Informe

Tarea obligatoria

Bases de datos 1

Joel Alayon

Leandro Alfonso

Introducción

En el presente informe se describe el desarrollo y resultados de nuestro proyecto enfocado en la creación de una base de datos, inserción de datos y consultas sobre los mismos utilizando un lenguaje de programación. La consigna plantea que fuimos seleccionados para trabajar en la formula 1 y se requiere crear un programa el cual sea capaz de crear una base de datos basada en un conjunto de archivos CSV extradidos de la plataforma ***KAGGLE*** e insertar todos los datos de los mismos en la base para poder realizar consultas sobre ellos.

Dicha base de datos trata sobre un campeonato mundial de fórmula 1 y consta con toda la información sobre las carreras, los pilotos, los constructores, la clasificación, los circuitos, los tiempos de vuelta, las paradas en boxes y los campeonatos desde 1950 hasta la última temporada de 2023. Para llevar a cabo este proyecto, seleccionamos cuidadosamente una tecnología específica, Java, que nos permitió interactuar de manera eficiente con sistemas de gestión de bases de datos relacionales.

El proceso principal por el que pasamos previo a comenzar el proyecto fue diseñar la base de datos correspondiente al conjunto de archivos que se nos brindó. Para ello tuvimos en cuenta que cada uno de los archivos hacía alusión a una tabla y rigurosamente nos fijamos como se relacionaba con el resto. Para tener una forma mucho más grafica de esto diseñamos un MER (Modelo entidad-relación), el cual nos ayudó bastante para entender las relaciones que tenían nuestras tablas entre sí.

Definido esto, creamos un documento de texto el cual contenía las instrucciones SQL para poder crear dichas tablas con sus respectivas claves primaras, claves foráneas y atributos. A continuación, desarrollamos un programa que nos permitió conectarnos a la base de datos, leer este documento, procesarlo para separar las instrucciones SQL y crear la respectiva base de datos desde el programa.

Una vez se encontraba lista la estructura de la base, el programa permitió leer los archivos CSV los cuales contenían cantidades de filas (cada fila representa un dato de esa tabla), procesar línea por línea dicho archivo y armar una instrucción SQL para poder insertar cada fila de datos en su respectiva tabla. Una vez que los datos se encontraban en la base, efectuamos diversas consultas para obtener la información que nos pedía la consigna.

A lo largo de este informe, se detallan las etapas clave de nuestro proyecto, incluyendo el diseño de la base de datos, la implementación del programa y el proceso de carga de datos.

Tecnología utilizada

Al momento de elegir la tecnología, ambos estuvimos de acuerdo en utilizar Java por diversas razones. La más importante es que Java contiene una variedad de herramientas y bibliotecas para interactuar con las bases de datos, por lo que fue una buena opción. Además de esto, Java es un lenguaje bastante popular y es bastante utilizado en el desarrollo de software y el cual venimos utilizando a lo largo de nuestra carrera estudiantil, es decir que estamos bastantes familiarizados con su sintaxis y características. Por lo tanto, nos permitió desarrollar el programa de forma más ágil y efectiva que con otros tipos de lenguajes.

Agregar también que java cuenta con una comunidad bastante activa y una amplia documentación, por lo que nos simplifico bastantes bloqueos que pudimos tener proceso.

Diseño de la base de datos

Para crear la estructura de la base de datos, el primer paso fue leer los archivos CSV proporcionados y analizamos el contenido de cada uno de ellos. Revisando los datos pudimos identificar los diferentes tipos de entidades y las relaciones entre ellas, cada una de ellas tiene sus respectivos atributos. Como herramienta visual para representar utilizamos un Modelo-Entidad-Relación (MER.) Una vez terminado esto y que pudimos tener una visión más clara, seguimos con diseñar las tablas de la base de datos.

Cada tabla representaba una entidad especifica y los atributos asociados a ella. Consideramos cuidadosamente los tipos de datos adecuados para cada atributo y aplicamos restricciones para algunos de estos atributos como por ejemplo los campos que hacían referencia a otras tablas para que no pudieran estar vacíos.

Posteriormente, creamos un conjunto de instrucciones SQL para ser capaces de crear las tablas en la base de datos. Estas instrucciones incluían la definición de cada tabla, sus columnas, restricciones, claves primarias y claves foráneas para representar las relaciones. En un principio pensamos en crear las tablas en un orden específico para que de esta manera no haya inconvenientes a la hora de crear una tabla que requería una relación con una inexistente. Pero luego nos dimos cuenta de que podríamos crear todas las tablas para luego agregar las claves foráneas, de esta manera no habría ningún inconveniente en el orden de creación de las tablas.

Esto último estuvo muy bueno y nos simplifico bastante a la hora de implementar el programa.

Implementación del programa

Al momento de cargar los datos desde los archivos CSV a la base de datos, desarrollamos un programa en Java que realizaba esta tarea de manera automatizada y eficiente. Además, este programa se encargaba de crear la base de datos utilizando las consultas que mencionamos en el bloque anterior. Es decir, el programa primero establecía la comunicación con la base y procesaba los archivos CSV realizando por cada fila un input en la tabla correspondiente de la base. Luego de esto ejecutaba ciertas consultas que la consigna requería.

En primer lugar, utilizamos la librería SQL de Java para poder establecer la conexión con la base y luego de esto se procesaron cada uno de los archivos necesarios para la creación de la base (los cuales están divididos en tablas y alters para no afectar al orden de la creación de las mismas) utilizando un método que se encarga de leer estos archivos y armar una instrucción para enviar a la base.

Completo esto, llamamos a un método que tiene como finalidad cargar la información de los csv en la base. Para cumplir con esto se encarga de leer uno por uno en una lista de archivos, para los cuales procesa línea por linear armando una instrucción de inserción la cual será enviada a la base. Esto lo logra ya que contamos con un helper que contiene metodos que se encargan de armar estos strings de instrucciones para cada tipo de tabla que tenemos, es decir, existe un método por cada tabla que tenemos. Esto se diseñó de esta manera ya que hacer uno solo y genérico parecía mucho más complicado y podría traer problemas al insertar datos que no queremos en columnas o tablas que no corresponden.

Finalmente realizamos algunas consultas ya que la consigna lo requería y para validar que estos datos y relaciones entre las tablas son correctas

Conclusiones

En conclusión, este proyecto nos sirvió como una experiencia valiosa en el ámbito de las bases de datos que nos permitió adquirir conocimientos sobre su funcionamiento e importancia. Aunque reconocemos que podríamos haber logrado un mejor resultado y no pudimos entregar en tiempo y forma, valoramos el aprendizaje obtenido a lo largo de esta tarea.

Durante el desarrollo tuvimos la oportunidad de aprender sobre el mundo de SQL y aprender nuevas cosas sobre el manejo de las bases de datos. Nos fue muy útil diseñar una estructura de base adecuada, sentimos que aprendimos bastante diseñando el MER y pensando cada detalle de la base en cuestión. También fue muy interesante procesar datos de archivos CSV y luego realizar consultas sobre lo procesado.

También aprendimos bastante sobre pensar en una solución a un problema planteado. A medida que fuimos encontrando desafíos, tomamos decisiones durante el proceso de diseño. Poner en cuestión el lenguaje a utilizar, la forma de procesar los archivos, cómo preparar las inserciones y luego como realizar las consultas fue una experiencia bastante enriquecedora.

Si bien no pudimos cumplir con los plazos establecidos, consideramos que esta tarea fue una valiosa oportunidad de aprendizaje. Nos permitió comprender el funcionamiento y la importancia de las bases de datos en el ámbito de la gestión de información. Además, nos brindó la oportunidad de aplicar los conceptos teóricos que aprendimos a lo largo del curso y ganar experiencia real en el desarrollo de bases de datos.