

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

October 15, 2022

Contents

1	Introducción (Max 250 Palabras) - (<i>Primera entrega</i>)	3
2	Características del proyecto de investigación (Max 500 Palabras) - (<i>Primera entrega</i>)	3
2.1	Titulo del proyecto de investigación (Max 100 Palabras) - (<i>Primera entrega</i>)	3
2.2	Objetivo general (Max 100 Palabras) - (<i>Primera entrega</i>)	3
2.2.1	Objetivos especificos (Max 100 Palabras) - (<i>Primera entrega</i>)	3
2.3	Alcance (Max 200 Palabras) - (<i>Primera entrega</i>)	4
2.4	Pregunta de investigación (Max 100 Palabras) - (<i>Primera entrega</i>) .	4
2.5	Hipotesis (Max 100 Palabras) - (<i>Primera entrega</i>)	4
3	Reflexiones sobre el origen de datos e información (Max 400 Palabras) - (<i>Primera entrega</i>)	5
3.1	¿Cual es el origen de los datos e información ? (Max 100 Palabras) - (<i>Primera entrega</i>)	5
3.2	¿Cuales son las consideraciones legales o eticas del uso de la información? (Max 100 Palabras) - (<i>Primera entrega</i>)	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? (Max 100 Palabras) - (<i>Primera entrega</i>)	5
3.4	¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto? (Max 100 Palabras) - (<i>Primera entrega</i>)	6

4	Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)(<i>Primera entrega</i>)	7
4.1	Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto (<i>Primera entrega</i>)	7
4.2	Diagrama modelo de datos (<i>Primera entrega</i>)	8
4.3	Imágenes de la Base de Datos (<i>Primera entrega</i>)	9
4.4	Código SQL - lenguaje de definición de datos (DDL) (<i>Primera entrega</i>)	10
4.5	Código SQL - Manipulación de datos (DML) (<i>Primera entrega</i>) . .	11
4.6	Código SQL + Resultados: Vistas (<i>Primera entrega</i>)	12
4.7	Código SQL + Resultados: Triggers (<i>Primera entrega</i>)	12
4.8	Código SQL + Resultados: Funciones (<i>Primera entrega</i>)	12
4.9	Código SQL + Resultados: procedimientos almacenados (<i>Primera entrega</i>)	13
5	Bases de Datos No-SQL (<i>Segunda entrega</i>)	16
5.1	Diagrama Bases de Datos No-SQL (<i>Segunda entrega</i>)	16
5.2	SMBD utilizado para la Base de Datos No-SQL (<i>Segunda entrega</i>)	16
6	Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos (<i>Tercera entrega</i>)	17
6.1	Ejemplo de aplicación de ETL y Bodega de Datos (<i>Tercera entrega</i>)	17
7	Lecciones aprendidas (<i>Tercera entrega</i>)	18
8	Bibliografía	19

1 Introducción (Max 250 Palabras) - (*Primera entrega*)

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.

2 Características del proyecto de investigación (Max 500 Palabras) - (*Primera entrega*)

2.1 Título del proyecto de investigación (Max 100 Palabras) - (*Primera entrega*)

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

2.2 Objetivo general (Max 100 Palabras) - (*Primera entrega*)

Diseñar y construir un modelo de clasificación, que permita determinar si el consumo de agua de un cliente de Fusagasugá Cundinamarca se encuentra dentro de las condiciones normales de medición.

2.2.1 Objetivos específicos (Max 100 Palabras) - (*Primera entrega*)

- Realizar un análisis exploratorio de los datos, mediante el uso de metodologías de estadística descriptiva, con el fin de conocer patrones y tendencias de consumo de agua por parte de los usuarios.
- Identificar y determinar mediante un análisis de correlación de variables, el grado de relación de las variantes implicadas en el consumo de agua de los clientes.
- Determinar la estructura de las vistas donde se alojará la información que se extraerá del reporte. Non tellus orci ac auctor augue. Elit at imperdiet dui accumsan sit. Ornare arcu dui vivamus arcu felis.
- Crear procedimientos de almacenado, orientados a la captura y alistamiento de la información de valores atípicos para reportes diarios.
- Implementar trigger para la ejecución de los procedimientos de almacenado y limpieza de tablas temporales.
- Definir el reporte resultado para el análisis de los datos atípicos presentados diariamente en el proceso de facturación.

2.3 Alcance (Max 200 Palabras) - (*Primera entrega*)

A través del análisis de 6 meses 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 desviaciones significativas.

2.4 Pregunta de investigación (Max 100 Palabras) - (*Primera entrega*)

¿En qué medida los consumos atípicos influyen en el porcentaje de agua no contabilizada en el Municipio de Fusagasugá?

2.5 Hipotesis (Max 100 Palabras) - (*Primera entrega*)

La mayor cantidad de perdidas en agua se encuentra en el consumo que se realiza en el predio, y se asocia por el consumo que supera el promedio del mismo uso, estrato, ubicación y barrio, cuyo origen es variado como daños en la red interna, manipulación de la acometida, etc.

3 Reflexiones sobre el origen de datos e información

(Max 400 Palabras) - (*Primera entrega*)

La información utilizada en el presente ejercicio, hace parte de la base de datos histórica de la entidad EMSERFUSA ESP del año 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/>. 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 ? (Max 100 Palabras) - (*Primera entrega*)

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? (Max 100 Palabras) - (*Primera entrega*)

La División Comercial, como fuente de información, autoriza su uso, solo para el objeto de estudio expuesto, mientras se conserve los principios de Habas 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? (Max 100 Palabras) - (*Primera entrega*)

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? (Max 100 Palabras) - (*Primera entrega*)

Vulputate nec hac convallis rutrum eu ante volutpat aliquam ullamcorper pulvinar tristique velit nulla, cubilia felis tempor aptent vitae rhoncus parturient euismod mauris potenti dignissim magna. Nunc nec cum quisque quam tincidunt mauris nascetur conubia placerat fusce consequat eget erat vulputate, est rhoncus etiam dignissim luctus volutpat facilisi molestie torquent at convallis accumsan. Magnis morbi auctor sapien fusce turpis leo ad libero vivamus, sem enim ultrices elementum curae accumsan vel hendrerit. Etiam elementum dui a sodales auctor lacus proin interdum, porttitor netus tortor blandit sociis facilisi ullamcorper, mi aenean euismod diam placerat dignissim class.

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

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

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.

TABLA FACTURACIÓN

VARIABLES	DESCRIPCIÓN	TIPO
ANO	Año de la toma de lectura	Categórica
PERIODO	Periodo (mes) de toma de lectura	Categórica
CICLO	Ciclo, agrupación por zona geografica de inmuebles	Categórica
CODIGOINTERNO	Consecutivo que identifica el inmueble	Numérica
TIPODOCUMENTO	Tipo de documento con el que se identifica el suscriptor	Numérica
NUMDOCUMENTO	Número de documento de identificación del suscriptor	Carácter
USO	Servicio Residencial, Industrial, Comercial, Oficial	Numérica
ESTRATO	Clasificación del uso residencial	Numérica
UBICACIÓN	Ubicación rural o urbana del predio	Categórica
BARRIO	Barrio de ubicación del predio	Categórica
SECTHIDRAULICO	Código de ubicación de la fuente de abastecimiento	Numérico
FECHAEXPFACTURA	Fecha de expedición de la factura	Fecha

La Tabla tipo de documento, contiene el descriptivo de los tipos de documentos

TABLA TIPO DOCUMENTO

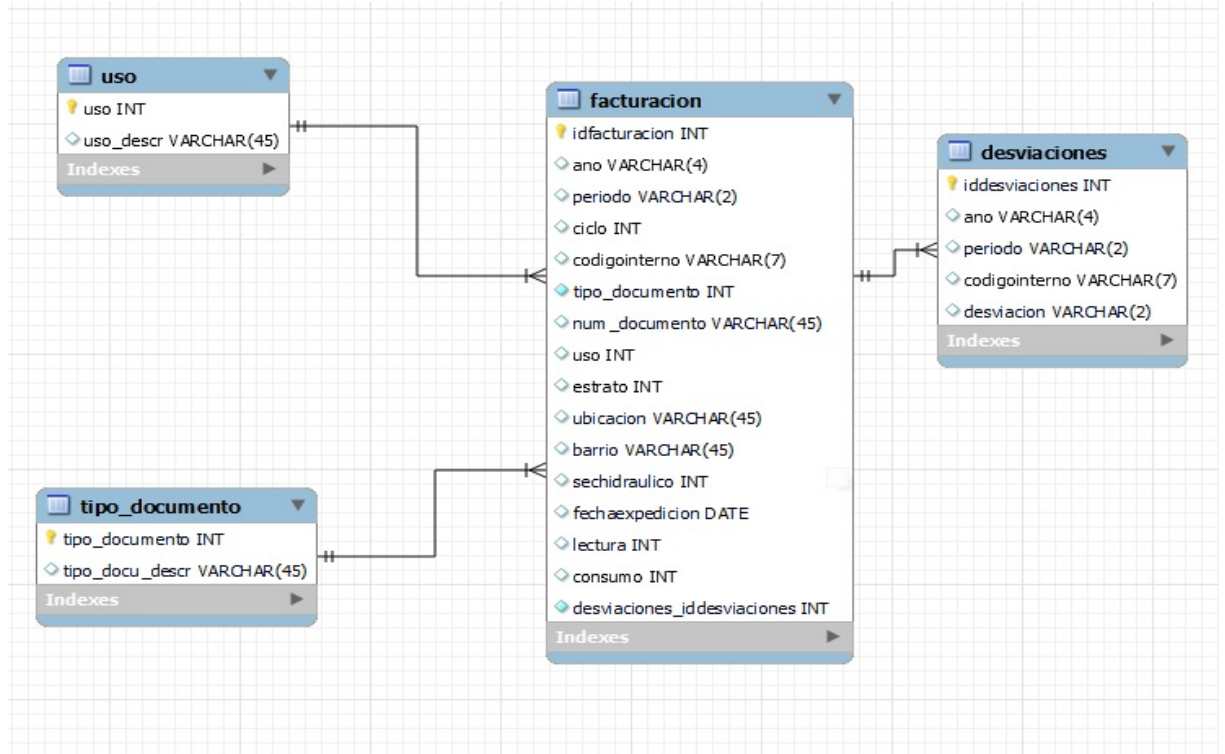
VARIABLES	DESCRIPCIÓN	TIPO
TIPODOCUMENTO	Tipo de documento con el que se identifica el suscriptor	Numérico
TIPODOCDESCRIP	Descripción de tipo de documento	Categórica

La Tabla Uso, tiene el descriptivo del identificador de los códigos de estratificación

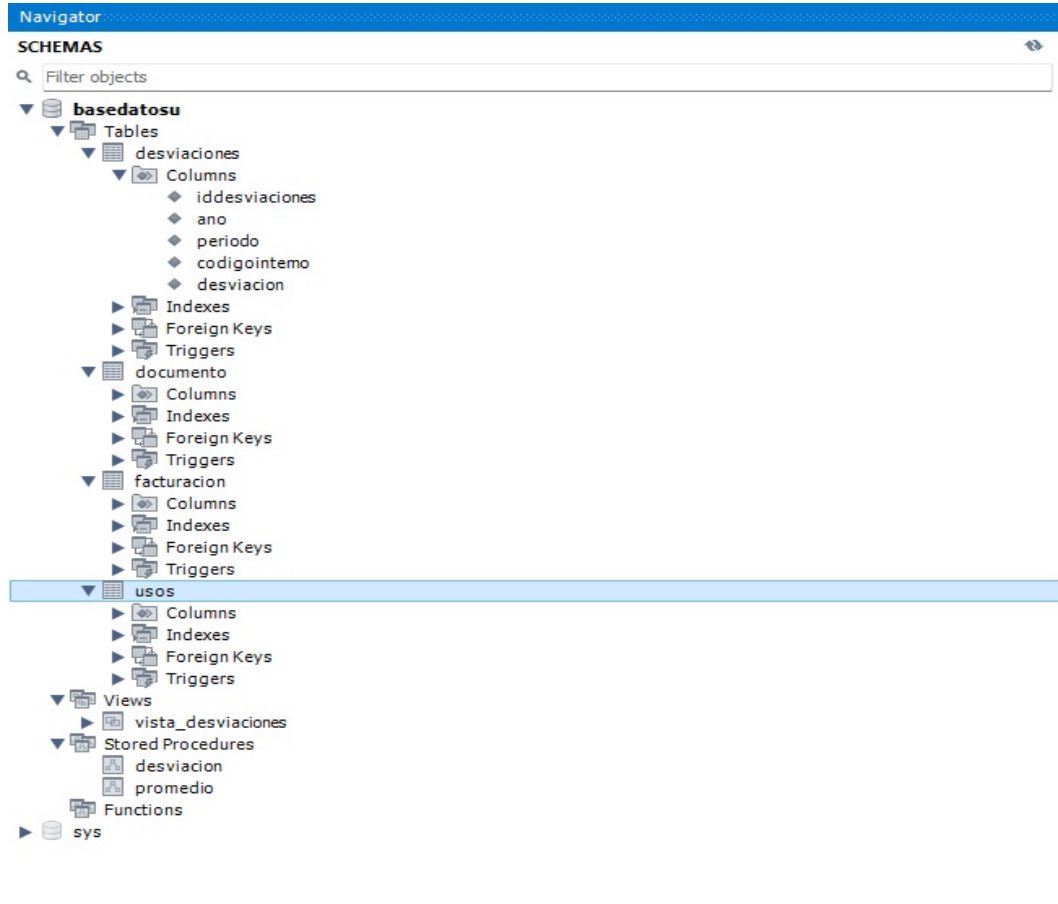
TABLA USO

VARIABLES	DESCRIPCIÓN	TIPO
IDUSO	Servicio Residencial, Industrial, Comercial, Oficial	Numérico
UDODESCRIPCION	Descripción del código de uso	Categórica

4.2 Diagrama modelo de datos (*Primera entrega*)



4.3 Imágenes de la Base de Datos (*Primera entrega*)



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

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
```

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=''
```

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
```

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
```

4.5 Código SQL - Manipulación de datos (DML) (Primera entrega)

```
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;
```

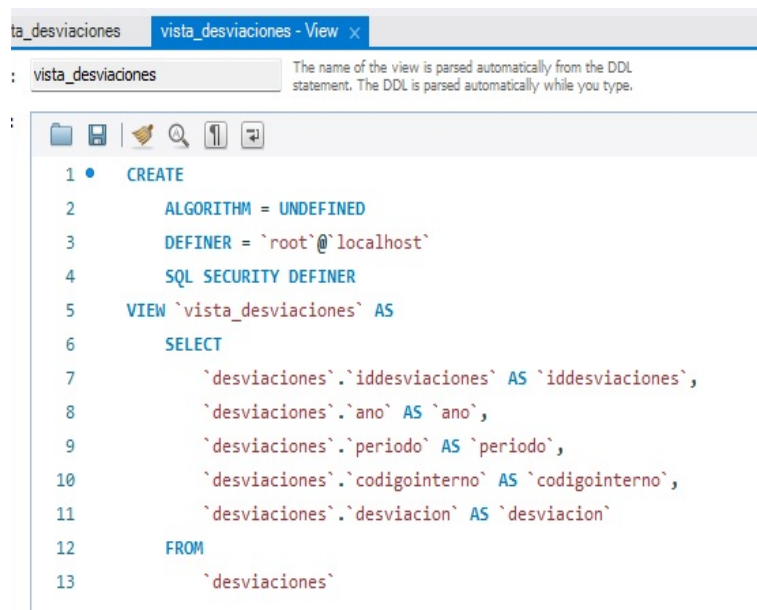
```
call promedio(1263845);  
call desviacion(1423345);
```

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

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

4.6 Código SQL + Resultados: Vistas (*Primera entrega*)

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`
```

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

4.7 Código SQL + Resultados: Triggers (*Primera entrega*)

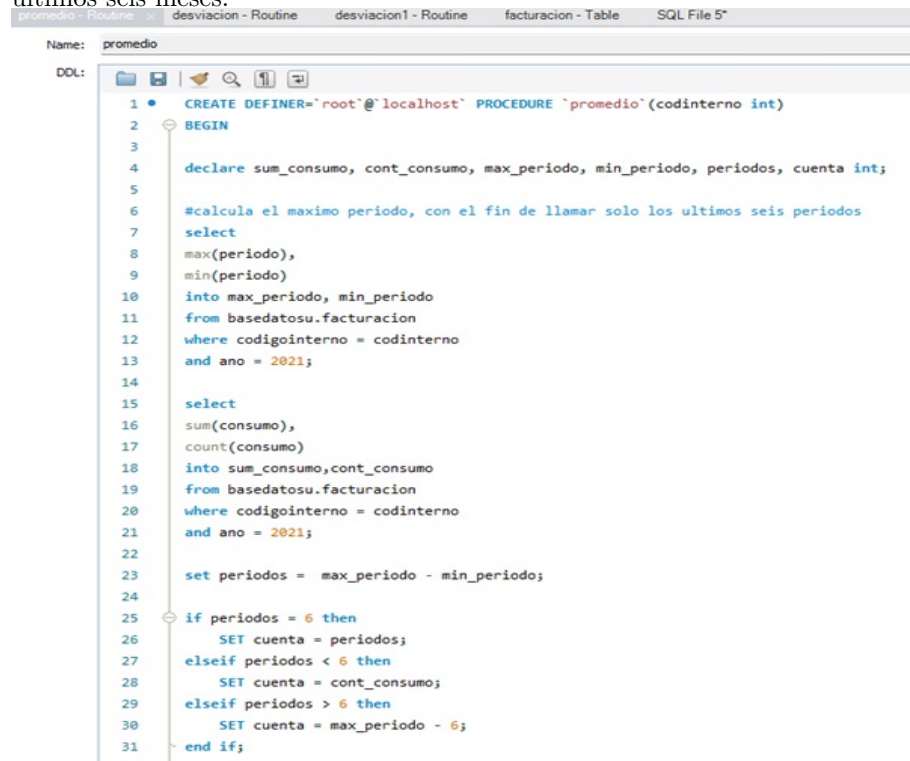
Aun el proceso sigue en desarrollo, por tal motivo no se ha visto la necesidad de la implementación de este recurso

4.8 Código SQL + Resultados: Funciones (*Primera entrega*)

Aun el proceso sigue en desarrollo, por tal motivo no se ha visto la necesidad de la implementación de este recurso

4.9 Código SQL + Resultados: procedimientos almacenados (Primera entrega)

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



```
promedio - Routine | desviacion - Routine | desviacion1 - Routine | facturacion - Table | SQL File 5*
Name: promedio
DDL:
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;
```

Procedimiento de almacenado para calcular la desviación significativa, del periodo de facturación:

desviacion - Routine x

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_perodo - 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_perodo,codinterno,desviacion);
61   END
```

5 Bases de Datos No-SQL (*Segunda entrega*)

5.1 Diagrama Bases de Datos No-SQL (*Segunda entrega*)

5.2 SMBD utilizado para la Base de Datos No-SQL (*Segunda entrega*)

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

6.1 Ejemplo de aplicación de ETL y Bodega de Datos (*Tercera entrega*)

7 Lecciones aprendidas (*Tercera entrega*)

8 Bibliografía