

Identificación de consumos de agua atípicos en el Municipio de Fusagasugá Cundinamarca

César Chavarriaga B.¹

Facultad de Ingeniería y Ciencias Básicas

Universidad Central

Maestría en Analítica de Datos

Curso de Bases de Datos

Bogotá, Colombia

November 24, 2022

Contents

1	Introducción	3
2	Características del proyecto de investigación	3
2.1	Titulo del proyecto de investigación	3
2.2	Objetivo general	3
2.2.1	Objetivos especificos	3
2.3	Alcance	4
2.4	Pregunta de investigación	4
2.5	Hipotesis	4
3	Reflexiones sobre el origen de datos e información	5
3.1	¿Cual es el origen de los datos e información ?	5
3.2	¿Cuales son las consideraciones legales o eticas del uso de la información?	5
3.3	¿Cuales son los retos de la información y los datos que utilizara en la base de datos en terminos de la calidad y la consolidación?	5
3.4	¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto?	6
4	Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)	7
4.1	Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto	7
4.2	Diagrama modelo de datos	7

4.3	Imágenes de la Base de Datos	8
4.4	Código SQL - lenguaje de definición de datos (DDL)	10
4.5	Código SQL - Manipulación de datos (DML)	11
4.6	Código SQL + Resultados: Vistas	12
4.7	Código SQL + Resultados: Triggers	13
4.8	Código SQL + Resultados: Funciones	13
4.9	Código SQL + Resultados: procedimientos almacenados	14
5	Bases de Datos No-SQL	18
5.1	Diagrama Bases de Datos No-SQL	18
5.2	SMBD utilizado para la Base de Datos No-SQL	19
6	Bibliografía	23

1 Introducción

El consumo de agua como servicio público es una de las necesidades básicas de cualquier comunidad. Actualmente las tarifas de medición del consumo se componen de un cargo fijo, un básico y un complementario; la unidad de medición de este consumo es el metro cúbico (m³). El municipio de Fusagasugá, una comunidad aledaña de Bogotá, con unas condiciones climatológicas más cálidas y con precipitaciones menos drásticas se destaca como un lugar óptimo para vacacionar; generando en una población flotante que debe ser tenida en cuenta especialmente por sus altos consumos en los lapsos de descanso.

Estos altos consumos pueden ser reales, es decir si hubo tal consumo, o no real que corresponde a fallas producidas por fugas, La Comisión de Regulación de Agua Potable y Saneamiento Básico CRA, define que existen las pérdidas de agua en los sistemas de acueducto se pueden clasificar, de forma general, en pérdidas técnicas y pérdidas comerciales. “Pérdidas técnicas: corresponden a aquellas que se presentan por fugas por fallas en los elementos de la red, tales como conductos, conexiones y tanques de almacenamiento;

Pérdidas comerciales: corresponden a consumos ilegales, consumos no medidos, consumos medidos no facturados y a errores de micromedición y facturación.”

Estos casos de consumos atípicos se enmarcan en los aumentos o reducciones, excesivas o repentinas del consumo, que comparado con el promedio histórico de los últimos seis periodos de facturación distan del promedio normal de consumo.

2 Características del proyecto de investigación

2.1 Título del proyecto de investigación

IDENTIFICACIÓN DE CONSUMOS DE AGUA ATÍPICOS EN EL MUNICIPIO DE FUSAGASUGÁ CUNDINAMARCA.

2.2 Objetivo general

Diseñar y construir un modelo de clasificación que permita realizar una precrítica efectiva en el mismo periodo de facturación del consumo de agua, orientada a disminuir las Pérdidas de Agua en un 10 %.

2.2.1 Objetivos específicos

- Identificar una muestra del ciclo de facturación (zona geográfica que se abastece de una de las plantas de tratamiento de agua), que presenta la mayor cantidad de consumos atípicos, en los dos últimos años.
- Establecer los perfiles de consumo de la muestra; de los dos últimos años, en el cual se tenga en cuenta la temporada del periodo de facturación (vacaciones, semana santa, feriados etc).

- Diseñar el modelo de clasificación que realice la precrítica a través de los perfiles de consumo.

2.3 Alcance

A través del análisis de los dos últimos años de facturación, se espera identificar el perfil de consumo atípicos, orientado a disminuir el porcentaje del índice de pérdidas de agua no contabilizada, teniendo en cuenta que el presente escrito solo se enfocara en uno de los componentes que afectan este índice, como son las pérdidas comerciales, de las desviaciones significativas que se detectan en la micromedición.

2.4 Pregunta de investigación

Teniendo en cuenta que el porcentaje de Pérdidas Comerciales, para los países en desarrollo están en un 40 % del 100 % de Pérdidas de Agua ¿En qué medida las pérdidas comerciales, de las desviaciones significativas que se detectan en la micromedición, influyen en el porcentaje de agua no contabilizada en el Municipio de Fusagasugá?

2.5 Hipotesis

El mayor índice de Pérdidas de Agua no Contabilizada se encuentra en las Pérdidas Comerciales, presentes en los consumos atípicos que se detectan en la micromedición y se consideran desviaciones significativas, cuyo origen de consumo se da en el predio y se asocia por el consumo que supera el promedio del mismo uso en los porcentajes establecidos por la CRA para el uso y estrato, cuyo origen es variado como daños en la red interna, manipulación de la acometida, etc.

3 Reflexiones sobre el origen de datos e información

La información utilizada en el presente ejercicio, hace parte de la base de datos histórica de la entidad EMSERFUSA ESP de los años 2020 y 2021, de la cual parte de ella se publica en la página de <https://www.datos.gov.co/> y la cual fue actualizada por última vez el 29 de marzo del 2021

[https://www.datos.gov.co/Vivienda-Ciudad-y-Territorio/
Consumo-y-facturaci-n-de-Agua-de-la-Empresa-de-Ser/jxfu-nayp/](https://www.datos.gov.co/Vivienda-Ciudad-y-Territorio/Consumo-y-facturaci-n-de-Agua-de-la-Empresa-de-Ser/jxfu-nayp/).

Sin embargo, es de aclarar que los datos relacionados objeto de estudio no comprometen la confidencialidad particular y de la entidad, dado que ya se han realizado publicaciones anteriores de periodos de facturación de los datos utilizados, a través de la plataforma publica de datos abiertos.

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

Como Empresa prestadora de Servicios Públicos Domiciliarios de Acueducto, Alcantarillado y Aseo en el Municipio de Fusagasugá, realiza la facturación de estos servicios a través de su CRM comercial, el cual es encargado de realizar la captura, procesamiento y alistamiento de la información, orientado a la administración y sostenibilidad del servicio y la entidad.

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

La División Comercial, como fuente de información, autoriza su uso, solo para el objeto de estudio expuesto, mientras se conserve los principios de Habeas Data y anonimato del titular de la información y esta no sea utilizada en un ámbito diferente al del presente objeto de estudio.

Con el fin de proteger los datos personales de los titulares de la información y con el fin de que los datos no puedan ser atribuidos a un individuo, se realiza un tratamiento de los datos para tal fin.

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

En el proceso de facturación se realiza la actualización de consumo de los servicios prestados, sin embargo, aunque el dato corresponde a una variable continua, debido al proceso de precritica, el valor ingresado puede ser un valor real o promedio, influenciando el valor de consumo objeto del análisis. Estos valores promedios, se asocian a las observaciones de lectura, las cuales influyen si, el dato se promedia o se ingresan por diferencia, sin embargo, es sensible al criterio del lector, por tal motivo para el presente análisis no se incluirá estas observaciones con el fin de que no influyan en el resultado de este.

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

Con el uso de un SGBD, se espera realizar dos labores; la primera será brindar una estructura acorde para el análisis descriptivo de la información orientada al análisis del fenómeno. La segunda brindara un espacio de almacenamiento para la toma de lectura en tiempo real orientado a la visualización de la información para la toma de decisiones antes de generar la factura. Esto último permitirá disminuir el volumen de PQR que se presentan en la entidad debido a la reducción de tiempos para evaluar las inspecciones generadas por las alertas.

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

El SMBD utilizado es MySQL 8.0, debido a que es un software libre y ofrece altas prestaciones de fiabilidad, velocidad y almacenamiento. Se crearan tres tablas, facturación, tipo de documento y uso.

La Tabla de Facturación, se considera la principal, ya que contiene los datos relevantes para el análisis. Cuenta con: 549079 registros, con información del año 2021 desde el mes de enero hasta noviembre.

4.2 Diagrama modelo de datos

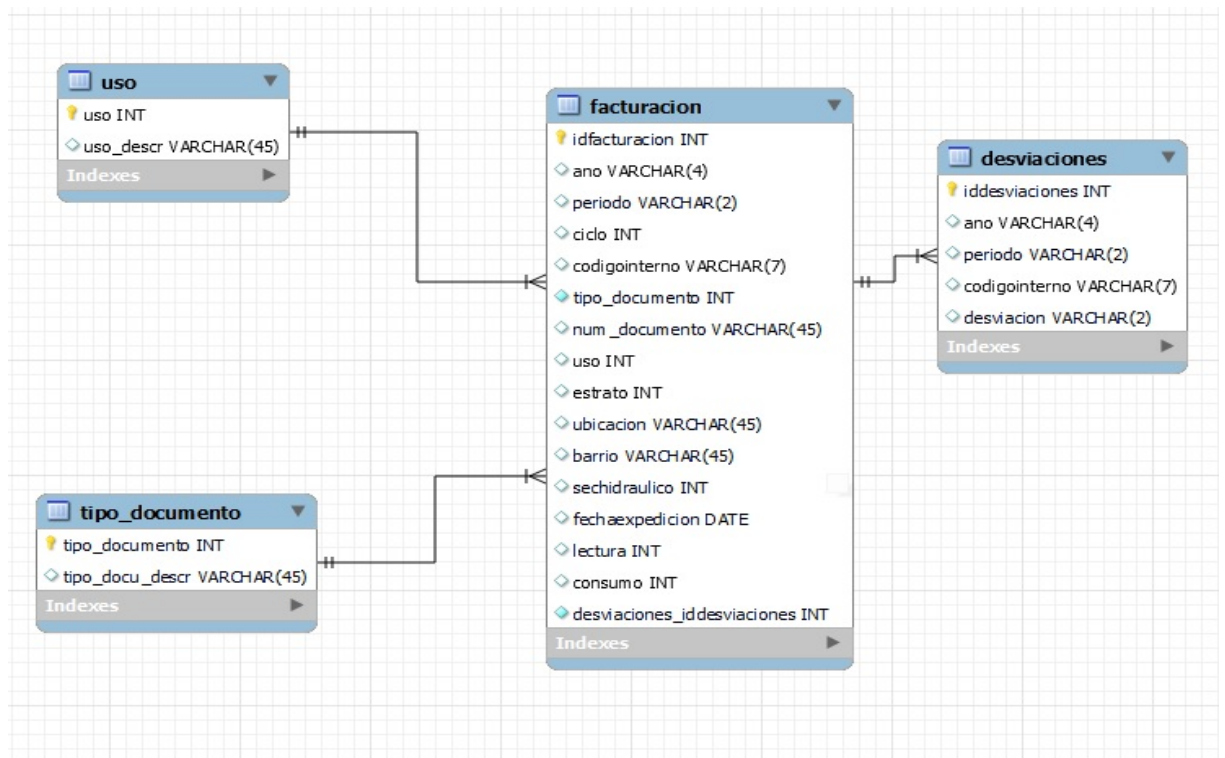


Figura 1: Modelo Entidad Relación de la Base de datos

4.3 Imágenes de la Base de Datos

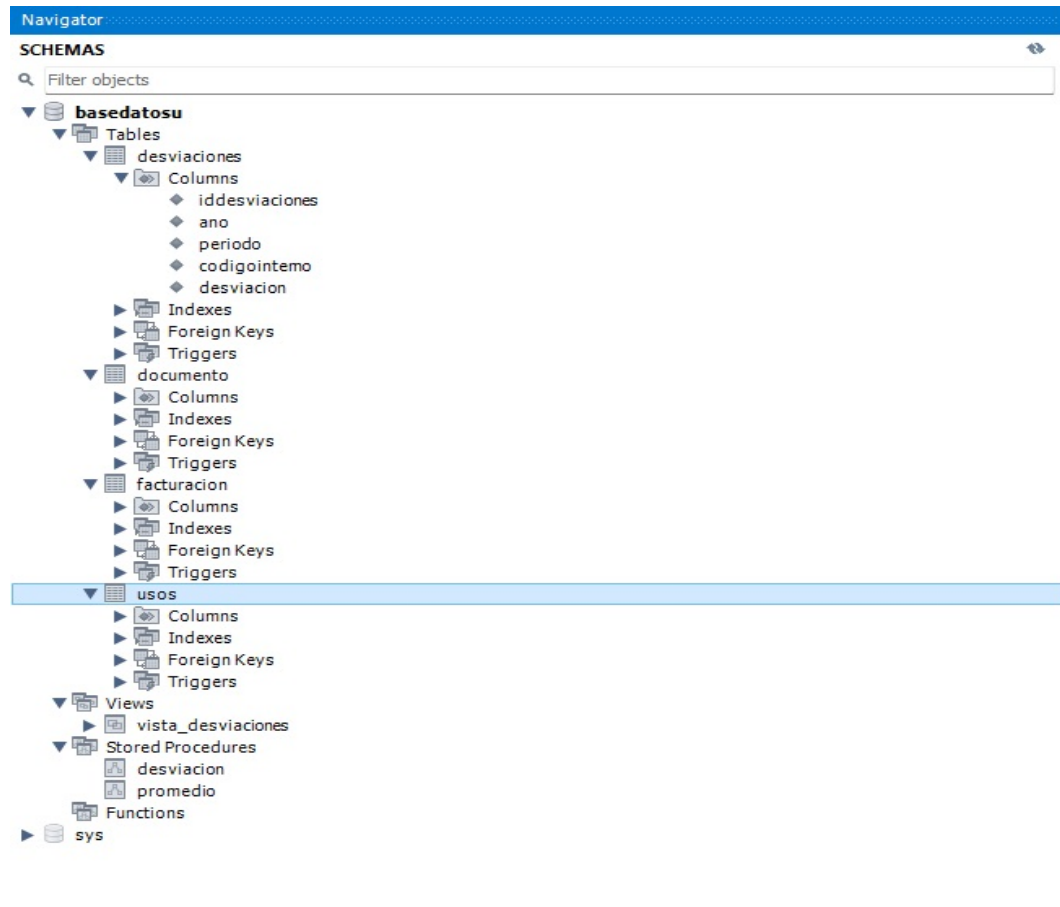


Figura 2: Base de datos

	tipo_docu	tipo_docu_descripcion
▶	1	TARJETA DE IDENTIDAD
	2	CEDULA CIUDADANIA
	3	NIT
	4	CEDULA EXTRANJERIA
	5	PASAPORTE
	6	SIN IDENTIFICAR

Figura 3: Tabla documento

	id_usos	uso_descripcion
▶	1	RESIDENCIAL
	10	INDUSTRIAL
	11	COMERCIAL
	12	OFICIAL

Figura 4: Tabla usos

	id_facturacion	ano	periodo	cido	codigointerno	tipo_doc	documento	uso	estrato	ubicacion	barrio	sechidraulico	fechaexpfactura	lectura	consumo
▶	0	2021	1	2	1207636	2	99991207636	1	4	URBANO	MANILA	704	2019-04-04	228	6
	1	2021	1	2	1263845	2	99991263845	1	4	URBANO	SANTA ANITA	704	2020-03-11	2063	6
	2	2021	1	2	1060565	2	99991060565	1	4	URBANO	MANDALAY	703	2019-08-27	130	9
	3	2021	1	2	1186352	2	99991186352	1	4	URBANO	COOTRANSFUSA	703	2020-08-06	30	9
	5	2021	1	2	1404391	2	99991404391	1	4	URBANO	PIEDRAGRANDE	704	2019-03-21	1190	6
	6	2021	1	2	1153154	2	99991153154	1	3	URBANO	BALMORAL	704	2020-03-06	434	15
	8	2021	1	2	1073220	2	99991073220	1	4	URBANO	MANILA	704	2016-06-20	1494	5
	11	2021	1	2	1084615	2	99991084615	1	4	URBANO	BALMORAL	704	2017-02-06	2771	29
	12	2021	1	2	1072131	2	99991072131	1	4	URBANO	VILLA CLARA	704	2017-02-01	404	3
	13	2021	1	2	1066257	2	99991066257	1	4	URBANO	PIEDRAGRANDE	704	2019-07-02	892	17
	15	2021	1	2	2462461	2	99992462461	1	4	URBANO	SANTA ANITA	704	2020-03-06	584	12
	16	2021	1	2	2423155	2	99992423155	1	3	URBANO	POTOSI	703	2019-05-06	535	7

Figura 5: Tabla facturacion





Result Grid				 Filter Rows:		Edit: 		
	iddesviaciones	ano	periodo	codigointerno	desviacion			
	10	2021	3	1200864	0			
	11	2021	3	2568187	0			
	12	2021	3	2568187	0			
	14	2021	3	1020320	0			
	15	2021	3	1243221	0			
	17	2021	3	1063619	0			
	18	2021	3	1084615	-1			
	19	2021	3	2440110	0			
	20	2021	3	1547173	0			
	21	2021	3	2440110	0			
	22	2021	3	1084615	-1			

Figura 6: Tabla desviaciones

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

```
DDL for bdproyecto.facturacion

1 CREATE TABLE `facturacion` (
2   `id_factura` int NOT NULL,
3   `facturacioncol` varchar(45) DEFAULT NULL,
4   `ano` varchar(4) DEFAULT NULL,
5   `periodo` varchar(2) DEFAULT NULL,
6   `ciclo` int DEFAULT NULL,
7   `codigointerno` varchar(7) DEFAULT NULL,
8   `tipo_doc` int DEFAULT NULL,
9   `num_documento` varchar(45) DEFAULT NULL,
10  `uso` varchar(1) DEFAULT NULL,
11  `estrato` varchar(2) DEFAULT NULL,
12  `ubicacion` varchar(45) DEFAULT NULL,
13  `barrio` varchar(45) DEFAULT NULL,
14  `sectorhidraulico` varchar(3) DEFAULT NULL,
15  `fechalectura` date DEFAULT NULL,
16  `lectura` int DEFAULT NULL,
17  `consumo` int DEFAULT NULL,
18  PRIMARY KEY (`id_factura`)
19 ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci
```

Figura 7: DDL facturacion

```
DDL for bdproyecto.tipo_documento

1 CREATE TABLE `tipo_documento` (
2   `tipo_documento` int NOT NULL,
3   `tipo_docu_des` varchar(45) DEFAULT NULL,
4   PRIMARY KEY (`tipo_documento`)
5 ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci COMMENT=''
```

Figura 8: DDL tipo documento

DDL for bdproyecto.usos

```
1 CREATE TABLE `usos` (  
2   `id_uso` int NOT NULL,  
3   `uso_descrip` varchar(45) DEFAULT NULL,  
4   PRIMARY KEY (`id_uso`)  
5 ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci
```

Figura 9: DDL usos

DDL for basedatosu.desviaciones

```
1 CREATE TABLE `desviaciones` (  
2   `iddesviaciones` int NOT NULL,  
3   `ano` varchar(4) DEFAULT NULL,  
4   `periodo` varchar(2) DEFAULT NULL,  
5   `codigointerno` varchar(7) DEFAULT NULL,  
6   `desviacion` varchar(2) DEFAULT NULL,  
7   PRIMARY KEY (`iddesviaciones`)  
8 ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci
```

Figura 10: DDL desviaciones

4.5 Código SQL - Manipulación de datos (DML)

```
select * from basedatosu.facturacion;  
select count(*) from basedatosu.facturacion;  
select * from basedatosu.documento;  
select * from basedatosu.usos;
```

```
select  
max(periodo),  
min(periodo)  
from basedatosu.facturacion  
where codigointerno = 1423345  
and ano = 2021;
```

```
select  
sum(consumo) as promedio  
from basedatosu.facturacion  
where codigointerno = 1423345;
```

```
select  
count(consumo) as cantidadper  
from basedatosu.facturacion
```

where codigointerno = 1423345;

```
select
distinct(codigointerno),count(codigointerno)
from basedatosu.facturacion
group by codigointerno
having count(codigointerno) > 5;
```

```
call promedio(1263845);
call desviacion(1423345);
```

```
select * from basedatosu.desviaciones;
```

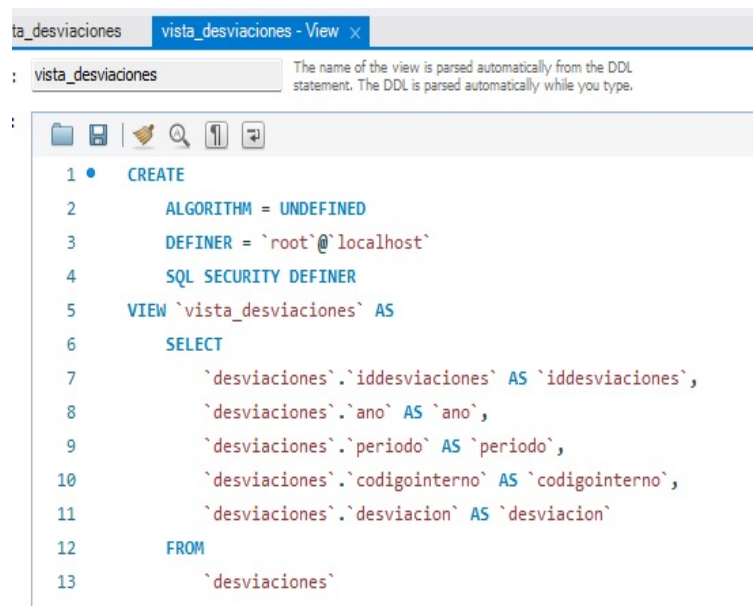
```
INSERT INTO desviaciones VALUES(iddesviaciones,
ano,periodo,codigointerno,desviacion);
```

```
DROP TRIGGER if exists 'actualizarDesviacion';
```

```
UPDATE 'basedatosu'.facturacion SET 'desviacion' = '0' WHERE
('idfacturacion' = '106781');
```

4.6 Código SQL + Resultados: Vistas

Se crea una vista para realizar la consulta de los suscriptores que presentan desviaciones en el periodo.



The screenshot shows a SQL IDE window titled 'vista_desviaciones - View'. The main editor displays the SQL code for creating the view. The code is as follows:

```
1 • CREATE
2     ALGORITHM = UNDEFINED
3     DEFINER = `root`@`localhost`
4     SQL SECURITY DEFINER
5     VIEW `vista_desviaciones` AS
6     SELECT
7         `desviaciones`.`iddesviaciones` AS `iddesviaciones`,
8         `desviaciones`.`ano` AS `ano`,
9         `desviaciones`.`periodo` AS `periodo`,
10        `desviaciones`.`codigointerno` AS `codigointerno`,
11        `desviaciones`.`desviacion` AS `desviacion`
12     FROM
13        `desviaciones`
```

Figura 11: vista

```
SELECT * FROM basedatosu.vista_desviaciones;
```

4.7 Código SQL + Resultados: Triggers

```
delimiter //  
• CREATE TRIGGER actualizarDesviacion  
  AFTER UPDATE ON basedatosu.facturacion  
  FOR EACH ROW  
  BEGIN  
    UPDATE `basedatosu`.`facturacion` SET `desviacion` = '-1'  
    WHERE ('id_facturacion' = '106781');  
  end//  
delimiter ;
```

Figura 12: trigger

El trigger detecta que se actualizo un registro en la tabla desviaciones, y procede a actualizar el registro en la tabla facturación.

4.8 Código SQL + Resultados: Funciones

```
1 • CREATE DEFINER='root'@'localhost' FUNCTION `desviacions_repe`(codigo int) RETURNS int  
2   READS SQL DATA  
3   DETERMINISTIC  
4   BEGIN  
5     declare codigoi varchar(7);  
6     declare numero INT;  
7     SELECT  
8       distinct codigointerno,  
9       COUNT(*)  
10      into codigoi, numero  
11      from basedatosu.desviaciones WHERE codigointerno = codigo  
12      group by codigointerno  
13      having count(*) > 1;  
14      RETURN numero;  
--
```

Figura 13: función

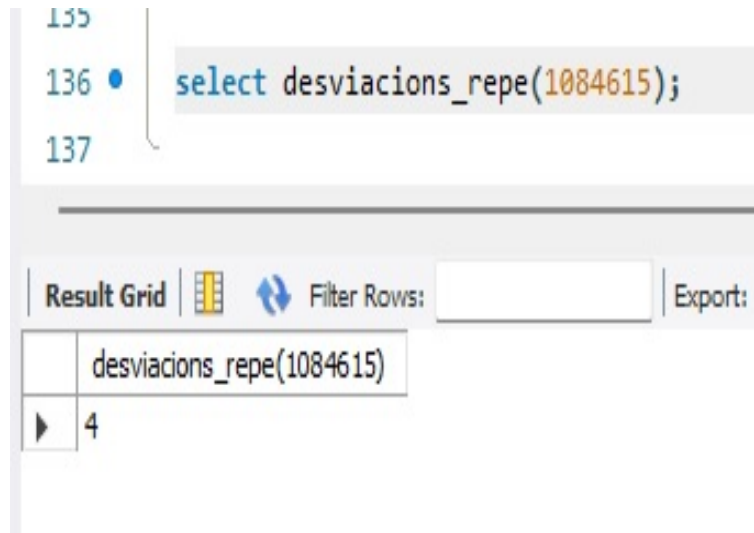
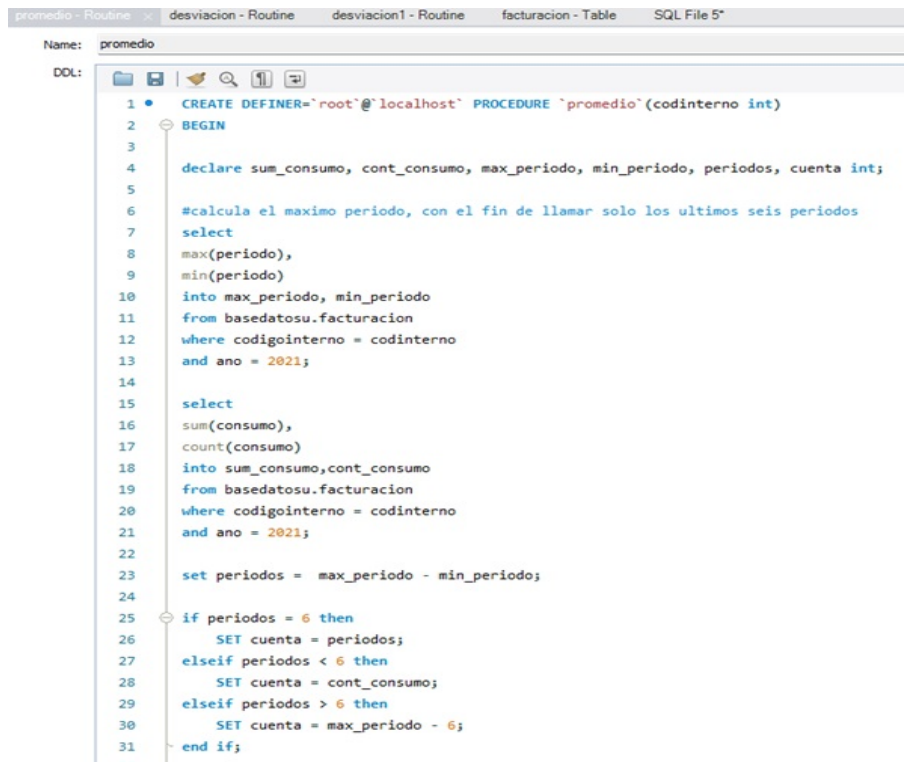


Figura 14: ejecución de la función

Se crea una función que permite revisar si existen repetidos en la tabla desviaciones, este servirá para alertar si el suscriptor ha presentado mas de una desviación en el mismo año.

4.9 Código SQL + Resultados: procedimientos almacenados

Procedimiento de almacenado para calcular el promedio de facturación de los ultimos seis meses:



The screenshot shows a SQL IDE with several tabs at the top: 'promedio - Routine', 'desviacion - Routine', 'desviacion1 - Routine', 'facturacion - Table', and 'SQL File 5*'. The 'promedio - Routine' tab is active. Below the tabs, the 'Name:' field contains 'promedio'. The 'DDL:' section shows the SQL code for creating the procedure. The code is as follows:

```
1 • CREATE DEFINER='root'@'localhost' PROCEDURE `promedio`(codinterno int)
2 BEGIN
3
4 declare sum_consumo, cont_consumo, max_periodo, min_periodo, periodos, cuenta int;
5
6 #calcula el maximo periodo, con el fin de llamar solo los ultimos seis periodos
7 select
8 max(periodo),
9 min(periodo)
10 into max_periodo, min_periodo
11 from basedatosu.facturacion
12 where codigointerno = codinterno
13 and ano = 2021;
14
15 select
16 sum(consumo),
17 count(consumo)
18 into sum_consumo, cont_consumo
19 from basedatosu.facturacion
20 where codigointerno = codinterno
21 and ano = 2021;
22
23 set periodos = max_periodo - min_periodo;
24
25 if periodos = 6 then
26     SET cuenta = periodos;
27 elseif periodos < 6 then
28     SET cuenta = cont_consumo;
29 elseif periodos > 6 then
30     SET cuenta = max_periodo - 6;
31 end if;
```

Figura 15: procedimiento de almacenado, calculo del promedio

El procedimiento de almacenado permite calcular la desviación significativa del periodo de facturación, identificado procede a realizar el insert en la tabla, la cual sirve como un histórico de estos sucesos.:

desviacion - Routine

Name: desviacion

The name of the routine is parsed automatically in statement. The DDL is parsed automatically while

DDL:

```
1 CREATE DEFINER='root'@'localhost' PROCEDURE `desviacion`(codinterno int)
2 BEGIN
3
4 declare prome, Mayor40, Menor40, sum_consumo, cont_consumo, max_periodo, min_periodo, periodos, cuenta, desviacion, cont_tabla int;
5
6 #calcula el maximo periodo, con el fin de llamar solo los ultimos seis periodos
7 select
8     max(periodo),
9     min(periodo)
10    into max_periodo, min_periodo
11   from basedatosu.facturacion
12  where codigointerno = codinterno
13     and ano = 2021;
14
15 select
16     sum(consumo),
17     count(consumo)
18    into sum_consumo, cont_consumo
19   from basedatosu.facturacion
20  where codigointerno = codinterno
21     and ano = 2021;
22
23 set periodos = max_periodo - min_periodo;
24
25 if periodos = 6 then
26     SET cuenta = periodos;
27 elseif periodos < 6 then
28     SET cuenta = cont_consumo;
29 elseif periodos > 6 then
```



```
desviacion - Routine x
Name: desviacion
DDL:
30     SET cuenta = max_periodo - 6;
31   end if;
32
33   set prome = round(sum_consumo / cuenta);
34
35   if prome > 40 then
36     SET Mayor40 = (40 * 35) / 100;
37     if prome >= Mayor40 then
38       set desviacion = -1;
39       select -1; #desviado
40     else
41       set desviacion = 0;
42       select 0; # no desviado
43     end if;
44   else
45     SET Menor40 = (39 * 65) / 100;
46     if prome >= Menor40 then
47       set desviacion = -1;
48       select -1; #desviado
49     else
50       set desviacion = 0;
51       select 0; # no desviado
52     end if;
53   end if;
54
55   select
56   count(*) +1
57   into cont_tabla
58   from basedatosu.desviaciones;
59
60   INSERT INTO desviaciones VALUES(cont_tabla,2021,max_periodo,codinterno,desviacion);
61   END
```

Figura 16: procedimiento de almacenado, para insertar en la tabla

5 Bases de Datos No-SQL

Para este proyecto se utiliza MongoDB, como solución del almacenamiento de datos o estructurados, debido a su flexibilidad ya que no sigue ningún esquema, tiene alta disponibilidad y tiene una amplia documentación de apoyo. Adicionalmente tiene otras ventajas, es gratuita, es de código abierto y la curva de aprendizaje es muy pequeña.

5.1 Diagrama Bases de Datos No-SQL

Cuando se realiza la visita en el predio del suscriptor, se hace necesario tomar evidencia fotográfica del estado del inmueble, orientado a encontrar indicios de la desviación significativa; el registro fotográfico se utiliza como prueba en las Inspecciones que se realizan. Con este fin se utiliza una base de datos no relacional, en la cual pueden reposar estos documentos.

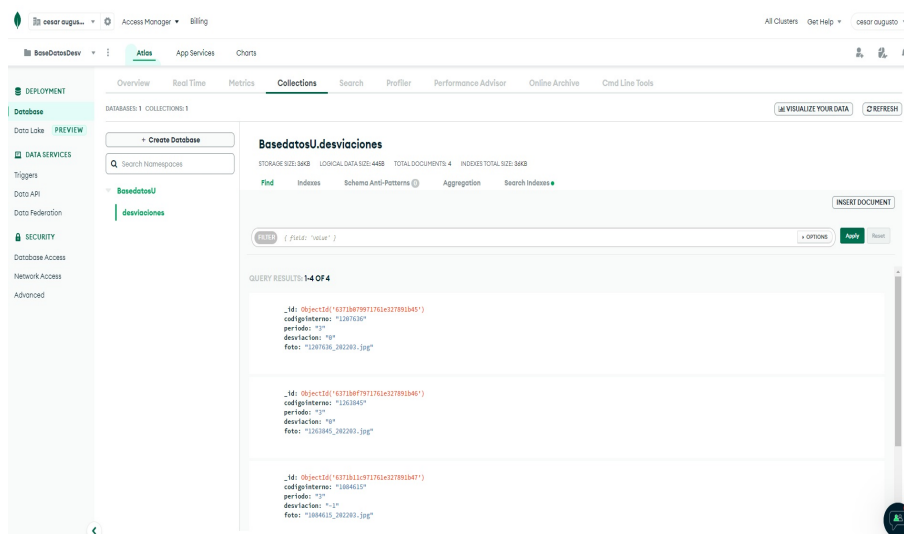


Figura 17: BD no-sql, repositorio de fotografias

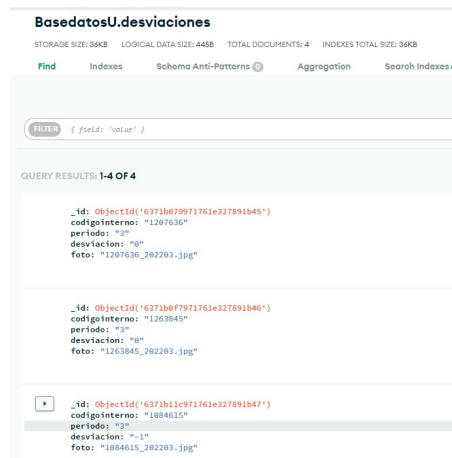


Figura 18: BD no-sql, repositorio de fotografias

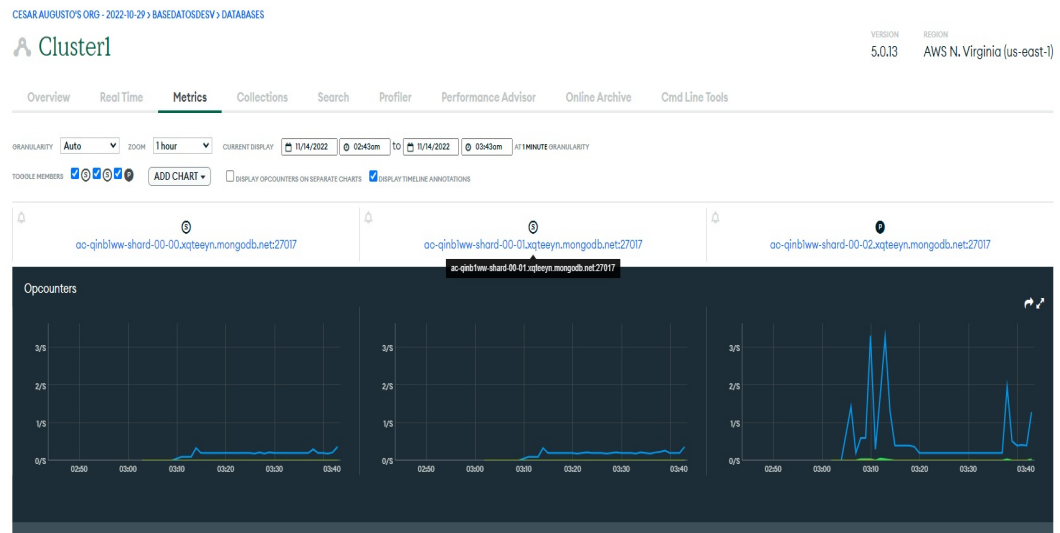


Figura 19: BD no-sql, metricas

5.2 SMBD utilizado para la Base de Datos No-SQL

Teniendo los datos en la nube, se hace necesario administrarlos de manera local, esto con el fin de poder realizar consultas con las aplicaciones locales que interactuaran tanto con la base de datos relacional como la no relacional. Para este fin se utilizará la interfaz MongoDB Compass.

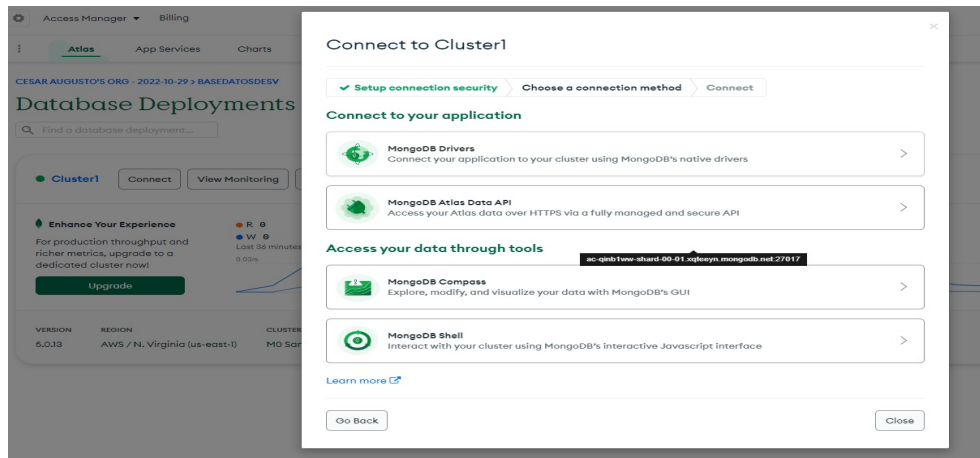


Figura 20: BD no-sql, descarga de MongoDB Compass

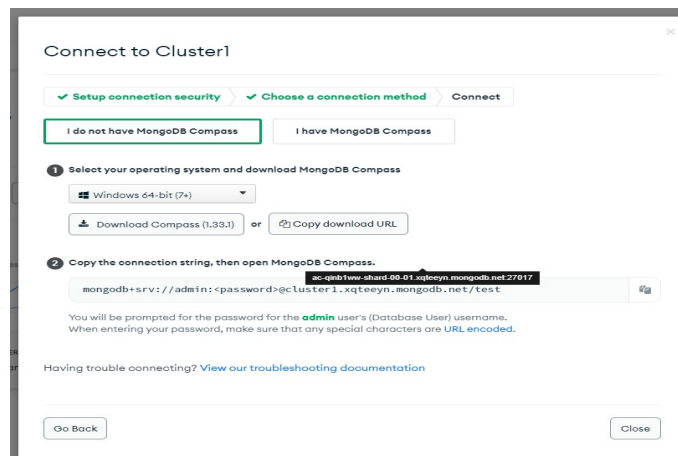


Figura 21: BD no-sql, descarga de MongoDB Compass

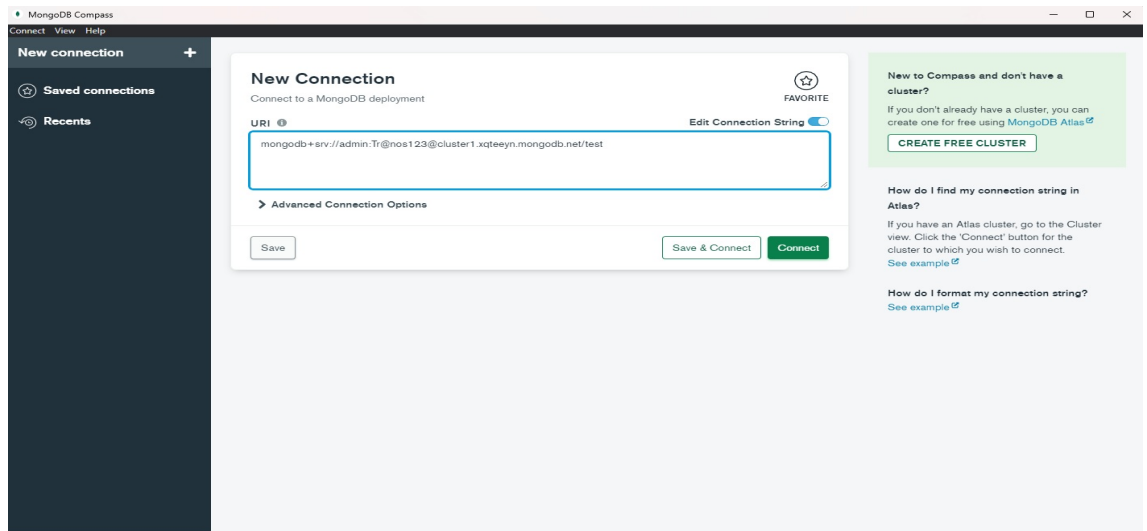


Figura 22: BD no-sql, configuración

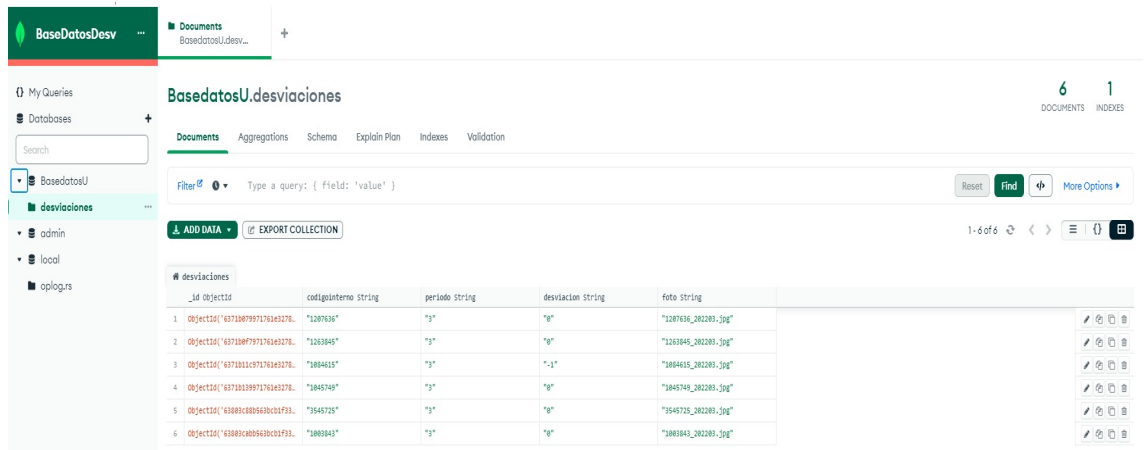


Figura 23: BD no-sql, vista local

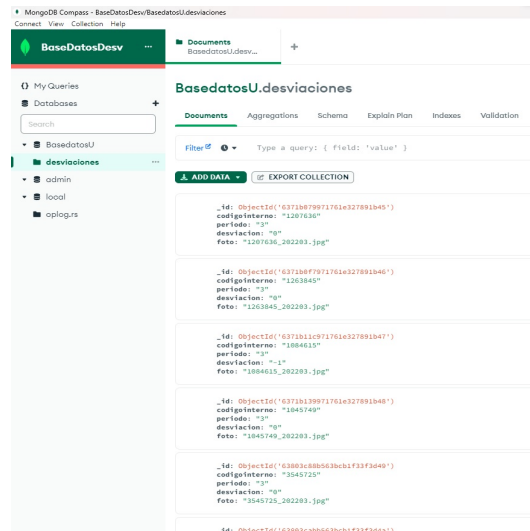


Figura 24: BD no-sql, vista local

Con la herramienta instalada, se realiza una consulta de los datos ingresados y se corrobora la sincronización con la información que se encuentra en la nube.

6 Bibliografía

- Acueducto de Bogotá EAAB(Agosto 2018), <https://www.acueducto.com.co>, Anexo: ABC DESVIACIONES.pdf
- Superservicios (2010), <https://www.youtube.com/watch?v=4OQTCvJHhww>
- Empresa de Acueducto y Alcantarillado de Yopal (Septiembre 2008): Procedimiento De Identificación, Trámite, Análisis Y Cierre De Desviaciones Significativas Positivas