ENTREGA 2 – SISTRANS - INFORME

Wilmer Manuel Arévalo González - 202214720 - w.arevalo Sara Sofía Cárdenas Rodríguez - 202214907 - ss.cardenas Daniel Felipe Díaz Moreno – 202210733 - d.diazm

1. ANÁLISIS

1.1. MODELO UML

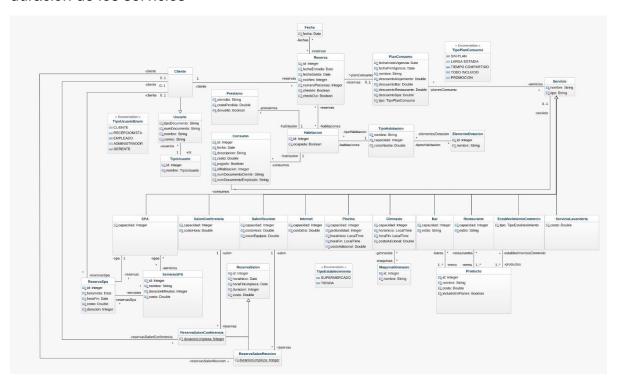
Se crearon asociaciones entre el usuario y más clases, ya que no todas las reservas apuntaban al cliente que las tenía (reservasSpa, reservasSalonConferencia, reservasSalonReunion)

Se creó una nueva clase llamada Fechas, con el fin de asociarse con las reservas y obtener el índice de ocupación de las habitaciones del hotel

A Consumos, se le añadió un atributo num_documento_empleado para apuntar al empleado que registró el consumo

A Servicios, se le añadió un atributo tipo para modelar los tipos de servicios mencionados en el enunciado

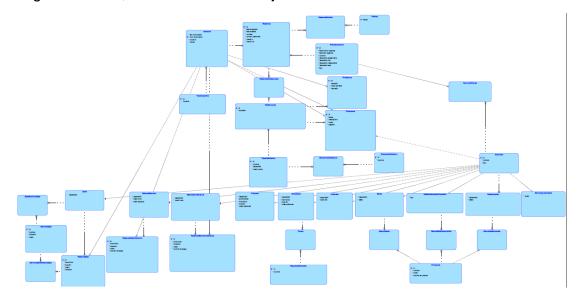
A ReservasSpa, se le añadió un atributo duración para facilitar la obtención de la duración de los servicios



Para obtener una resolución óptima, consulte el archivo UML.pdf ubicado en la carpeta "modelos" dentro del directorio "docs".

1.2. MODELO E/R

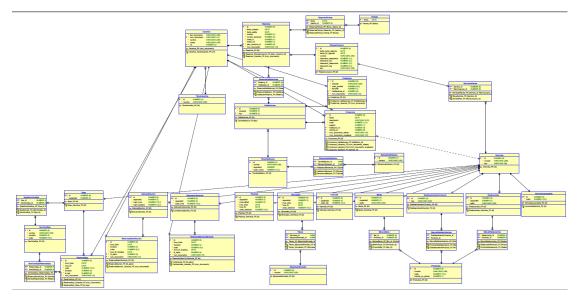
De igual manera, los cambios se reflejan en este modelo



Para obtener una resolución óptima, consulte el archivo modeloER.pdf ubicado en la carpeta "modelos" dentro del directorio "docs".

1.3. MODELO RELACIONAL

De igual manera, los cambios se reflejan en este modelo



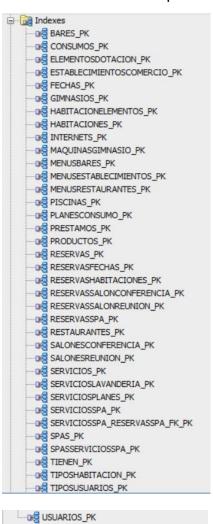
Para obtener una resolución óptima, consulte el archivo modeloRelacional.pdf ubicado en la carpeta "modelos" dentro del directorio "docs"

2. DISEÑO DE LA APP

2.1. ÍNDICES

2.1.1. ÍNDICES DE ORACLE

Oracle creó 37 índices para modelar las 37 llaves primarias de todas las tablas.



Cabe resaltar que, para las tablas de asociación, como hay dos columnas en sus llaves primarias, se generaron índices compuestos sobre ambas columnas. Por ejemplo, en el menú de los bares



Entonces, estos índices fueron creados ya que correspondes a las PK de todas las tablas, y evidentemente aumentan la velocidad de todos los requerimientos, dado que las PK son columnas susceptibles al uso de joins y a consultas de todo tipo. Además, cuentan con una altísima selectividad

2.1.2. ÍNDICES CREADOS

Se crearon 25 índices, especialmente sobre atributos usados para join, como FKs, y sobre columnas frecuentemente consultadas para rangos, como los costos o fechas

```
CREATE INDEX reservas_fecha_entrada ON reservas(fecha_entrada);
CREATE INDEX reservas_fecha_salida ON reservas(fecha_salida);
CREATE INDEX reservas noches ON reservas(noches);
CREATE INDEX reservas num documento ON reservas(num documento);
CREATE INDEX consumos_fecha ON consumos(fecha);
CREATE INDEX consumos_costo ON consumos(costo);
CREATE INDEX consumos_servicio_id ON consumos(servicio_id);
CREATE INDEX consumos_habitacion_id ON consumos(habitacion_id);
CREATE INDEX consumos_num_documento_cliente ON consumos(num_documento_cliente);
CREATE INDEX consumos_num_documento_empleado ON consumos(num_documento_empleado);
CREATE INDEX servicios_tipo ON servicios(tipo);
CREATE INDEX salonesReunion_costo_hora ON salonesReunion(costo_hora);
CREATE INDEX salonesConferencia_costo_hora ON salonesConferencia(costo hora);
CREATE INDEX piscinas_costo_adicional ON piscinas(costo_adicional);
CREATE INDEX gimnasios costo adicional ON gimnasios(costo adicional);
CREATE INDEX internets_costo_dia ON internets(costo_dia);
CREATE INDEX productos_costo ON productos(costo);
CREATE INDEX serviciosLavanderia_costo ON serviciosLavanderia(costo);
  EATE INDEX serviciosSpa_costo ON serviciosSpa(costo);
CREATE INDEX reservasSalonReunion_duracion ON reservasSalonReunion(duracion);
CREATE INDEX reservasSalonReunion_num_documento ON reservasSalonReunion(num_documento);
CREATE INDEX reservasSalonConferencia_duracion ON reservasSalonConferencia(duracion);
CREATE INDEX reservasSalonConferencia_num_documento ON reservasSalonConferencia(num_documento);
 REATE INDEX reservasSpa_duracion ON reservasSpa(duracion);
CREATE INDEX reservasSpa_num_documento ON reservasSpa(num_documento);
```

RFC1: DINERO RECOLECTADO POR SERVICIOS EN CADA HABITACIÓN EN EL ÚLTIMO AÑO.

Se crearon tres índices en Consumos, uno para la fecha, otro para el costo y otro para el id de la habitación

Se desearía crear un índice de hash en id_habitación, dados los valores únicos, pero no fue posible. Se optó por B+ secundario

Para los demás, e creó un índice B+ de tipo secundario, para acceder rápidamente a rangos y a valores de estos atributos

Bajo selectividad para la fecha, hay 730 días entre 2022 y 2023, por lo que la alta selectividad es del 0.13 %

Bajo selectividad para el costo, hay 1 E10 posibles valores en pesos colombianos, pero normalmente oscilan entre 0 y 1 millon de pesos, por lo que la alta selectividad es del 0.0001 %

Bajo selectividad para id_habitación, como hay más de 1k habitaciones, la selectividad es menor que 0.001%

RFC2: 20 SERVICIOS MÁS POPULARES

Nuevamente se usa el índice de fecha en consumo, para evaluarlo entre dos fechas

RFC3: ÍNDICE DE OCUPACIÓN DE CADA UNA DE LAS HABITACIONES DEL HOTEL

Se crearon tres índices en Reservas, uno para la fecha de entrada, otro para la fecha de salida y otro para la cantidad de noches

Se creó un índice B+ de tipo secundario, para acceder rápidamente a rangos y a valores de estos atributos

La selectividad para las fechas es la misma

Para la cantidad de noches, se tiene mínimo una noche y máximo 21, por lo que la selectividad es del 4.76%

RFC4: SERVICIOS QUE CUMPLEN CON CIERTA CARACTERÍSTICA

Precio

Se crearon ocho índices en las tablas que heredan de servicios (Piscinas, Gimnasios, Internets, Salones, ServiciosLavanderia), para atributos que modelan precios (costo, costo_hora, costo_dia, costo_adicional)

Para los servicios sin costo asociado, se crearon dos índices sobre productos y serviciosSpa sobre la columna costo

Se creó un índice B+ de tipo secundario, para acceder rápidamente a rangos y a valores de estos atributos

Nuevamente la selectividad para costos es la misma

Empleado

Se creó un índice sobre la columna num_documento_empleado en Consumos

Se desearía crear un índice de hash, dados los valores únicos, pero no fue posible. Se optó por B+ secundario

Bajo selectividad, como hay más de 1k empleados, la selectividad es menor que 0.001%

Tipo

Se creó un índice sobre la columna tipo en Servicios

Se desearía crear un índice de hash, dados los valores únicos, pero no fue posible. Se optó por B+ secundario

Bajo selectividad, como hay 10 tipos, la selectividad es del 10%

RFC5: CONSUMO POR UN USUARIO DADO, EN UN RANGO DE FECHAS

Nuevamente se usa el índice de fecha en consumo, para evaluarlo entre dos fechas

RFC6: OPERACIÓN DE HOTELANDES

Se usaron los índices por defecto de Oracle, ya que eran solo PKs

RFC7: BUENOS CLIENTES

Se creó un índice en Reservas, para el num_documento

Se desearía crear un índice de hash, dados los valores únicos, pero no fue posible. Se optó por B+ secundario

Se crearon dos índices en Consumos, para servicio_id y para num_documento_cliente

Se desearía crear dos índices de hash, dados los valores únicos, pero no fue posible. Se optó por B+ secundario

Bajo selectividad de num_documento y num_documento_cliente, como hay más de 1k clientes, la selectividad es menor que 0.001%

Bajo selectividad de servicio_id, como hay más de 1k servicios, la selectividad es menor que 0.001%

RFC8: SERVICIOS QUE NO TIENEN MUCHA DEMANDA

Nuevamente se usa el índice de servicios_id en Consumos

RFC9: CONSUMO EN HOTELANDES

Nuevamente se usan los índices de consumos

RFC10: CONSUMO EN HOTELANDES v2

Nuevamente se usan los índices de consumos

RFC11: FUNCIONAMIENTO

Nuevamente se usan los índices en reservas para las fechas y los de consumos

RFC12: CLIENTES EXCELENTES

Se crearon cuatro índices en ReservasSalonReunión y ReservasSalonConferencia, para sus atributos num_documento y duración

En ReservasSpa, se crearon dos índices para num_documento y duración de igual forma

Se desearía tres índices de hash para num_documento, dados los valores únicos, pero no fue posible. Se optó por B+ secundario

Se crearon tres índices B+ de tipo secundario, para acceder rápidamente a rangos y a valores de duración

Bajo selectividad de num_documento y num_documento_cliente, como hay más de 1k clientes, la selectividad es menor que 0.001%

Las duraciones en minutos rondan entre 15 y 720, por lo que la selectividad es del 0.13%

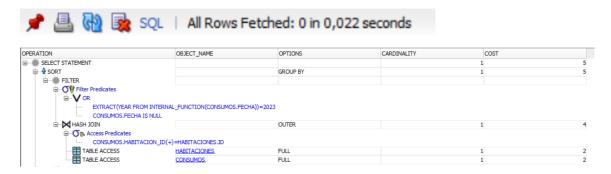
2.2. DISEÑO DE LAS CONSULTAS

Las sentencias SQL de consulta se encuentran en scripts/Sentencias

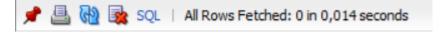
La distribución de los datos se hizo con grandes intervalos. Para fechas, se escogió todo el 2023. Para tipos, se escogió Bares y ServiciosLavandería. Se usaron usuarios predefinidos en la población de la anterior iteración. Se usarons costos no superiores a 100.000 pesos.

A continuación, se presentan los planes de consulta y tiempos de ejecución para la población reducida. Aun así, los planes deberían mantenerse en su estructura.

RFC1: DINERO RECOLECTADO POR SERVICIOS EN CADA HABITACIÓN EN EL ÚLTIMO AÑO.



RFC2: 20 SERVICIOS MÁS POPULARES

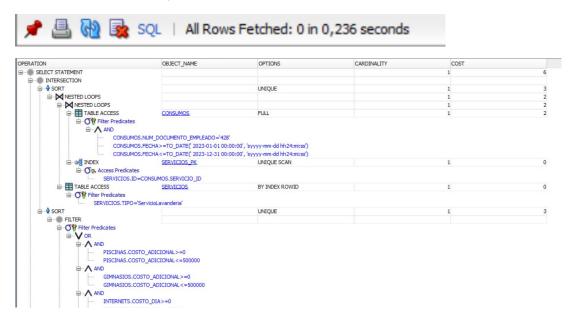


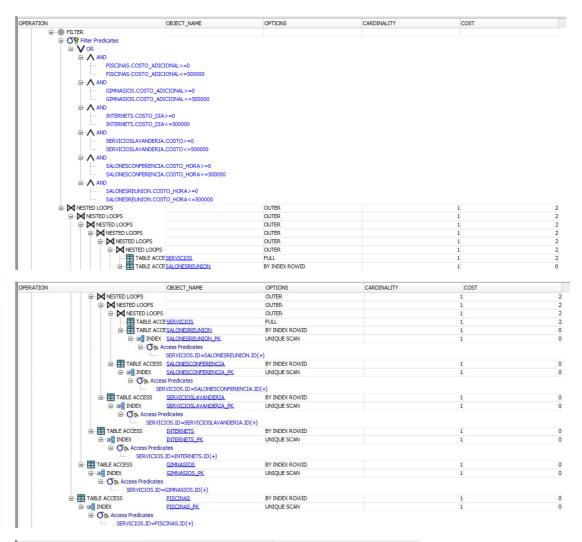


RFC3: ÍNDICE DE OCUPACIÓN DE CADA UNA DE LAS HABITACIONES DEL HOTEL

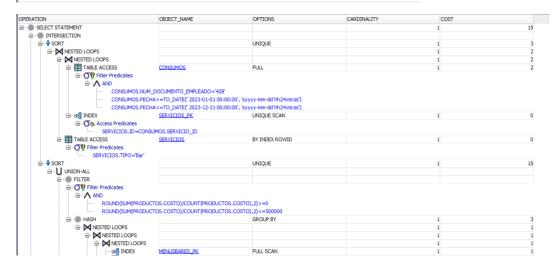


RFC4: SERVICIOS QUE CUMPLEN CON CIERTA CARACTERÍSTICA









⊟ ∧ AN					
		OS.COSTO)/COUNT(PRODUCTOS.CO			
_	ROUND(SUM(PRODUCTO	DS.COSTO)/COUNT(PRODUCTOS.CO			
HASH			GROUP BY	1	3
	ESTED LOOPS			1	1
	NESTED LOOPS			1	1
P	NESTED LOOPS			1	1
		MENUSBARES_PK	FULL SCAN	1	1
	TABLE ACCESS		BY INDEX ROWID	1	0
		SERVICIOS_PK	UNIQUE SCAN	1	0
	⊕ Ot Acces				
		VICIOS.ID=MENUSBARES.BAR_ID	I INTO UE CCAN	•	0
-		PRODUCTOS_PK	UNIQUE SCAN	1	0
	Access Predica				
=		S.PRODUCTO_ID=PRODUCTOS.ID	BY INDEX ROWID	1	0
FILTER	TABLE ACCESS	PRODUCTOS.	BY INDEX ROWID	1	U
Filter	Deadleston				
□ AN					
- / \		OS.COSTO)/COUNT(PRODUCTOS.CO	STO) 2)>=0		
		DS.COSTO)/COUNT(PRODUCTOS.CO			
HASH	KOOND (SON) (FRODUCTO)3.co310)/co0N1(FR000C103.c0	GROUP BY	1	3
	ESTED LOOPS		GROOF DI	1	1
	NESTED LOOPS			1	1
	NESTED LOOPS			1	1
1 4		MENUSESTABLECIMIENTOS_PK	FULL SCAN	1	1
	TABLE ACCESS	SERVICIOS	BY INDEX ROWID	1	0
		SERVICIOS_PK	UNIQUE SCAN	1	0
			ONLOGE SCAN	-	· ·
	⊕ − O ® Acces	s rieucates			
□ ∧ A					
		CTOS.COSTO)/COUNT(PRODUCTO			
L.		CTOS.COSTO)/COUNT(PRODUCTO	OS.COSTO),2)<=500000		
⊟ ■ HASH			GROUP BY	1	3
	NESTED LOOPS			1	1
	NESTED LOOPS			1	1
				1	-
	INDEX	MENUSBARES_PK	FILL SCAN		
			FULL SCAN	1	
	TABLE ACCE		BY INDEX ROWID	1	0
	□ □ □ INDEX	SERVICIOS_PK	UNIQUE SCAN	1	0
		cess Predicates			
	L .	SERVICIOS.ID=MENUSBARES.BAR	_ID		
	⊞@ INDEX	PRODUCTOS_PK	UNIQUE SCAN	1	0
	Ē Ō‰ Access Pre				
	MENUSB	ARES.PRODUCTO_ID=PRODUCTO	S.ID		
L.F	TABLE ACCESS	PRODUCTOS	BY INDEX ROWID	1	0
	I ADEL ACCESS	CAMPOCIOS	DI INDEX ROVID	1	
FILTER	2.0				
□ O Filte					
-	ROUND(SUM(PRODU	CTOS.COSTO)/COUNT(PRODUCTO	OS.COSTO),2)>=0		
L	ROUND(SUM(PRODU	CTOS.COSTO)/COUNT(PRODUCTO	OS.COSTO),2)<=500000		
HASH			GROUP BY	1	3
	NESTED LOOPS			1	1
	NESTED LOOPS			1	
	NESTED LOOPS	MENI ICECTADI CONTRACTO	K FILL SCAN	1	1
		MENUSESTABLECIMIENTOS_P		1	1
	TABLE ACCE	SS SERVICIOS	BY INDEX ROWID	1	0
	⊟-□€ INDEX	SERVICIOS_PK	UNIQUE SCAN	1	0
		cess Predicates			
	_ <u> </u>				
<u> </u>	AND				
		TOS.COSTO)/COUNT(PRODUCTOS	5.COSTO).2)>=0		
		TOS.COSTO)/COUNT(PRODUCTOS			
HASH			GROUP BY	1	3
			GROOF BT		
	IESTED LOOPS			1	1
	NESTED LOOPS			1	1
E	NESTED LOOPS			1	1
		MENUSBARES_PK	FULL SCAN	1	1
	TABLE ACCES	S SERVICIOS	BY INDEX ROWID	1	0
		SERVICIOS_PK	UNIQUE SCAN	1	0
		ess Predicates		-	
		ERVICIOS.ID=MENUSBARES.BAR_	ID		
	⊞ u∰ INDEX	PRODUCTOS PK	UNIQUE SCAN	1	0
			UNIQUE SCAN	1	0
	☐ OM Access Pred		-		
_		RES.PRODUCTO_ID=PRODUCTOS			
	TABLE ACCESS	PRODUCTOS.	BY INDEX ROWID	1	0
FILTER					
Ē O ₩ Filte	er Predicates				
□ A					
		TOS.COSTO)/COUNT(PRODUCTOS	COSTO) 2)>=0		
-					
		TOS.COSTO)/COUNT(PRODUCTOS			
⊟ ● HASH			GROUP BY	1	3
<u></u>	NESTED LOOPS			1	1
	NESTED LOOPS			1	1
	■ NESTED LOOPS			1	1
		MENUSESTABLECIMIENTOS PK	FULL SCAN	1	
	INDEX				1
	TABLE ACCES		BY INDEX ROWID	1	0
	□ □ □ INDEX	SERVICIOS_PK	UNIQUE SCAN	1	0
		ess Predicates			

RFC5: CONSUMO POR UN USUARIO DADO, EN UN RANGO DE FECHAS

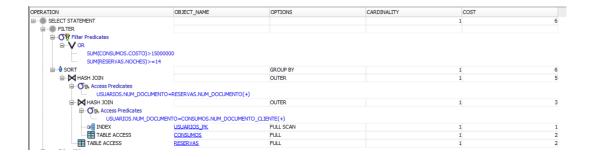


RFC6: OPERACIÓN DE HOTELANDES

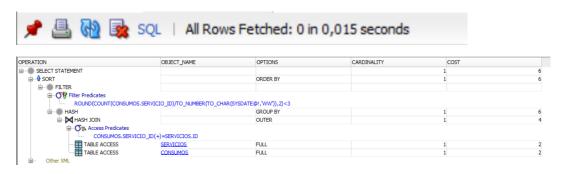


RFC7: BUENOS CLIENTES

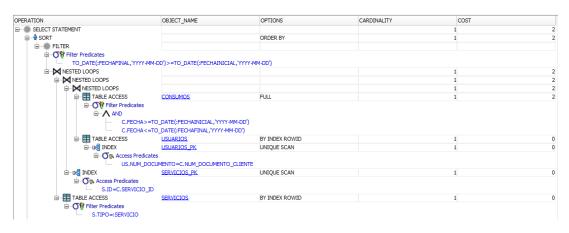




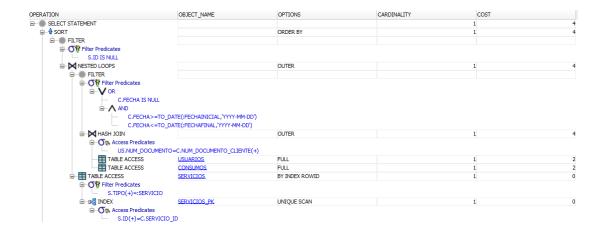
RFC8: SERVICIOS QUE NO TIENEN MUCHA DEMANDA



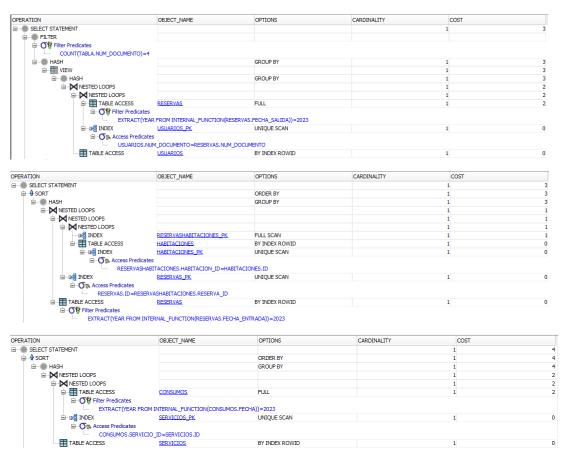
RFC9: CONSUMO EN HOTELANDES



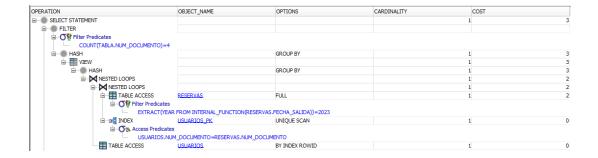
RFC10: CONSUMO EN HOTELANDES v2



RFC11: FUNCIONAMIENTO

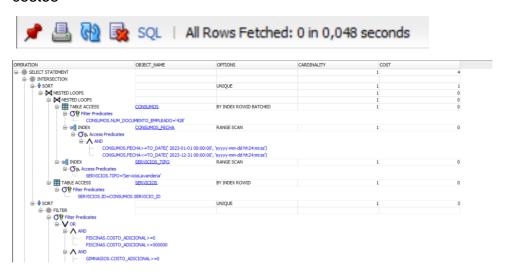


RFC12: CLIENTES EXCELENTES



2.2.1. Análisis de eficiencia

Puede evidenciarse que los índices en las columnas objetivo cumplieron su función de reducir costos y tiempos en las consultas. Por ejemplo, uno de los requerimientos más complejos fue el 4. Con ayuda de índices, se redujo su tiempo a 0.048 segundos y a un costo de 4. Esto representa una mejora del 20.33% en tiempo y del 66,66% en costos



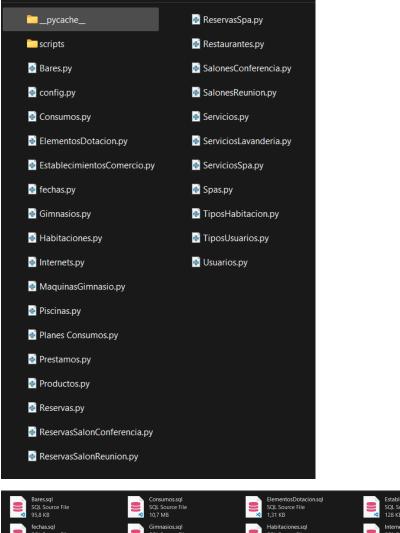
3. DISEÑO Y CARGUE MASIVO DE DATOS

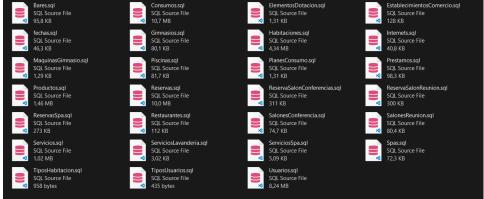
Cada relación en la base de datos está respaldada por un archivo Python dedicado que se encarga de generar datos pseudoaleatorios, como nombres, correos electrónicos y valores numéricos dentro de rangos específicos. Estas operaciones se realizan mediante el uso de las librerias Faker y Random.

Además, para gestionar las claves foráneas, se emplean variables de configuración predefinidas que se encuentran en el archivo config.py. Esto asegura que las tuplas generadas se mantengan en un rango de 1 a n para cada relación, evitando problemas relacionados con la integridad referencial de los datos.

Los datos falsos generados se formatean de manera que coincidan con las consultas de inserción SQL correspondientes. Finalmente, se generan un número predefinido de estas consultas y se escriben en un archivo SQL. En otras palabras, cada archivo Python crea un archivo SQL que contiene sus consultas relacionadas.

No obstante, las relaciones de un tamaño menor a 50 tuplas fueran escritas a mano, usando valores razonables





Clases de asociación

En el caso de las asociaciones donde uno de los extremos tiene una cardinalidad de "muchos", también se elige aleatoriamente la cantidad de elementos relacionados, asegurando que esta cantidad esté dentro de un rango válido, respetando las restricciones de los tipos de datos definidas en los modelos relacionales

Un caso especial de clases de asociación fueron los menús de Bares, Establecimientos de Comercio y Restaurantes, donde se distribuyeron los productos.

Para esto, el menú de un bar serían tuplas contiguas en memoria, y los menús de dos bares contiguos tendrán productos en común. Entonces, un bar podría tener un menú de 50 productos, y 20 de esos productos también estarán en el menú de otro bar, para asemejarse a la vida real.

