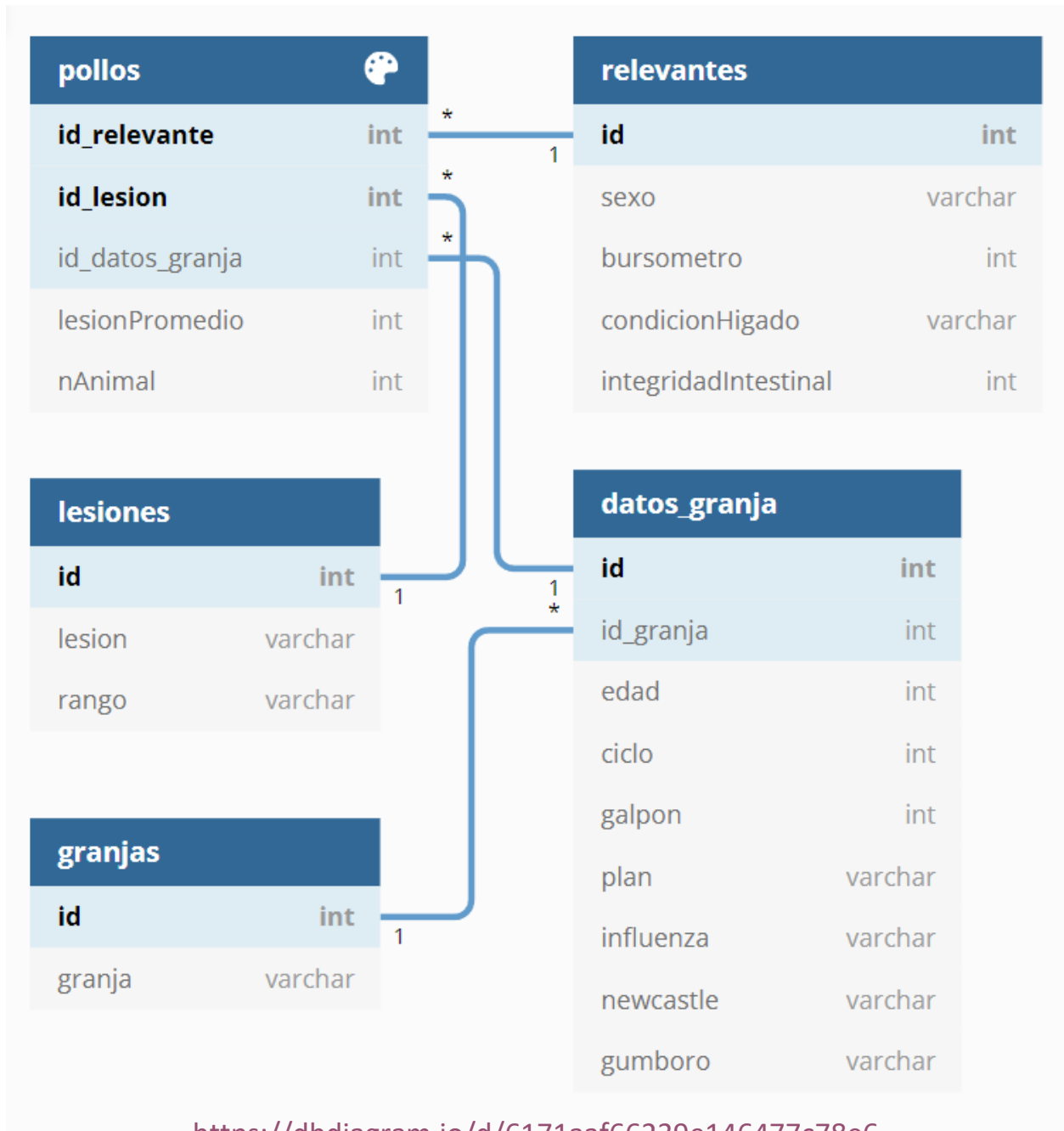


DISEÑO DE BASE DE DATOS:



<https://dbdiagram.io/d/6171aaf66239e146477c78e6>

Originalmente se tenía una sola tabla de 17 columnas. Ahora se crean 5 tablas que distribuyen las mismas 17 columnas. Aunque haría una puesta en común con el equipo o con el cliente para resolver las siguientes preguntas:

- ¿Es necesario guardar la columna 'nAnimal'?
- ¿Es necesario guardar los registros que en 'lesionPromedio' se tenga 0? Recordando que en la tabla es cero, pero en el archivo original era un dato vacío.
- Teniendo en cuenta que 'newcastleVacuna' siempre almacena el mismo dato ¿Es necesario implementar esta columna?

Asumiendo que se deben mantener todas las columnas, volvamos al modelo generado. Al analizar la tabla resultante del proceso de ETL, se identifica claramente que a la hora de ingresar los datos de un solo pollo se generan 33 registros, uno por cada lesión, y de estos registros, 14 de sus columnas son exactamente iguales para cada uno, además que la columna 'lesionRango' no es necesaria tenerla en cuenta si se sabe que el rango de las lesiones no debe cambiar, por ejemplo, el rango con el que se mida 'piernasPalidas' es igual para todos los pollos.

Como se acaba de explicar, el rango de una lesión, debería ser siempre la misma, no importa el pollo al que se le realice la prueba, así que guardar este dato iterativamente no aporta mucho, lo que se debe conocer es la lesión para poder identificar su rango. Para eso, se crea una tabla 'lesiones', que guarda tanto el nombre de la lesión como el rango de esta. Luego la llave primaria de esta tabla será una llave foránea en la tabla 'pollos', así nos evitaremos estar guardando las dos columnas para cada pollo, solo se almacena esta llave foránea que representa los dos.

Como se dice en el enunciado, el sexo, bursometro, condicionHigado e integridadIntestinal son atributos especiales para realizar seguimiento, por eso se les asigna su propia tabla 'relevantes'. La llave primaria id de esta tabla se usará como llave foránea en la tabla 'pollos'. También pasa lo mismo que con la tabla 'lesiones', hacer esto evita que la tabla 'pollos' en diferentes registros se guarden las mismas 4 columnas y en este caso solo se guardaría una que representa las 4 a la vez.

En el archivo de Excel original, la metadata de cada 5 pollos estaba dispuesta en una tabla en la parte superior del archivo. Se crea la tabla 'datos_granja' en el modelo para que guarde específicamente esta información. También la llave primaria de esta tabla será una llave foránea de la tabla 'pollos', esto con el fin de las tablas anteriores, evitar tener muchos registros con las mismas X columnas. Aquí también se debe tener en cuenta que no se repite la misma información para un solo pollo si no para 5.

Aunque no es necesario del todo, se crea una tabla que guarde específicamente el nombre de las granjas, solo con el fin de facilitar los procesos de analítica a la hora de filtrar datos por granja, es decir, es más fácil tener una tabla con los nombres de las granjas para saber cuáles se pueden filtrar que primero hacer una búsqueda(query) para saber este dato. La llave primaria de esta tabla también es llave foránea en la tabla 'datos_granja'.

Por último, la tabla 'pollos' es la que guarda 'lesionPromedio' y 'nAnimal' de cada pollo. Cuenta con otros atributos, que son las llaves foráneas previamente mencionadas. Esta es la tabla que representa la tabla resultante del ETL, en lugar de tener 17 columnas ya se cuenta con solo 5, pero 3 de estas son llaves foráneas que permiten acceder a las 15 columnas que no se encuentran directamente en 'pollos'. La ventaja de este sistema es que, si tomamos por ejemplos los atributos de la tabla 'relevantes' y los colocáramos directamente en la tabla 'pollos', siempre tendríamos 33 registros con 4 columnas iguales, pero con el nuevo sistema, se agrega un registro con los 4 atributos en la tabla 'relevantes' y solo se crea una columna en 'pollos' con 33 registros repetidos. Se emplea como llave primaria compuesta id_lesion y id_relevante, esto debido a que la combinación de estas dos llaves me permiten reconocer la lesión específica de cualquier pollo, que es al fin y al cabo el objetivo de la tabla 'pollos'.