

Seguimiento de Animales en Peligro de Extinción

Grupo 4

Representación del modelo de proyecto en una tabla general

En base al modelo del proyecto se identificaron las columnas necesarias para organizar la información en una tabla general en Excel, para luego normalizarlas y posteriormente integrarla en el script de SQL.

En esta versión se usaron nombres de columnas más descriptivos para evitar confusiones. Sin embargo, al momento de integrar el diseño de las tablas en SQL los nombres serán más simples. Por ejemplo, aquí se usa `especie_nombre_cientifico`, pero en SQL, dentro de la tabla `Especies`, solo se usará `nombre_cientifico`.

- Se incluyó `registro_id` como identificador único de cada registro.
- Para nombrar a las especies se definieron dos columnas:
 - Nombre común: `especie_nombre_comun`
 - Nombre científico: `especie_nombre_cientifico`
- Para el estado de conservación:
 - Nombre del estado: `estado Conservacion_nombre`
 - Descripción del estado: `nombre_cientifico`
- Para las ubicaciones:
 - Nombre de la ubicación: `ubicacion_nombre`
 - Coordenadas de la ubicación: `ubicacion_coordenadas`
- Para la fecha y hora del avistamiento: `avistamiento_fecha_hora`
- Para el observador del avistamiento:
 - Nombre del observador: `observador_nombre`
 - Correo electrónico del observador: `observador_email`
- Para registrar las amenazas de cada especie se definió `amenaza`

Tabla general de conservación

registro_id	especie_nombre_cientifico	especie_nombre_comun	estado Conservacion_nombre	estado Conservacion_descripcion	ubicacion_nombre
1	Rhinoceros sondaicus	Rinoceronte de Java	En Peligro	La caza furtiva para la venta ilegal de sus cuernos es la principal causa de su peligro, así como la destrucción de su hábitat.	Isla De Java, Indonesia
2	Gorilla beringei beringei	Gorila de Montaña	En Peligro	La pérdida de su hábitat y la caza furtiva son factores clave que ponen en riesgo su supervivencia.	Montañas Virunga, Ruanda
3	Phocoena sinus	Vaquita Marina	En Peligro	Se encuentra en peligro crítico de extinción debido a la pesca incidental con redes ilegales.	Golfo De California, México
4	Panthera tigris sumatrae	Tigre de Sumatra	En Peligro	La deforestación y la caza para el comercio de sus pieles y órganos utilizados en la medicina tradicional han reducido drásticamente sus poblaciones.	Isla De Sumatra, Indonesia
5	Pongo pygmaeus	Orangután de Borneo	En Peligro	La tala de bosques y la expansión agrícola han llevado a la destrucción de su hábitat, afectando gravemente su población.	Isla de Borneo, Indonesia

Tabla 1a. Tabla general de conservación (parte 1)

ubicacion_coordenadas	avistamientos_fecha_hora	observador_nombre	observador_email	amenaza	impacto_amenaza
"-6.595275, 107.556447"	"2024-01-01 13:00:00;2024-05-03 14:00:00"	Lionel Messi	lionelmessi@gmail.com	"pérdida de hábitat, caza furtiva"	Alto
"-1.429596, 29.547217"	"2024-01-01 13:00:00;2024-05-03 14:00:00"	Moises Caicedo	moisescalcedo@gmail.com	"pérdida de hábitat, caza furtiva"	Medio
"26.874306, -112.407191"	"2024-01-01 13:00:00;2024-05-03 14:00:00"	Taylor Swift	taylorswift@gmail.com	"pesca ilegal"	Alto
"-0.076812, 101.376294"	"2024-01-01 13:00:00;2024-05-03 14:00:00"	Charlie Yamal	charlieyamal@gmail.com	"pérdida de hábitat, caza furtiva"	Alto
"0.950946, 114.983948"	"2024-01-01 13:00:00;2024-05-03 14:00:00"	William Martin	wiliammartin@gmail.com	"pérdida de hábitat"	Alto

Tabla 1b. Tabla general de conservación (parte 2)

Normalización aplicada

1. Primera Forma Normal – Celdas atómicas

A partir de la primera forma normal se comienza a separar la tabla general en varias tablas.

En la tabla se identificaron varios problemas. El campo `avistamiento_fecha_hora` incluía varias fechas dentro de una misma celda, el campo `amenaza` contenía varias amenazas y en el campo `ubicacion_coordenadas` se mezclaban la latitud y la longitud en un mismo registro.

Correcciones:

- Cada avistamiento pasó a representarse como un registro independiente, con una única fecha y hora. Además, cada avistamiento representa la relación entre una especie y una ubicación.
- Se creó una tabla separada para las amenazas, junto con una intermedia que relaciona especies y amenazas, dado que existe una relación de muchos a muchos entre ambas.
- Las coordenadas de la ubicación se dividieron en dos columnas de latitud y longitud.

Tablas:

ESPECIES				
registro_id	especie_nombre_cientifico	especie_nombre_comun	estado Conservacion_nombre	estado Conservacion_descripcion
1	Rhinoceros sondaicus	Rinoceronte de Java	En Peligro	La caza furtiva para la venta ilegal de sus cuernos es la principal causa de su peligro, así como la destrucción de su hábitat.
2	Gorilla beringei beringei	Gorila de Montaña	En Peligro	La pérdida de su hábitat y la caza furtiva son factores clave que ponen en riesgo su supervivencia.
3	Phocoena sinus	Vaquita Marina	En Peligro	Se encuentra en peligro crítico de extinción debido a la pesca incidental con redes ilegales.

4	Panthera tigris sumatrae	Tigre de Sumatra	En Peligro	La deforestación y la caza para el comercio de sus pieles y órganos utilizados en la medicina tradicional han reducido drásticamente sus poblaciones.
5	Pongo pygmaeus	Orangután de Borneo	En Peligro	La tala de bosques y la expansión agrícola han llevado a la destrucción de su hábitat, afectando gravemente su población.

UBICACIONES				
ubicacion_id	ubicacion_nombre	latitud	longitud	
1	Isla De Java, Indonesia	"-6.595275"	"107.556447"	
2	Montañas Virunga, Ruanda	"-1.429596"	"29.547217"	
3	Golfo De California, México	"26.874306"	"-112.407191"	
4	Isla De Sumatra, Indonesia	"-0.076812"	"101.376294"	
5	Isla de Borneo, Indonesia	"0.950946"	"114.983948"	

OBSERVADORES		
observador_id	observador_nombre	observador_email
1	Lionel Messi	lionelmessi@gmail.com
2	Moises Caicedo	moisesc caicedo@gmail.que
3	Taylor Swift	taylorswift@gmail.la
4	Charlie Yamal	charlieyamal@gmail.tra
5	William Martin	wiliam martin@gmail.su

AMENAZAS	
amenaza_id	categoria
1	Perdida de Habitat
2	Pesca Ilegal
3	Caza Furtiva

AVISTAMIENTOS				
avistamiento_id	especie_id	ubicacion_id	observador_id	avistamiento_fecha_hora
1	1	1	1	"2024-01-01 13:00:00"
2	1	1	1	"2024-05-03 14:00:00"
3	2	2	2	"2024-01-01 13:00:00"
4	2	2	2	"2024-05-03 14:00:00"
5	3	3	3	"2024-01-01 13:00:00"
6	3	3	3	"2024-05-03 14:00:00"
7	4	4	4	"2024-01-01 13:00:00"
8	4	4	4	"2024-05-03 14:00:00"
9	5	5	5	"2024-01-01 13:00:00"
10	5	5	5	"2024-05-03 14:00:00"

ESPECIES_AMENAZAS		
especie_id	amenaza_id	impacto_amenaza
1	1	Alto
1	3	Medio
2	1	Alto
2	3	Alto
3	2	Alto
4	1	Alto
4	3	Medio
5	1	Bajo

2. Segunda Forma Normal – eliminar dependencias parciales

En la segunda forma normal se busca evitar las dependencias parciales. En la tabla de avistamientos podía ocurrir que el nombre y el correo del observador dependieran únicamente del observador y no de todo el registro, lo que llevaba a repeticiones innecesarias. Como desde la primera forma normal se separaron los observadores en su propia tabla y en avistamientos quedó solo su identificador, este problema ya estaba resuelto y el modelo cumple con la segunda forma normal.

3. Tercera Forma Normal – eliminar dependencias transitivas

En la tabla Especies existían dependencias transitivas ya que estado Conservacion_descripcion dependía de estado Conservacion_nombre y no directamente de la clave principal de especie.

Correcciones:

- Se creó una tabla independiente llamada EstadoConservacion, con un identificador, el nombre del estado y su descripción.
- En la tabla de especies quedó únicamente el identificador del estado como clave foránea, en lugar de repetir el nombre y la descripción.
- Para permitir que una especie tenga distintos estados de conservación según la región, se creó la tabla Especie_Estado.
- Para resolver las relaciones de muchos a muchos, se crearon las tablas intermedias Especie_Ubicacion y Especies_Amenazas.

Tablas finales hasta la 3FN para SQL

1. EstadoConservacion

- PK: estado_id
- Atributos: nombre, descripción

EstadoConservacion		
estado_codigo	nombre	descripcion
NP	NO PELIGRO	Tiene una amplia población sin amenazas directas
EP	EN PELIGRO	Su población ha disminuido drásticamente y presenta 1 amenaza directa o más
CA	CASI AMENAZADO	Podría estar en riesgo en un futuro cercano

2. Especie

- PK: especie_id
- Atributos: nombre_cientifico, nombre_comun

ESPECIES		
especies_id	nombre_cientifico	nombre_comun
1	Rhinoceros sondaicus	Rinoceronte de Java
2	Gorilla beringei beringei	Gorila de Montaña
3	Phocoena sinus	Vaquita Marina
4	Panthera tigris sumatrae	Tigre de Sumatra
5	Pongo pygmaeus	Orangután de Borneo

3. Especie_Estado

- PK: especie_id, estado_codigo
- Atributos: region
- FK: especie_id y estado_codigo
- Esta tabla intermedia resuelve la relación muchos a muchos entre Especie y EstadoConservacion. Una especie puede tener distintos estados de conservación según la región o el nivel de análisis. Un mismo estado de conservación puede aplicar a muchas especies.

ESPECIE_ESTADO		
especie_id	estado_codigo	region
1	EP	Global
1	NP	Indonesia
2	EP	Global
3	EP	Global
4	EP	Global
5	EP	Global

4. Ubicación

- PK: ubicacion_id
- Atributos: nombre, latitud, longitud

UBICACION			
ubicacion_id	nombre	latitud	longitud
1	Isla De Java, Indonesia	"-6.595275"	"107.556447"
2	Montañas Virunga, Ruanda	"-1.429596"	"29.547217"
3	Golfo De California, México	"26.874306"	"-112.407191"
4	Isla De Sumatra, Indonesia	"-0.076812"	"101.376294"
5	Isla de Borneo, Indonesia	"0.950946"	"114.983948"

5. Especie_Ubicacion

- PK: especie_id, ubicacion_id
- Atributos: region
- FK: especie_id y ubicación_id
- Esta tabla intermedia resuelve la relación muchos a muchos entre Especie y EstadoConservacion, Una especie puede encontrarse en varias ubicaciones.

ESPECIE_UBICACION		
especie_id	ubicacion_id	region
1	1	Indonesia
2	2	África Central
3	3	Golfo de California
4	4	Indonesia
5	5	Indonesia

6. Observador

- PK: observador_id
- Atributos: nombre, email

OBSERVADOR		
observador_id	nombre	email
1	Lionel Messi	lionelmessi@gmail.com
2	Moises Caicedo	moisescacedo@gmail.que
3	Taylor Swift	taylorswift@gmail.la
4	Charlie Yamal	charlieyamal@gmail.tra
5	William Martin	wiliammartin@gmail.su

7. Amenaza

- PK: amenaza_id
- Atributos: categoría

AMENAZA	
amenaza_id	categoría
1	Perdida de Habitat
2	Pesca Ilegal
3	Caza Furtiva

8. Especies_Amenazas

- PK compuesta: especie_id, amenaza_id
- FK: especie_id y amenaza_id que hacen referencia a las tablas Especies y Amenazas. Se da una relación de muchos a muchos porque una especie puede tener varias amenazas y una misma amenaza puede afectar a varias especies. Esta relación se resuelve mediante la tabla intermedia Especies_Amenazas.
- Atributo: impacto

ESPECIES_AMENAZAS		
especie_id	amenaza_id	impacto
1	1	Alto
1	3	Medio
2	1	Alto
2	3	Alto
3	2	Alto
4	1	Alto
4	3	Medio
5	1	Bajo

9. Especies_Amenazas_Registro

- PK: registro_id
- Atributos: fecha_registro
- FK: especie_id, observador_id y amenaza_id que hacen referencia a las tablas Especies, Observador y Amenazas respectivamente.

Especies_Amenazas_Registro				
registro_id	especie_id	amenaza_id	fecha_registro	observador_id
1	1	1	"2024-01-01"	1
2	1	1	"2024-01-01"	2
3	1	3	"2024-01-01"	3
4	2	1	"2024-01-01"	4
5	2	3	"2024-01-01"	5
6	3	2	"2024-01-01"	1
7	3	2	"2024-01-01"	2

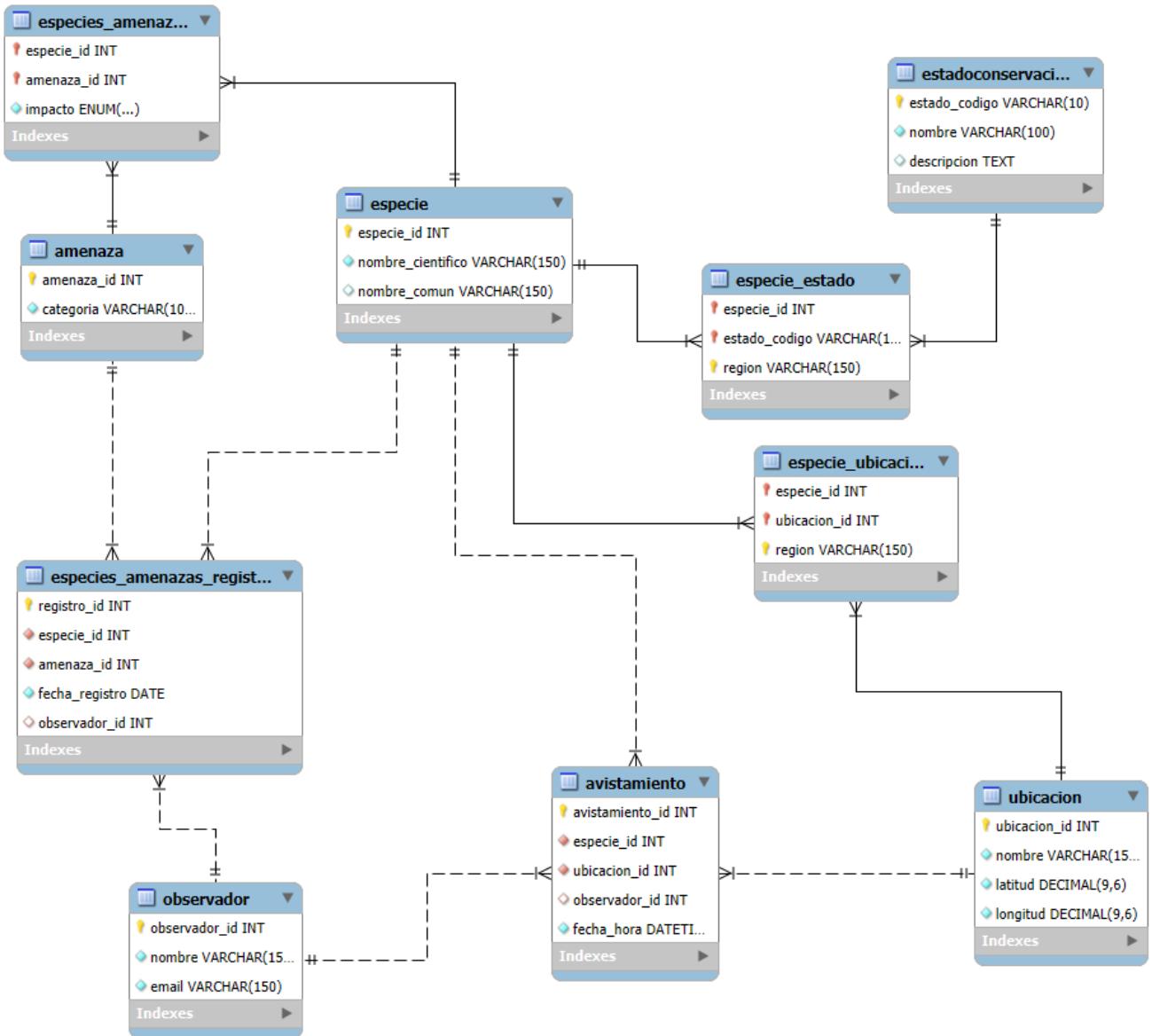
10. Avistamiento

- PK: avistamiento_id
- Atributo: fecha_hora
- FK: especie_id que hace referencia a la tabla Especies. Se da una relación de muchos a uno porque varios avistamientos pueden corresponder a una misma especie.
- FK: ubicacion_id que hace referencia a la tabla Ubicaciones. Se da una relación de muchos a uno porque varios avistamientos pueden ocurrir en una misma ubicación.

- FK: observador_id que hace referencia a la tabla Observadores. Se da una relación de muchos a uno porque un observador puede registrar varios avistamientos.

AVISTAMIENTO				
avistamiento_id	especie_id	ubicacion_id	observador_id	fecha_hora
1	1	1	1	"2024-01-01 13:00:00"
2	1	1	1	"2024-05-03 14:00:00"
3	2	2	2	"2024-01-01 13:00:00"
4	2	2	2	"2024-05-03 14:00:00"
5	3	3	3	"2024-01-01 13:00:00"
6	3	3	3	"2024-05-03 14:00:00"
7	4	4	4	"2024-01-01 13:00:00"
8	4	4	4	"2024-05-03 14:00:00"
9	5	5	5	"2024-01-01 13:00:00"
10	5	5	5	"2024-05-03 14:00:00"

Modelo de datos documentado con diagramas



Optimización de consultas

1. Especies con mayor declive poblacional

- Explain antes del índice

	id	select_type	table	partitions	type	possible_keys	key	key_len	ref	rows	filtered	Extra
▶	1	SIMPLE	ec	NULL	const	PRIMARY	PRIMARY	42	const	1	100.00	NULL
	1	SIMPLE	ee	NULL	ref	PRIMARY,estado_codigo	estado_codigo	42	const	9	10.00	Using where; Using index
	1	SIMPLE	e	NULL	eq_ref	PRIMARY	PRIMARY	4	g4 Conservacion.ee.especie_id	1	100.00	NULL

- Implementación del índice

```
CREATE INDEX idx_estado_region ON Especie_Estado (estado_codigo, region);
```

- Explain después del índice

	id	select_type	table	partitions	type	possible_keys	key	key_len	ref	rows	filtered	Extra
▶	1	SIMPLE	ec	NULL	const	PRIMARY	PRIMARY	42	const	1	100.00	NULL
	1	SIMPLE	ee	NULL	ref	PRIMARY,idx_estado_region	idx_estado_region	644	const,const	5	100.00	Using where; Using index
	1	SIMPLE	e	NULL	eq_ref	PRIMARY	PRIMARY	4	g4 Conservacion.ee.especie_id	1	100.00	NULL

- Explicación

Antes, MySQL buscaba por el estado de conservación (EP) y luego filtraba las regiones, lo que hacía revisar más filas de las necesarias.

Con el índice en (estado_codigo, region) la base de datos encuentra directamente las especies que cumplen ambas condiciones, reduciendo el tiempo de búsqueda.

2. Regiones con mayor biodiversidad

- Explain antes del índice

	id	select_type	table	partitions	type	possible_keys	key	key_len	ref	rows	filtered	Extra
▶	1	SIMPLE	e	NULL	index	PRIMARY	nombre_cientifico	602	NULL	10	100.00	Using index; Using temporary; Using filesort
	1	SIMPLE	eu	NULL	ref	PRIMARY,ubicacion_id	PRIMARY	4	g4 Conservacion.e.especie_id	1	100.00	Using index

- Implementación del índice

```
CREATE INDEX idx_region ON Especie_Ubicacion (region);
```

- Explain después del índice

	id	select_type	table	partitions	type	possible_keys	key	key_len	ref	rows	filtered	Extra
▶	1	SIMPLE	e	NULL	index	PRIMARY	nombre_cientifico	602	NULL	10	100.00	Using index; Using temporary; Using filesort
	1	SIMPLE	eu	NULL	ref	PRIMARY,ubicacion_id,IDX_REGION	PRIMARY	4	g4 Conservacion.e.especie_id	2	100.00	Using index

- Explicación

Antes, MySQL tenía que recorrer todos los registros y crear tablas temporales para agrupar y ordenar las especies por región.

Con el índice en región, la información ya está organizada y el agrupamiento se hace más rápido, lo que reduce el tiempo de búsqueda cuando la base de datos es más grande.

3. Amenazas más comunes

- Explain antes del índice

	id	select_type	table	partitions	type	possible_keys	key	key_len	ref	rows	filtered	Extra
▶	1	SIMPLE	a	NULL	ALL	PRIMARY	NULL	NULL	NULL	5	100.00	Using temporary; Using filesort
1	SIMPLE	ear	NULL	NULL	ref	especie_id,amenaza_id	amenaza_id	4	g4_conservacion.a.amenaza_id	1	100.00	NULL
1	SIMPLE	e	NULL	NULL	eq_ref	PRIMARY	PRIMARY	4	g4_conservacion.ear.especie_id	1	100.00	NULL
1	SIMPLE	ea	NULL	NULL	eq_ref	PRIMARY,amenaza_id	PRIMARY	8	g4_conservacion.ea.especie_id,g4_conservad...	1	100.00	NULL

- Implementación del índice

```
CREATE INDEX idx_especie_amenaza
ON Especies_Amenazas_Registro (especie_id, amenaza_id);
```

- Explain después del índice

	id	select_type	table	partitions	type	possible_keys	key	key_len	ref	rows	filtered	Extra
▶	1	SIMPLE	a	NULL	ALL	PRIMARY	NULL	NULL	NULL	5	100.00	Using temporary; Using filesort
1	SIMPLE	ea	NULL	NULL	ref	PRIMARY,amenaza_id	amenaza_id	4	g4_conservacion.a.amenaza_id	1	100.00	NULL
1	SIMPLE	e	NULL	NULL	eq_ref	PRIMARY	PRIMARY	4	g4_conservacion.ea.especie_id	1	100.00	NULL
1	SIMPLE	ear	NULL	NULL	ref	amenaza_id,IDX_idx_especie_amenaza	IDX_idx_especie_amenaza	8	g4_conservacion.ea.especie_id,g4_conservacio...	1	100.00	Using index

- Explicación

Antes, MySQL revisaba todos los registros para agrupar y contar las amenazas por especie.

Con el índice, los datos ya están organizados por especie y amenaza, lo que acelera el agrupamiento y reduce el tiempo de búsqueda cuando la base tiene muchos registros.