

CODING BOOTCAMPS ESPOL

Modulo 2: Database SQL and query optimization

Proyecto Integrador del Módulo

Integrantes

RODAS CRUZ MIA FIORELLA

RODRIGUEZ BUSTAMANTE EMILY ALEJANDRA

TINGO BORBOR CARLOS RAUL

UGALDE ALMEIDA KATTY LORENA

VILLACÍS MORÁN CRISTINA BELÉN

ALLISON BRIGETTE VINUEZA GUTIERREZ

Profesora:

Nicole Agila

Mentora:

Anabella Sanchez

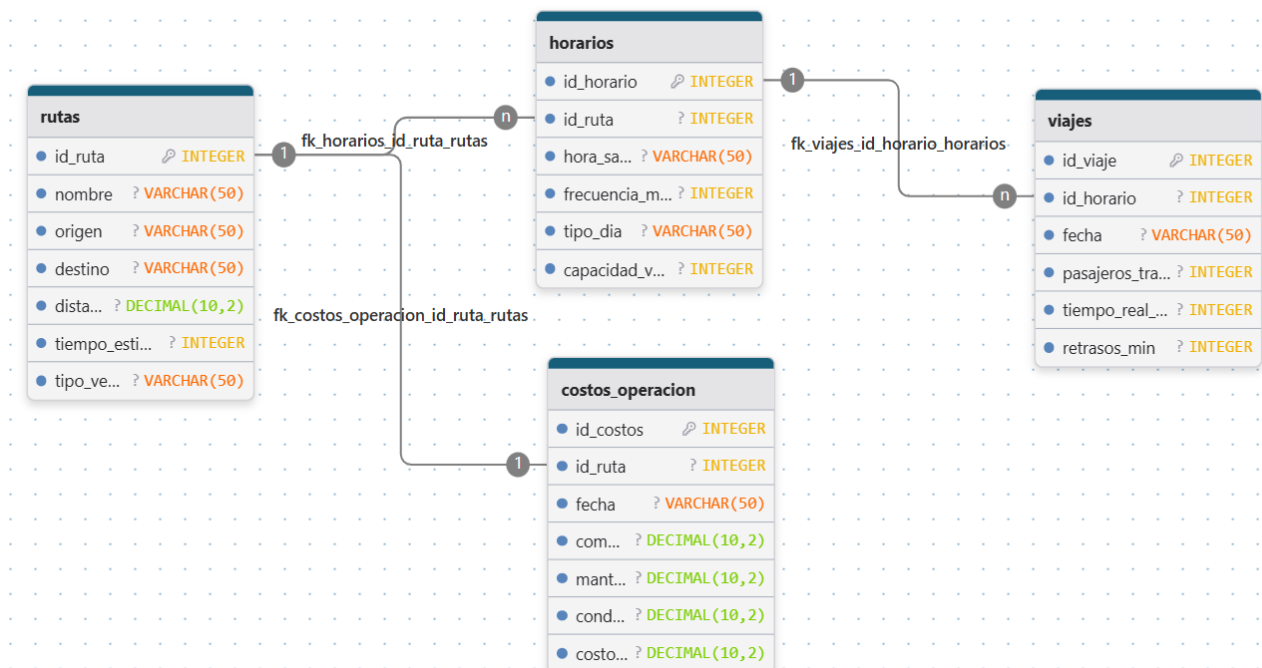
2025 - 2026

¿De qué trata el proyecto?

El objetivo de nuestro proyecto fue diseñar y optimizar una base de datos para analizar las redes de transporte urbano. Buscamos mejorar la eficiencia de las rutas y horarios para reducir los tiempos de viaje y la congestión, basándonos en datos reales. Para lograr esto, creamos una base de datos que permite registrar, analizar y generar reportes.

1. Analizar los datos CSV para identificar entidades, atributos y relaciones

En base a los CSV correspondientes se pudo identificar 5 entidades que son rutas, horarios, costos_operacion y viajes con tipos de datos como integer, varchar y decimal.



2. Aplicar las reglas de normalización hasta Tercera Forma Normal (3FN)

1FN

El modelo cumple con 1FN que asegura atomicidad en cada celda.

2FN

El modelo cumple con 2FN ya que no tiene dependencias parciales.

3FN

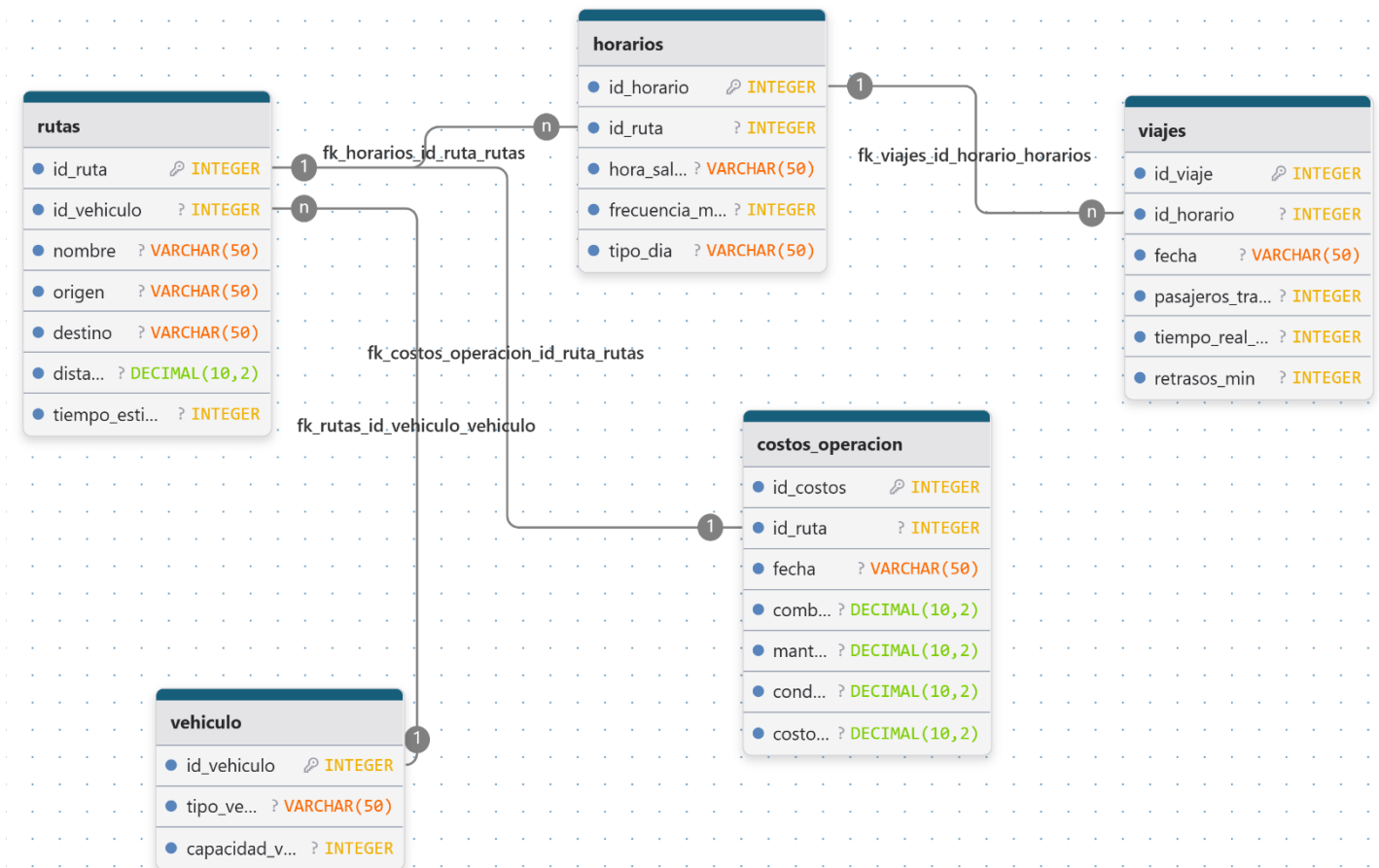
El modelo presenta una dependencia transitiva en la tabla de horarios y rutas; en horarios la columna capacidad_vehiculo que depende de id_ruta y no a su PK id_horario en el caso de la tabla de rutas la columna tipo_vehiculo no tendría mucha relacion por lo cual se recomienda agruparla en una nueva tabla llamada vehiculo.



3. Diseñar un modelo relacional eficiente y sin redundancias

El diseño evita duplicación de información, ya que cada tabla concentra un conjunto específico de datos (vehículos, rutas, horarios, viajes, costos).

Las relaciones establecidas mediante claves primarias y foráneas garantizan la consistencia y trazabilidad, facilitando consultas complejas como ocupación de rutas, retrasos promedio y costo por pasajero.



Planificación urbana – Horas pico

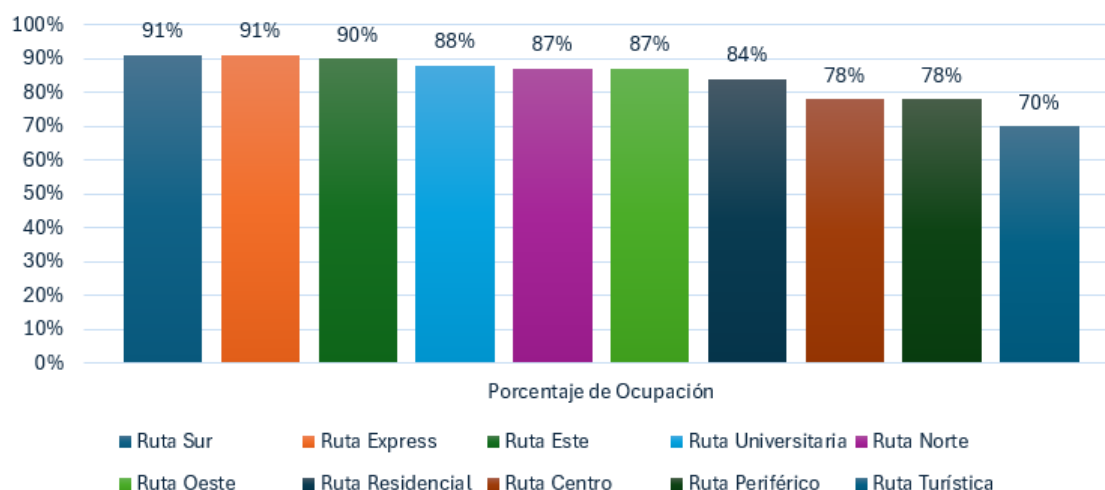
A partir de la siguiente consulta se busca calcular el porcentaje promedio de ocupación de los vehículos en las horas pico (06:00–09:00 y 17:00–19:00), agrupando los resultados por cada ruta para identificar cuáles presentan mayor demanda en esos intervalos de tiempo.

```
75 -- Planificación urbana
76 -- Horas pico
77 • SELECT T2.nombre,
78      CONCAT(TRUNCATE(AVG(T1.pasajeros_transportados / T3.capacidad) * 100, 0), '%')
79      AS porcentaje_ocupacion
80 FROM Viajes AS T1
81 JOIN Horarios AS T4 ON T1.id_horario = T4.id_horario
82 JOIN Rutas AS T2 ON T4.id_ruta = T2.id_ruta
83 JOIN Vehiculos AS T3 ON T2.tipo_vehiculo = T3.tipo_vehiculo
84 WHERE TIME(T4.hora_salida) >= '06:00:00' AND TIME(T4.hora_salida) < '09:00:00'
85        OR TIME(T4.hora_salida) >= '17:00:00' AND TIME(T4.hora_salida) < '19:00:00'
86 GROUP BY T2.nombre
87 ORDER BY porcentaje_ocupacion DESC;
88
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

nombre	porcentaje_ocupacion
Ruta Sur	91%
Ruta Express	91%
Ruta Este	90%
Ruta Universitaria	88%
Ruta Norte	87%
Ruta Oeste	87%
Ruta Residencial	84%
Ruta Centro	78%
Ruta Periférico	78%
Ruta Turística	70%

Planificación urbana – Horas pico



Conclusión

La consulta muestra que las rutas Sur, Express y Este superan el 90% de ocupación, evidenciando la necesidad de asignar vehículos de mayor capacidad y frecuencias más cortas en esos tramos.

nombre	tipo_vehiculo	capacidad_vehiculo
Ruta Norte	Autobús	40
Ruta Sur	Articulado	80
Ruta Este	Minibús	25
Ruta Oeste	Autobús	40
Ruta Centro	Articulado	80
Ruta Periférico	Articulado	80
Ruta Residencial	Minibús	25
Ruta Express	Autobús	40
Ruta Universitaria	Minibús	25
Ruta Turística	Minibús	25

Gestión temporal – Horarios menor uso

En la siguiente consulta se desea detectar los horarios y días con menor cantidad promedio de pasajeros, lo que sirve como base para ajustar la oferta de transporte y optimizar recursos.

```
89 -- Gestión temporal
90 -- Horarios con menor uso de transporte
91 • SELECT h.hora_salida,
92        h.tipo_dia,
93        ROUND(AVG(v.pasajeros_transportados),2) AS promedio_pasajeros
94 FROM Viajes v
95 JOIN Horarios h ON v.id_horario = h.id_horario
96 GROUP BY h.hora_salida, h.tipo_dia
97 ORDER BY promedio_pasajeros ASC;
```

hora_salida	tipo_dia	promedio_pasajeros
10:15:00	Fin_semana	10.50
09:30:00	Fin_semana	12.50
09:45:00	Fin_semana	14.00
13:00:00	Fin_semana	15.00
08:00:00	Fin_semana	15.50
10:00:00	Fin_semana	15.60
16:00:00	Laboral	18.50
12:00:00	Laboral	19.50
09:15:00	Fin_semana	24.50
09:00:00	Fin_semana	25.67
09:00:00	Laboral	28.00
08:30:00	Fin_semana	28.50
05:30:00	Laboral	28.50
08:30:00	Laboral	32.00
10:30:00	Laboral	34.00

Conclusión

Los resultados señalan que, durante los fines de semana, especialmente entre las 09:30 y 10:15, la demanda promedio es la más baja (entre 10 y 15 pasajeros). Esto sugiere que en esas franjas horarias podría reducirse la frecuencia de viajes para optimizar recursos.

Hora	Tipo de día	Promedio de Pasajeros
10:15:00	Fin semana	10,5
9:30:00	Fin semana	12,5
9:45:00	Fin semana	14,00
13:00:00	Fin semana	15,00
8:00:00	Fin semana	15,50
10:00:00	Fin semana	15,60
16:00:00	Laboral	18,50
12:00:00	Laboral	19,50
9:15:00	Fin semana	24,50

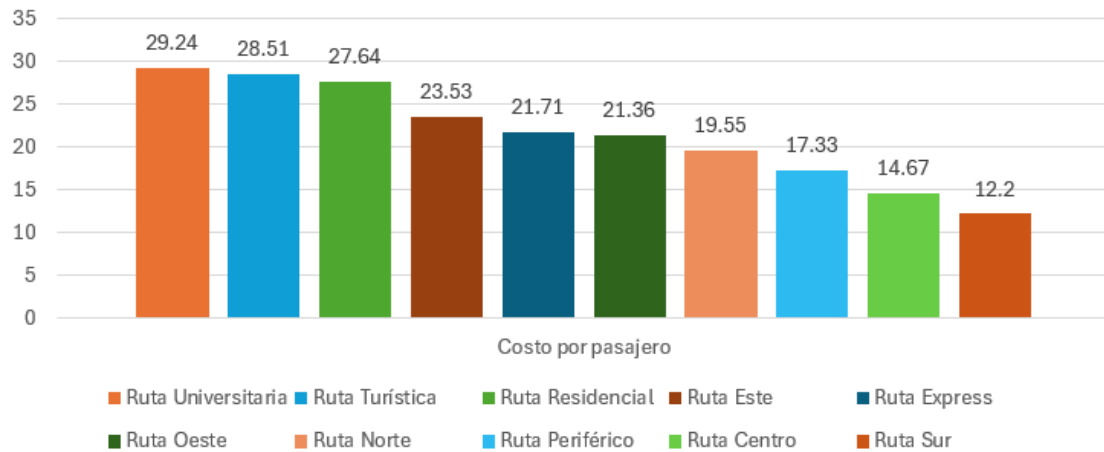
Métricas – Costo por pasajero

Una métrica clave para evaluar la optimización de las rutas es el costo por pasajero. La consulta permite identificar las rutas con mayor costo por pasajero a este indicador se suman el tiempo promedio de viaje y el retraso promedio, que permiten analizar eficiencia y puntualidad operativa.

```
102 -- Métricas
103 -- Costo Por Pasajero
104 • SELECT T2.nombre,
105 TRUNCATE(SUM(T1.costo_total) / SUM(T4.pasajeros_transportados), 2) AS costo_por_pasajero
106 FROM costos_operacion AS T1
107 INNER JOIN rutas AS T2 ON T1.id_ruta = T2.id_ruta
108 INNER JOIN horarios AS T3 ON T2.id_ruta = T3.id_ruta
109 INNER JOIN viajes AS T4 ON T3.id_horario = T4.id_horario AND T1.fecha = T4.fecha
110 GROUP BY T2.nombre
111 ORDER BY costo_por_pasajero DESC;
```

nombre	costo_por_pasajero
Ruta Universitaria	29.24
Ruta Turística	28.51
Ruta Residencial	27.64
Ruta Este	23.53
Ruta Express	21.71
Ruta Oeste	21.36
Ruta Norte	19.55
Ruta Periférico	17.33
Ruta Centro	14.67
Ruta Sur	12.20

Métricas – Costo por pasajero



Conclusión

Los resultados indican mayores valores en rutas como la Universitaria (29,24), Turística (28,51) y Residencial (27,64), reflejando menor eficiencia, mientras que la Sur (12,20) y la Periférico (16,70) son más rentables.

Usando la métrica costo por pasajero las rutas como la Universitaria y Turística presentan los valores más altos, indicando menor eficiencia.

Función ventana

Esta consulta utiliza una función ventana con SUM() OVER para calcular el acumulado de pasajeros transportados en cada ruta a lo largo del tiempo, sin perder el detalle por fecha.

```
113 -- FUNCION VENTANA
114 -- Evolución de pasajeros acumulados por ruta
115 -- Esta consulta muestra cómo se acumulan los pasajeros transportados
116 -- en cada ruta a lo largo del tiempo (fecha).
117 -- Permite analizar la tendencia de crecimiento y detectar picos de demanda.
118
119 • SELECT r.nombre AS ruta,
120        v.fecha,
121        v.pasajeros_transportados,
122        SUM(v.pasajeros_transportados)
123          OVER (PARTITION BY r.nombre ORDER BY v.fecha) AS acumulado_pasajeros
124 FROM Viajes v
125 JOIN Horarios h ON v.id_horario = h.id_horario
126 JOIN Rutas r ON h.id_ruta = r.id_ruta
127 ORDER BY r.nombre, v.fecha;
```



```

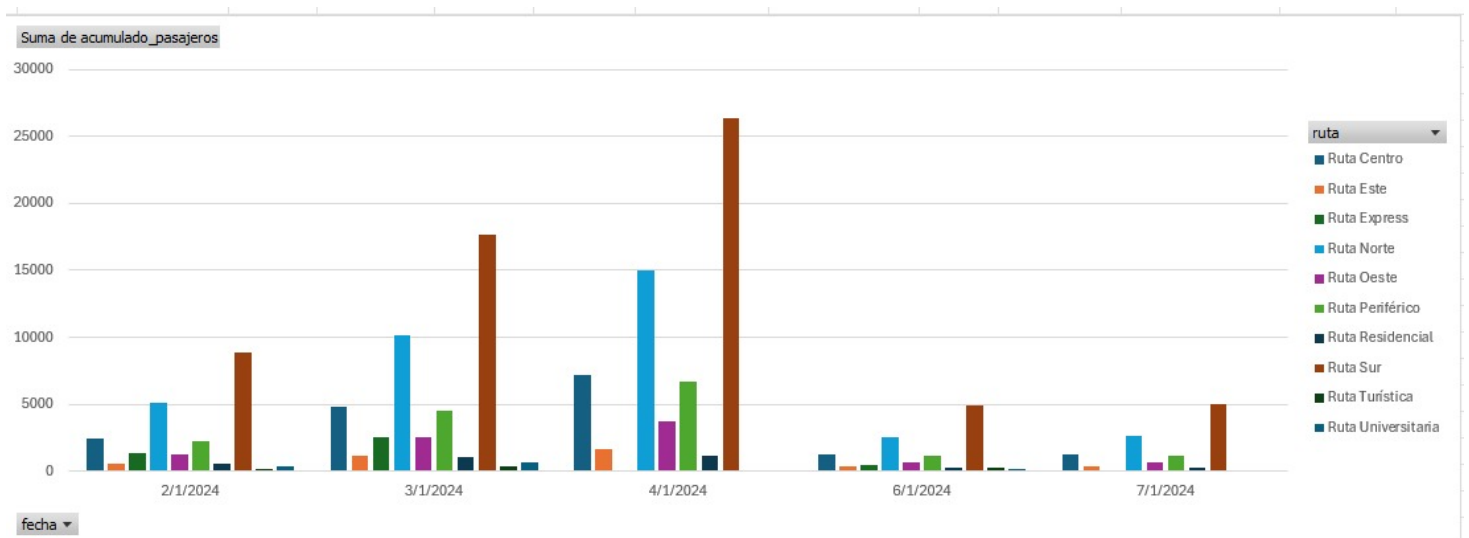
124 FROM Viajes v
125 JOIN Horarios h ON v.id_horario = h.id_horario
126 JOIN Rutas r ON h.id_ruta = r.id_ruta
127 ORDER BY r.nombre, v.fecha;

```

Result Grid Filter Rows: Export: Wrap Cell Content:				
	ruta	fecha	pasajeros_transportados	acumulado_pasajeros
▶	Ruta Centro	2024-01-02	62	403
	Ruta Centro	2024-01-02	66	403
	Ruta Centro	2024-01-02	65	403
	Ruta Centro	2024-01-02	72	403
	Ruta Centro	2024-01-02	68	403
	Ruta Centro	2024-01-02	70	403
	Ruta Centro	2024-01-03	61	800
	Ruta Centro	2024-01-03	67	800
	Ruta Centro	2024-01-03	71	800
	Ruta Centro	2024-01-03	65	800
	Ruta Centro	2024-01-03	69	800
	Ruta Centro	2024-01-03	64	800
	Ruta Centro	2024-01-04	66	1191
	Ruta Centro	2024-01-04	70	1191
	Ruta Centro	2024-01-04	63	1191
	Ruta Centro	2024-01-04	68	1191
	Ruta Centro	2024-01-04	64	1191
	Ruta Centro	2024-01-04	60	1191
	Ruta Centro	2024-01-06	28	1219
	Ruta Centro	2024-01-07	29	1248
	Ruta Este	2024-01-02	25	118

Análisis total acumulado de pasajeros

De esta manera se observa no solo cuántos pasajeros viajaron en un día específico, sino también cómo va creciendo el total acumulado día tras día. Este enfoque es útil porque permite analizar la tendencia de la demanda, identificar picos de crecimiento y evaluar la evolución del servicio por ruta, lo que facilita decisiones sobre asignación de recursos y planificación operativa.



Subconsulta

Esta subconsulta realiza un análisis de puntualidad al calcular el retraso promedio por ruta y compararlo con el promedio global del sistema. Para ello, se obtiene el AVG(retrasos_min) de cada ruta y se aplica un HAVING que filtra únicamente aquellas que superan el retraso promedio de todos los viajes. Los resultados muestran que las rutas Sur (4.00 min), Periférico (3.70 min), Norte (3.63 min) y Centro (3.60 min) presentan mayores retrasos que el promedio global, lo que indica que requieren especial atención en la planificación operativa y posibles ajustes en horarios o frecuencias.

```
135 • SELECT r.nombre,  
136         ROUND(AVG(v.retrasos_min),2) AS retraso_promedio  
137 FROM Viajes v  
138 JOIN Horarios h ON v.id_horario = h.id_horario  
139 JOIN Rutas r ON h.id_ruta = r.id_ruta  
140 GROUP BY r.nombre  
141 HAVING retraso_promedio > (  
142     SELECT ROUND(AVG(retrasos_min),2)  
143     FROM Viajes  
144 );
```

nombre	retraso_promedio
Ruta Norte	3.63
Ruta Sur	4.00
Ruta Centro	3.60
Ruta Periférico	3.70



Optimización de consulta

Se implementó un índice en la columna fecha de la tabla viajes para agilizar las búsquedas y joins por fecha, y un índice compuesto en costos_operacion (id_ruta, fecha) para mejorar el rendimiento en consultas que combinan ruta y fecha. Posteriormente, se utilizó el comando EXPLAIN sobre la consulta de costo por pasajero, lo que permitió verificar que el motor de MySQL aprovecha estos índices, reduciendo la cantidad de filas escaneadas y mejorando la eficiencia en la ejecución. Este proceso asegura que las consultas sean más rápidas y escalables conforme crezca el volumen de datos.

```
151 • CREATE INDEX idx_fecha_viajes ON viajes(fecha);
152
153 -- índice compuesto en costos_operacion para consultas por ruta y fecha
154 • CREATE INDEX idx_ruta_fecha_costos ON costos_operacion(id_ruta, fecha);
155
156 • EXPLAIN
157 SELECT r.nombre AS ruta,
158        TRUNCATE(SUM(c.costo_total) / NULLIF(SUM(v.pasajeros_transportados),0), 2) AS costo_por_pasajero
159 FROM costos_operacion c
160 JOIN Rutas r ON c.id_ruta = r.id_ruta
161 JOIN Horarios h ON r.id_ruta = h.id_ruta
162 JOIN Viajes v ON h.id_horario = v.id_horario AND c.fecha = v.fecha
163 GROUP BY r.nombre
164 ORDER BY costo_por_pasajero DESC;
```

Result Grid											
Filter Rows:											
Export: Wrap Cell Contents:											
id	select_type	table	partitions	type	possible_keys	key	key_len	ref	rows	filtered	Extra
1	SIMPLE	r	NULL	ALL	PRIMARY	id_ruta	5	transporte_urbano.r.id_ruta	10	100.00	Using temporary; Using filesort
1	SIMPLE	h	NULL	ref	PRIMARY, id_ruta	id_ruta	5	transporte_urbano.h.id_ruta	7	100.00	Using index
1	SIMPLE	v	NULL	ref	id_horario, idx_fecha_viajes	id_horario	5	transporte_urbano.h.id_horario	2	100.00	NULL
1	SIMPLE	c	NULL	ref	idx_ruta_fecha_costos	idx_ruta_fecha_costos	8	transporte_urbano.r.id_ruta, transporte_urban...	1	100.00	NULL

Conclusión

El modelo de base de datos y las consultas desarrolladas permitieron analizar de manera integral la operación del sistema de transporte urbano. Los resultados muestran patrones de alta demanda en rutas como Sur, Express y Este, baja utilización en ciertos horarios de fin de semana y diferencias significativas en la eficiencia económica y puntualidad de las rutas. Esta información constituye una herramienta clave para la toma de decisiones, ya que facilita la optimización de frecuencias, la asignación de vehículos adecuados y la mejora en el control de costos, contribuyendo a un servicio más eficiente y orientado a las necesidades de los usuarios.