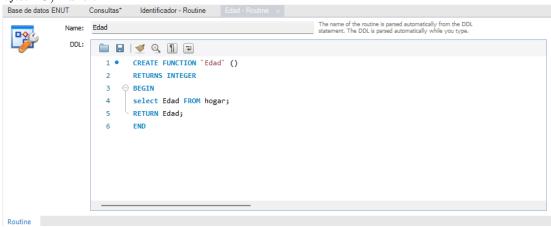
4.7 Código SQL + Resultados: Triggers

(Gráfico 9) Trigger:

```
Base de datos ENUT Consultas* × Identificador - Routine
    iii II | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/2 | 1/
         55
         56
         57
         58
         59
         60
         61
                                                     DELIMITER $$
         62 • CREATE TRIGGER Identificador
         63
                                                      BEFORE UPDATE ON vivienda
                                                      FOR EACH ROW
         65 ⊝ BEGIN
         66
                                                          SELECT ID FROM vivienda;
                                           END$$
         67
                                                        DELIMITER;
         68
         69
         70
```

4.8 Código SQL + Resultados: Funciones

(Gráfico 10) Función:



${\bf 4.9}\quad {\bf C\'odigo~SQL+Resultados:~procedimientos~almacenados}$

(Gráfico 11) Prodecimiento:

```
Name: Identificador

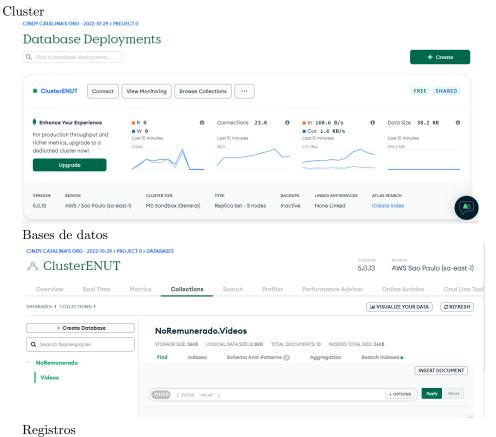
The name of the routine is parsed automatically from the DDL statement. The DDL is parsed automatically while you type.

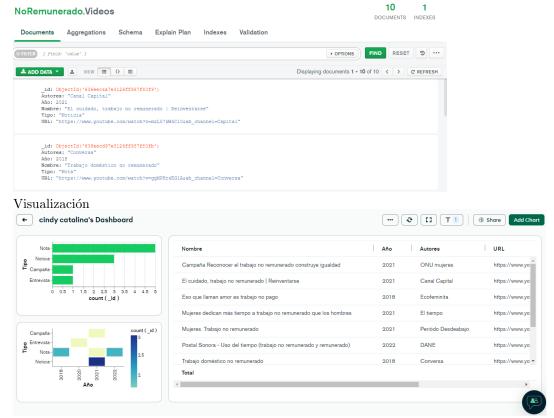
1 • CREATE DEFINER="root"@'localhost" PROCEDURE 'Identificador'()
2 BEGIN
3 SELECT ID FROM Hogar;
4 END
```

Bases de Datos No-SQL 5

Para la estructuración de un modelo que se adapte a las necesidades de la investigación se ha desarrollado un clúster en MongoDB Atlas para registrar los videos de entrevistas, noticias, documentales, notas y otras herramientas visuales alrededor de trabajo no remunerado.

5.1 Diagrama Bases de Datos No-SQL





Tablero:

https://charts.mongodb.com/charts-project-0-tynax/public/dashboards/635d3402-fb36-4cc5-8e95-8ea226783a22

5.2 SMBD utilizado para la Base de Datos No-SQL

Se ha utilizado MongoDB para el manejo de Bases de Datos No SQL, este SMBD tiene las ventajas de ser ligero, tener un buen rendimiento, ser de bajo costo y tener una gran documentación.

Algunas de las funcionalidades más importantes por destacar son: nuevos motores de funcionamiento en Clúster para adopción de un modelo Cloud y el soporte de los principales lenguajes de programación.

6 Lecciones aprendidas

- Los procesos de selección, limpieza y estandarización de los datos son momentos clave que ahorrarán tiempo.
- Garantizar el cumplimiento de las reglas de normalización permite una carga de datos eficiente.
- La carga masiva requiere de una estructura de datos eficiente con el objetivo de no sobrecargar los recursos disponibles.
- El modelo entidad relación debe ser puesto a prueba con el objetivo de validar su lógica y mejorarlo en caso de reduccir el exceso de la información.
- Las bases de datos sql permiten un uso eficiente de la información que se puede estructurar en tablas, otras necesidades de almacenamiento requieren soluciones innovadoras como lo son los SMBD tipo no sql.

7 Bibliografía

CEPAL. (2020). La economía del cuidado como acelerador del cambio estructural con igualdad. Notas Para La Igualdad, 30 (Observatorio de Igualdad de Género de America Latina y el Caribe). https://oig.cepal.org/sites/default/files/no30_esp_-_economia_del_cuidado.pdf

CEPAL. (s.f). Observatorio de Igualdad de Género de América Latina y el Caribe. https://oig.cepal.org/es/indicadores/tiempo-total-trabajo

Cristina Benavente Alejandra Valdés B, M. R. (2014). Desarrollo Social Políticas públicas para la igualdad de género Un aporte a la autonomía de las mujeres. www.cepal.org/es/suscripciones

DANE. (2018). Boletín técnico Encuesta Nacional de Uso del Tiempo (ENUT) 2016-2017. 28. https://www.dane.gov.co/files/investigaciones/boletines/ENUT/Bol_ENUT_2016_2017.pdf

DANE, D. A. N. E. (2021). Boletín Técnico Encuesta Nacional del Uso del Tiempo (ENUT) 2020-2021. 1-59.

Silberschatz, A., Korth, H., SudarshanS. (2005). Database management. In Disability and Rehabilitation (Vol. 27, Issue 16). https://doi.org/10.1080/09638280500030605