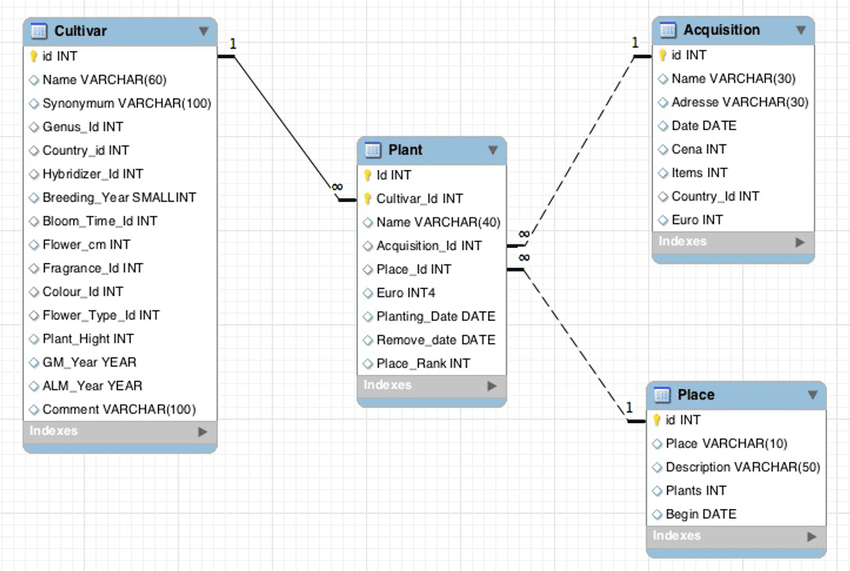
**Configuración de bases de datos locales**

**Configuración de bases de datos en MySQL**

Una vez que instales el servidor de **MySQL** en tu equipo, el siguiente paso es configurar la arquitectura de tu base de datos. Algo muy útil en las bases de datos relacionales es el uso de diagramas entidad-relación que modelan las tablas de la base de datos y la relación entre estas.

A partir de estos diagramas, podemos detectar los tipos de relaciones entre tablas (uno a muchos, muchos a muchos, uno a uno) y por ende el tipo de llaves primarias que almacenará.

Por supuesto, el diseño de bases de datos es todo un proceso que va más allá de los alcances de este curso, sin embargo, con que entiendas las nociones básicas es más que suficiente.



Dentro del diagrama, también debemos analizar, cuantos campos tendrá cada tabla y de qué tipo son.

**Creación de bases de datos**

Una vez definida la arquitectura, debes crear una base de datos dentro del servidor, también llamada esquema. Para crear una base de datos, usamos el comando:

* CREATE DATABASE <nombre>;

**Creación de tablas**

Para crear tablas, debes especificar un nombre para la tabla, para cada campo y el tipo asociado a cada uno de estos.

* CREATE TABLE <nombre> (  
   id\_<nombre> TIPO PRIMARY KEY,  
   campo TIPO,  
   …  
   campo TIPO  
  );

Por convención, el primer campo de una tabla, por lo general, representa a la llave primaria, y recibe el nombre id. Por ejemplo, id\_empleado, id\_usuario, etc. Sin embargo, esto dependerá de tus necesidades.

Puedes tener más de una llave primaria, siempre y cuando la combinación de éstas identifique de forma única a los registros de la tabla.

Si necesitas hacer referencia a un campo dentro de otra tabla, al final de la definición de las tablas, deberás añadir la llave foránea con la siguiente sintaxis:

* FOREIGN KEY (campo\_\_destino) REFERENCES tabla(campo\_origen);

**Inserción de registros**

Para insertar registros en una tabla, se usa el comando:

* INSERT INTO <TABLA> (campo1,campo2,...) VALUES (valor1,valor2,...);

Por ejemplo:

* INSERT INTO empleados (id\_empleado,nombre) VALUES (1,”Juanito”);

Para cargas masivas de datos, emplearemos la herramienta de importación de **MySQL Workbench**, sin embargo, debes tener la estructura de las tablas ya definida. Se aceptan varios formatos para cargar información, el más usado es **CSV** de manera que si tu fuente de datos está en este formato, será sencillo cargarlos.

**Configuración de bases de datos en MongoDB**

Para **MongoDB** configuraremos un cluster mediante la plataforma MongoDB Atlas que permite configurar servidores de bases de datos, con solo unos clics. Por supuesto hay una versión gratuita y una de paga.

Para los fines del curso la versión gratuita, nos vendrá bastante bien. Durante el work trabajaremos en la configuración de este servidor, de momento te dejamos esta [liga](https://www.mongodb.com/cloud/atlas) por si quieres conocer más de esta plataforma.

**Inserción de registros**

En **MongoDB** no es necesario que generemos la estructura de la base de datos pues no existe como tal ese concepto, recordemos que son bases de datos no relacionales. Por lo tanto, necesitamos tener la base de nuestros documentos previamente definida y con valores concretos.  
Para cargar datos en **Compass** podemos usar archivos en formato CSV o JSON debido a la naturaleza de los documentos.

De aquí podemos notar la preferencia del uso de **MongoDB** en ciertos proyectos con una estructura previamente definida, pues la carga de datos consiste en subir un archivo sin necesidad de especificar la estructura primero.

**Carga y exportación masiva de datos**

En general, cuando hablamos de análisis de datos es poco común que se nos proporcione una base de datos con unos cuantos registros, pues esto no garantiza un análisis correcto. Debido a esto es común que se carguen archivos con grandes volúmenes de datos y por lo general el primer reto al que nos enfrentamos es a la limpieza de datos.

De esta forma, es importante que antes de cargar un archivo CSV o JSON a tu base de datos, verifiques que su estructura sea correcta, que tenga los campos que necesitas, que la codificación sea la adecuada entre otras.

Aquí te dejamos una referencia para que limpies datos desde consola. Pregunta al experto otras alternativas en caso de dudas.

* https://www.datacamp.com/community/tutorials/shell-commands-data-scientist

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_