



# ENTREGA FINAL

SQL FLEX

VICTORIA MARQUEZ

## **SQL y meta análisis**

### **Índice**

Introducción.....	3
Situación problemática.....	3
Objetivo.....	4
Modelo de negocio.....	4
Listado de tablas.....	4
Diagrama conceptual de entidad de relaciones.....	8
Diagrama de entidad de relaciones con ingeniería inversa.....	8
Importación de datos.....	9
Vistas.....	10
Funciones.....	10
Procedimientos.....	11
Triggers.....	11
Backup.....	11
Conclusiones.....	12

## **Introducción**

El cambio climático es uno de los factores de perturbación antrópica más importantes que afectan actualmente a los ecosistemas. Tanto las plantas como sus polinizadores insectos se ven afectados por el cambio climático. Las alteraciones en las fenologías de plantas y polinizadores debido al cambio climático pueden provocar desajustes en sus interacciones, lo que a su vez puede comprometer los servicios de polinización y el funcionamiento de los ecosistemas. Uno de los objetivos de mi proyecto post doctoral es realizar un meta análisis para examinar los efectos del cambio climático sobre las fenologías de floración de plantas y de los polinizadores en estudios realizados en distintos ecosistemas y con diferentes enfoques metodológicos, con el fin de identificar patrones, sesgos metodológicos y posibles desajustes planta-polinizador a nivel global.

Un meta-análisis es una técnica estadística muy utilizada en ecología que tiene como objetivo sintetizar cuantitativamente los resultados de múltiples estudios independientes sobre un mismo tema. Su objetivo es identificar patrones generales, estimar efectos promedio y evaluar la consistencia de los resultados entre estudios. En lugar de basarse en una sola investigación, el meta-análisis reúne datos de muchos trabajos científicos previos, y mediante herramientas estadísticas permite responder preguntas con mayor poder y robustez. Es especialmente útil cuando los estudios individuales muestran resultados variables o contradictorios.

## **Situación problemática**

En las etapas iniciales de un meta-análisis, se generan bases de datos extensas, producto de la recopilación de información de muchos estudios. Estas bases pueden incluir cientos o miles de registros relacionados con especies, ubicaciones, variables fenológicas, parámetros climáticos, años de estudio, entre otros. Dado este volumen y complejidad, el uso de bases de datos relacionales y en particular el lenguaje SQL resulta sumamente útil para organizar, consultar, cruzar y analizar esta información de forma robusta, eficiente y reproducible. Por ejemplo, una base de datos relacional permite: relacionar especies con

sitios de muestreo y condiciones climáticas, clasificar estudios según enfoque metodológico, año o región y consultar con rapidez tendencias específicas o generar subconjuntos de datos para análisis más focalizados.

### Objetivo

Propongo como objetivo diseñar una base de datos funcional que permita integrar datos de estudios fenológicos para analizar desajustes planta-polinizador, generar informes automatizados, validar integridad de los datos y brindar soporte a investigaciones ecológicas y meta-análisis.

### Modelo de negocio

Esta base es útil para instituciones científicas, grupos de investigación o universidades que estudian fenología, polinización y cambio climático. También permite generar reportes automatizados para tomadores de decisiones en conservación.

### Listado de tablas.

- Tabla estudio: contiene información básica de cada trabajo incluido en la base de datos. Cada estudio está vinculado a uno o más **sitios** y a una o múltiples **especies** de plantas, polinizadores o ambos.

Columna	Tipo de dato	Primary key	Foreign key	Not Null	Autoincremental	Notas
<a href="#">id_estudio</a>	INT	True		True	True	id de los estudios
titulo	VARCHAR(100)			True		nombre del estudio
DOI	VARCHAR(100)					Digital Object Identifier
tipo_de_estudio	VARCHAR(100)			True		información sobre los estudios primarios
year_estudio	YEAR			True		año de publicación del estudio

- Tabla planta: almacena datos sobre las especies vegetales estudiadas. Son polinizadas por una o más **especies de polinizadores**

Columna	Tipo de dato	Primary key	Foreign key	Not Null	Autoincremental	Notas
---------	--------------	-------------	-------------	----------	-----------------	-------

id_planta	INT	True		True	True	id de las plantas
nombre_cientifico	VARCHAR(150)			True		nombre científico de cada planta
familia	VARCHAR(100)			True		familia botánica de cada planta
forma_de_vida	VARCHAR(100)			True		información sobre ecológica sobre las plantas
id_estudio	INT		True	True		references to tabla estudio

- Tabla polinizador: almacena datos sobre los polinizadores registrados en los estudios.

Columna	Tipo de dato	Primary key	Foreign key	Not Null	Autoincremental	Notas
id_polinizador	INT	True		True	True	id de los polinizadores
nombre_cientifico	VARCHAR(100)			True		nombre científico de cada polinizador
grupo_funcional	VARCHAR(100)					información sobre ecológica sobre los polinizadores
id_estudio	INT		True	True		references to tabla estudio

- Tabla interacción: almacena datos sobre las interacciones entre plantas y polinizadores observadas en los estudios. Una especie de **plantas y polinizadores** en un **sitio** y fecha determinada

Columna	Tipo de dato	Primary key	Foreign key	Not Null	Autoincremental	Notas
id_interaccion	INT	True		True	True	id de las interacciones
id_planta	INT		True	True		references to planta
id_polinizador	INT		True	True		references to polinizador
id_estudio	INT		True	True		references to tabla estudio

- Tabla sitio: registra los sitios donde se llevaron a cabo los estudios. Lugar donde se registran las interacciones, la **fenología** y los datos de **clima**

Columna	Tipo de dato	Primary key	Foreign key	Not Null	Autoincremental	Notas
<a href="#">id_sitio</a>	INT	True		True	True	id de los sitios
nombre_sitio	VARCHAR(100)			True		nombre de cada sitio
país	VARCHAR(100)					país donde se encuentra el sitio
latitud	DECIMAL(9,6)					referencias geográficas
longitud	DECIMAL(9,6)					referencias geográficas
tipo_ecosistema	VARCHAR(100)					información ecológica de los sitios
<a href="#">id_estudio</a>	INT		True	True		references to estudio

- Tabla **fenologia\_planta**: registra la fenología de la floración de las plantas observadas. Relaciona **plantas**, **sitios** y fechas.

Columna	Tipo de dato	Primary key	Foreign key	Not Null	Autoincremental	Notas
<a href="#">id_fenologia_planta</a>	INT	True		True	True	id de la información fenológica
inicio_floracion	DATE					fecha inicio floración
pico_floracion	DATE					fecha del pico de floración
fin_floracion	DATE					fecha del fin de la floración
<a href="#">id_planta</a>	INT		True	True		references to planta
<a href="#">id_sitio</a>	INT		True	True		references to sitio

- Tabla **fenologia\_polinizador**: registra la actividad fenológica de los polinizadores. Relaciona **polinizadores**, **sitios** y fechas.

Columna	Tipo de dato	Primary key	Foreign key	Not Null	Autoincremental	Notas
<a href="#">id_fenologia_polinizador</a>	INT	True		True	True	id de la información fenológica
inicio_actividad	DATE					fecha inicio

d						actividad
pico_actividad	DATE					fecha del pico de actividad
fin_actividad	DATE					fecha del fin de la actividad
id_polinizador	INT		True	True		references to polinizador
id_sitio	INT		True	True		references to sitio

- Tabla clima: contiene datos climáticos asociados a cada sitio de muestreo. Vinculados a cada **sitio** y pueden usarse para analizar su efecto en la **fenología**

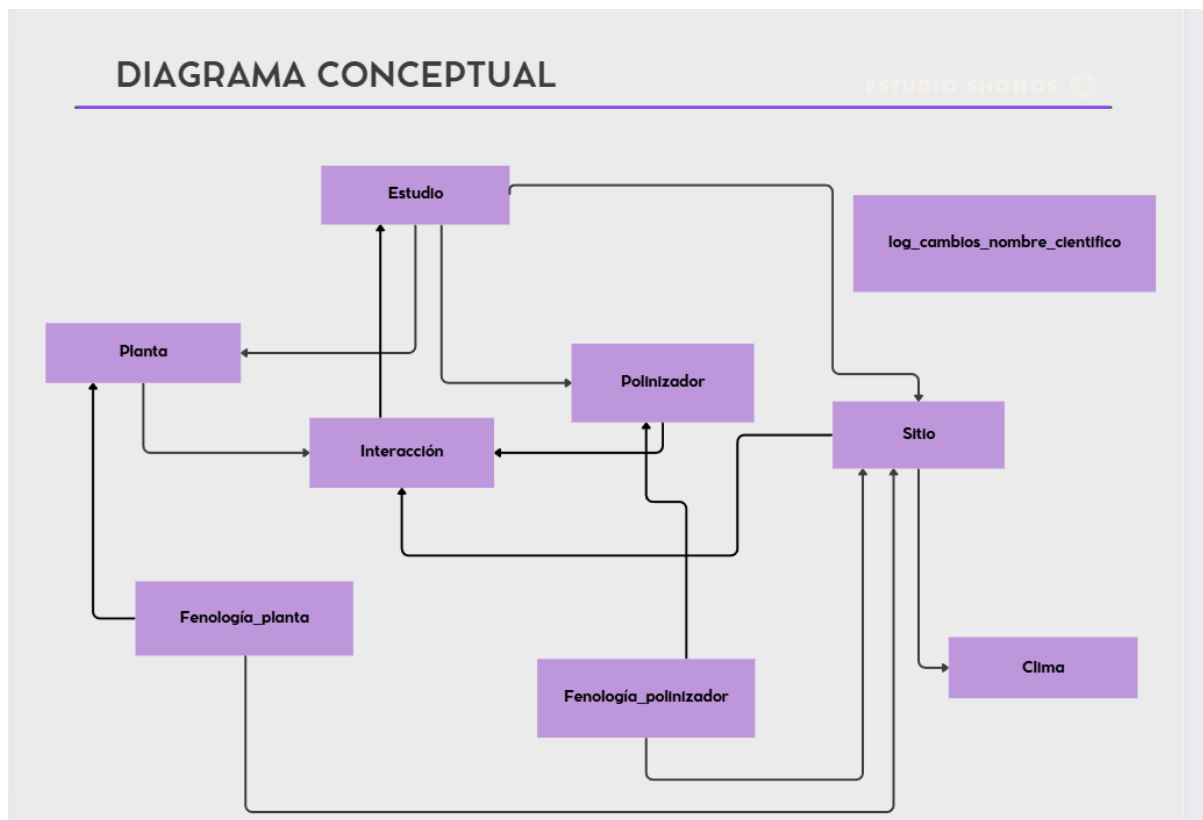
Columna	Tipo de dato	Primary key	Foreign key	Not Null	Autoincremental	Notas
id_clima	INT	True		True	True	id de datos climáticos
temperatura_media	FLOAT					temperatura media
precipitacion	FLOAT					precipitación
id_sitio	INT		True	True		references to sitio

- Tabla log\_cambios\_nombre\_cientifico: contiene los nombres científicos actualizados y los no actualizados tanto de plantas como de polinizadores. Audita cambios en el nombre científico (por ejemplo si se actualiza la taxonomía)

Columna	Tipo de dato	Primary key	Foreign key	Not Null	Autoincremental	Notas
id_log	INT	True		True	True	id de los cambios
nombre_anterior	VARCHAR(150)			True		nombre científico viejo
nombre_nuevo	VARCHAR(150)			True		nombre científico actualizado
tabla	VARCHAR(50)			True		tabla en la cual se realizó el cambio: planta o polinizador
id_registro	INT			True		id_planta o id_polinizador que cambio
usuario	VARCHAR(					guarda el

	100)					usuario que realizó el cambio
--	------	--	--	--	--	-------------------------------

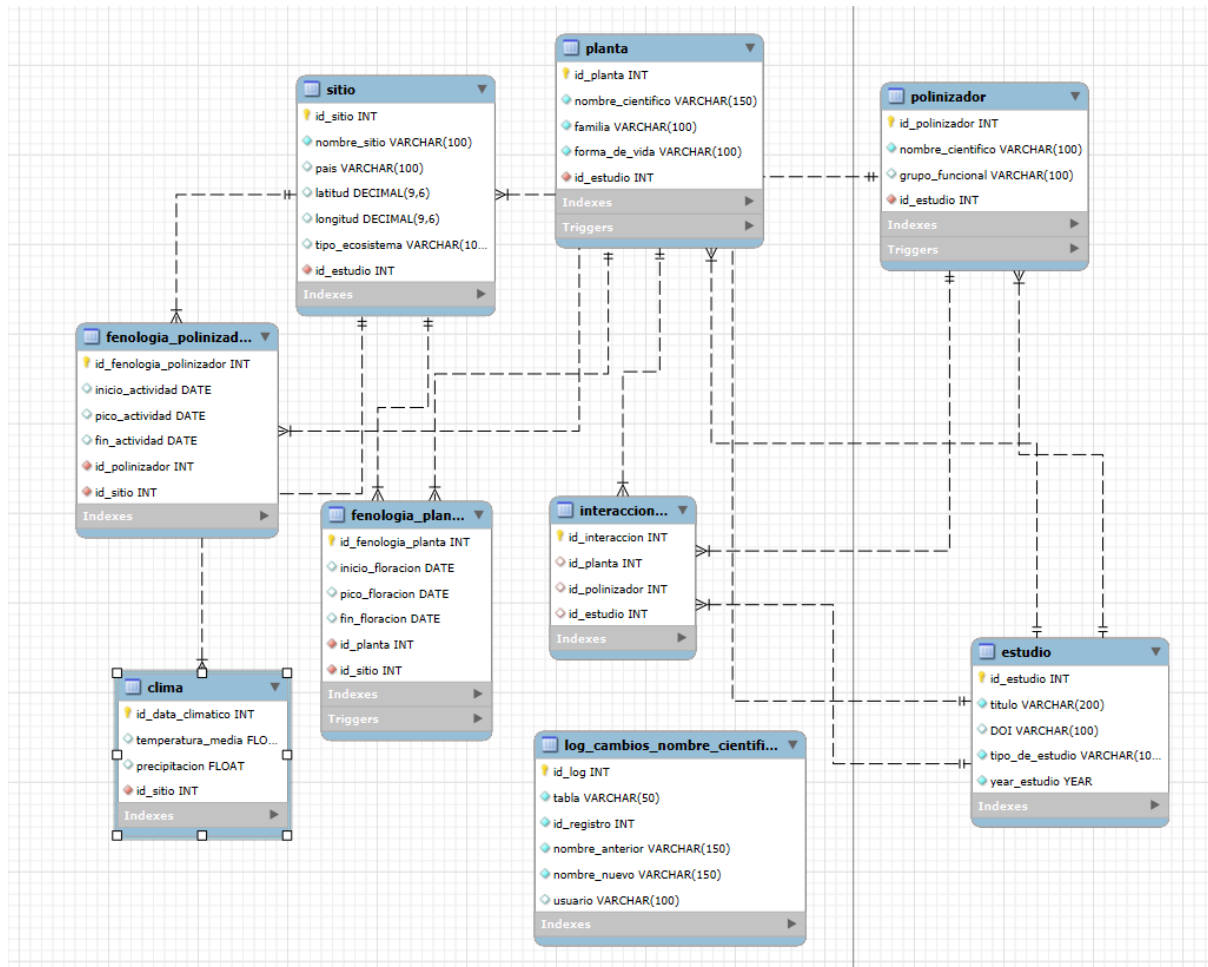
### Diagrama conceptual de entidad de relaciones



**Fig.1** Diagrama de entidad de relaciones.

### Diagrama de entidad de relaciones con ingeniería inversa





## Importación de datos

En este caso, dada la especificidad y complejidad de los datos propuestos utilicé Python para crear archivos csv con valores aleatorios, pero relacionados entre sí. Luego realicé la importación de los datos en MySQL Workbench de la siguiente manera.

1. Seleccioné la tabla destino de los datos que voy a importar
2. Click derecho sobre el nombre de la misma y seleccione *Table Data Import Wizard*
3. En *File Path* elegí la carpeta de origen de mis archivos csv y el archivo en sí mismo que quiero importar
4. Seleccioné la opción *Use existing table* por que mis tablas ya están creadas

5. En la siguiente pestaña seleccioné el *Encoding* utf-8 y finalmente importo la información

Algo importante a chequear es que la identidad, número y nombre de las columnas en el archivo csv coincidan con las columnas de nuestra tabla en MySQL.

### **Vistas**

`vista_fenologia_mismatch`: tiene como objetivo mostrar las fenológicas de las plantas y polinizadores que estén en lo mismos sitios. Es útil para detectar posibles desajustes y solapamientos fenológicos entre plantas y polinizadores. Las tablas involucradas son: `fenologia_planta`, `planta`, `sitio`, `fenologia_polinizador` y `polinizador`.

`vista_interacciones_taxonomia`: me permite ver todas las interacciones planta-polinizador junto a la información taxonómica (nombre científico de polinizador y planta). Las tablas involucradas son: `planta`, `polinizador` e `interacciones`.

`vista_estudios_detalle`: resume por estudio las especies vegetales y polinizadoras involucradas, junto con los sitios donde se realizaron. Sirve para hacer análisis bibliométricos que pueden ser importantes en meta-análisis. Las tablas involucradas son: `estudio`, `planta`, `polinizador`, `sitio`.

### **Funciones**

`duracion_floracion_planta` (`p_id_feno_planta` INT): calcula la duración total en días de la floración de una planta a partir de su fecha de inicio y fin. Tabla involucrada: `fenologia_planta`

`numero_interacciones_por_estudio`(`p_id_estudio` INT): retorna la cantidad de interacciones registradas en un estudio específico. Tabla involucrada: `interacciones`

## Procedimientos

`sp_insertar_interaccion(p_id_planta, p_id_polinizador, p_id_estudio)`: inserta una nueva interacción entre una planta y un polinizador asociada a un estudio. Tabla involucrada: `interacciones`

`ver_fenologia(p_nombre_cientifico)`: devuelve las fechas de floración (inicio, pico y fin) de una planta según su nombre científico. Permite consultar rápidamente la fenología de una especie vegetal específica. Tablas involucradas: `planta` y `fenologia_planta`.

## Triggers

`trg_validar_fechas_fenologia`: verifica al insertar en `fenologia_planta` que la fecha de inicio no sea posterior a la de fin. Garantizar la integridad temporal de los registros fenológicos. Tablas involucradas: `fenologia_planta`. Trigger: BEFORE INSERT

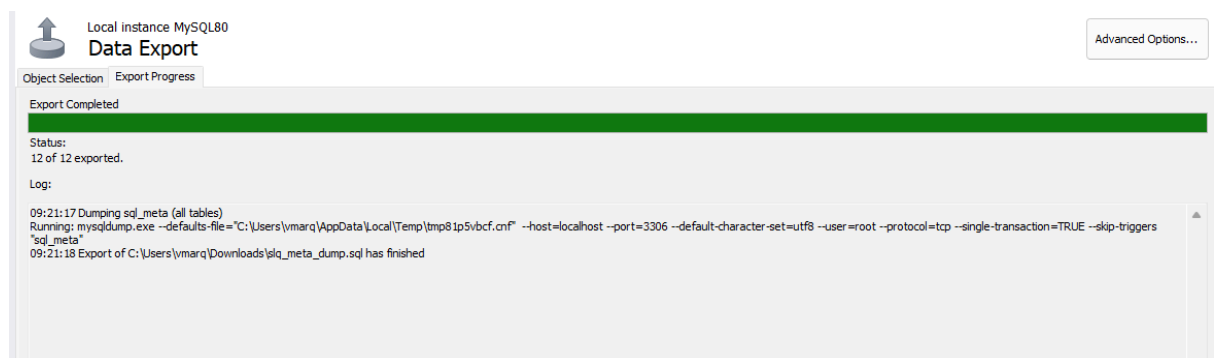
`trg_log_cambio_nombre_planta`: registra en una tabla log cada vez que se modifica el nombre científico de una planta. Llevar un historial de cambios en nomenclatura científica, útil para trazabilidad de datos. Tablas involucradas: `planta` y `log_cambios_nombre_cientifico`. Trigger BEFORE UPDATE

`trg_log_cambio_nombre_polinizador`: igual que el anterior, pero para los nombres científicos de polinizadores. Tablas involucradas: `polinizador` y `log_cambios_nombre_cientifico`. Trigger BEFORE UPDATE

## Backup

1. Para realizar un backup (exportación de datos en MySQL Workbench) voy a *Server* en el panel izquierdo superior y selecciono *Data Export*
2. En *Data Export* selecciono el o los esquema/s a exportar
3. En mi caso decidí realizar un backup del esquema `sql_meta` tanto de la data como de la estructura. (selecciono en el panel derecha inferior: *Data Structure and Data*) y selecciono todas las tablas de la base de datos en el panel de la derecha.

4. Selecciono los objetivo a exportar y marco con un tilde: *Dump Stored Procedures and Functions*, *Dump Events* y *Dump Triggers* (todos los objetos).
5. En Export Options selecciono *Export to Self-Contained File* (archivo comprimido) y en los tres puntitos selecciono la carpeta donde quiero exportar en mi caso *Downloads* y el nombre que le quiero darle al backup (sql\_meta\_dump). Luego selecciono la opción que aparece abajo *Create Dump in a Single Transaction* y *Include Create Schema*
6. Hago la Exportación



## Conclusiones

Utilizando la vista **vista\_fenologia\_mismatch** muestra las fenología de las plantas y polinizadores que están en lo mismos sitio. La vista arroja 48 registros de especies en los mismos sitios, sin embargo esto no significa necesariamente que las especies están interactuando. Es más, si consideramos que el desfase fenológico se produce cuando la floración es anterior a la actividad de los polinizadores o cuando la floración termina antes de que comience la actividad ( $\text{inicio\_floracion} > \text{fin\_actividad}$  o  $\text{fin\_floracion} < \text{inicio\_actividad}$ ). Vemos que 43 de los registros anterior cumple con esa condiciones y solo 5 registros indican una posible interacción ( $\text{inicio\_floracion} \leq \text{fin\_actividad}$  o  $\text{fin\_floracion} \geq \text{inicio\_actividad}$ ).

nombre_sitio	planta	inicio_floracion	pico_floracion	fin_floracion	polinizador	inicio_actividad	pico_actividad	fin_actividad
Bosque del Norte #48	Tabebuia reticulata	2021-03-20	2021-03-30	2021-04-28	Melipona quadrifasciata_44	2021-04-05	2021-04-24	2021-06-01
Estación Costero #19	Bauhinia officinalis	2020-01-15	2020-01-27	2020-03-01	Trigona spinipes_1	2020-02-19	2020-03-04	2020-04-03
Parque del Norte #6	Myrcia pubescens	2022-06-12	2022-06-28	2022-07-28	Cetonia aurata_26	2022-05-02	2022-05-22	2022-07-01
Parque del Norte #6	Bauhinia pubescens	2022-06-11	2022-07-01	2022-08-13	Cetonia aurata_26	2022-05-02	2022-05-22	2022-07-01
Estación Sur #1	Acacia glabra	2020-04-24	2020-05-01	2020-06-27	Bombus terrestris_33	2020-04-23	2020-06-01	2020-08-20

Utilizando la **vista\_interacciones\_taxonomia** puedo ver si hay alguna forma de vida de las plantas que esté más vinculada a un grupo funcional de polinizador através las interacciones. En este caso, encontré que las mariposas están más vinculadas a los arbustos y las abejas solitarias a los árboles. Este último grupo funcional de polinizadores fue a su vez el más representado en las interacciones planta-polinizador.

Con respecto a las plantas y usando la función **duracion\_floracion\_planta** puedo saber cuales son las plantas que tiene floración más larga

	planta	inicio_floracion	fin_floracion	duracion_dias
►	Erythrina reticulata	2022-01-06	2022-03-24	77
	Acacia officinalis	2023-02-09	2023-04-25	75
	Cecropia reticulata	2022-03-23	2022-06-05	74

Algo interesante es que si comparamos con las plantas con mayor número de interacciones, estas no coinciden. Por lo tanto, una floración más larga no necesariamente significa una mayor cantidad de interacciones.