

Práctica 7: Población de Bases de Datos.

Integrantes: Arcos Morales Ramón : 319541478
 Casarrubias Casarrubias Victor Manuel : 421003581
 Castillo Hernández Eduardo : 420003557
 López Asano Miguel Akira : 320219089
 Rivera Lara Sandra Valeria : 320039823

Fecha: 21/10/2024

Ejercicio 1. Uso de Mockaroo.

La herramienta que se usó para la generación de datos fue Mockaroo, la cuál es una herramienta para generar una gran cantidad de datos randomizados a escala. Para utilizar dicha herramienta se hizo uso de los schemas que ofrece la pagina; es decir, una clase de template que en la que se pueden poner atributos y los datos que se quiere que generen. Por ejemplo, para elegir generar ids se usó un conjunto de datos y configuraciones definidos por Mockaroo como row numbers, para que se le asignara una serie de ids únicos a cada tupla. Para poder conectar los diferentes datos y schemas, se hizo uso de los datasets, la cuál es una forma de almacenar los datos generados de un schema por Mockaroo y guardarlos localmente para ser utilizados en otros schemas, como por ejemplo, para poder conectar los atributos que hacen referencia a tuplas de otras tablas, las cuales son llaves foráneas. Por último, para poder generar el código de inserción en SQL se hizo uso de los schemas de igual manera, ya que existe la opción de cambiar el formato de guardado local en SQL, que cambia automáticamente el código generado por Mockaroo de datos en CSV a SQL, luego con esos archivos, solo se guardó en el archivo DML para su uso en nuestra base de datos.

País.

Se generaron 3000 países y se aplicó *shell scripting* para ordenar y eliminar duplicados.

Resultado final, 160 países únicos.

El atributo nombrePaís fue generado con el tipo *Country* de Mockaroo.

Relación con llave primaria.

Atleta.

Se generaron 1000 atletas.

Atletas representativos de Estados Unidos, Reino Unido, Australia y Canadá.

El atributo numeroPasaporte fue generado con una secuencia de 10 caracteres de dígitos aleatorios y letras en mayúsculas.

El atributo nombrePaís fue generado con la restricción de los cuatro países antes mencionados.

El atributo fechaNacimiento fue considerado a partir del 1 de enero de 1980 hasta el 31 de diciembre del 2009.

El atributo nacionalidad fue generado con una distribución de peso con los siguientes valores: Estados Unidos 200, Canadá 50, Australia 40, Reino Unido 30, todos los demás países 1.

El atributo nombre fue generado con el tipo de dato *First Name*.

El atributo primerApellido fue generado con el tipo de dato *Last Name*.

El atributo segundoApellido fue generado con el tipo de dato *Last Name*.

El atributo género fue generado con el tipo de dato *Gender(abbrev)*.

Relación con llave primaria y foránea.

TelefonoAtleta.

Se generaron 1000 teléfonos correspondientes uno a uno con atletas.

El atributo teléfono fue generado con el tipo de dato *Phone*.

Relación con llave compuesta.

CorreoAtleta

Se generaron 1000 correos correspondientes uno a uno con atleta.

Relación con llave compuesta.

Se crearon los campos no visibles `--nombre`, `--usuario` y `--dominio`: el primero recupera el nombre del atleta, el segundo se genera con el tipo de dato *Username* y el último se genera con una lista personalizada con los valores *gmail*, *outlook*, *hotmail*, *yahoo*, *icloud* y una distribución de peso.

Con los datos anteriores se genera el atributo correo con la siguiente formula en *Ruby*

```
1 lower("#{__nombre}.__#{__usuario}@#{__dominio}.com")
```

Disciplina

Se generó el CSV a mano de las 51 disciplinas oficiales según la página oficial de los juegos olímpicos de *Los Angeles 2028* [IOC, 2024].

Se generaron las 99 disciplinas restantes de la siguiente manera:

El atributo nombreDisciplina con el tipo de dato *Buzzword*

El atributo categoria con una expresión regular que escoge entre Individual o Equipo.

El CSV fue después convertido a SQL con una herramienta externa.

Relación con llave primaria.

Entrenador

Se generaron 1000 entrenadores.

La construcción fue similar a los datos de la relación atleta, siendo la única diferencia que un entrenador puede haber nacido desde el 1 de enero de 1950.

Relación con llave primaria y foránea.

Patrocinador

Se generaron 520 patrocinadores y se aplicó *shell scripting* para ordenar y eliminar repetidos.

Resultado final, 500 patrocinadores únicos.

El atributo nombrePatrocinador fue generado con el tipo de dato *Fake Company Name*.

Relación con llave primaria.

Localidad

Se generó el CSV a mano de 20 localidades oficiales.

Se generaron las 130 localidades restantes de la siguiente manera:

El atributo idLocalidad con un *Row Number*.

El atributo nombrePaís con un *Dataset Column* a la relación pais.

El atributo calle con un *Street Name*.

El atributo número con *Street Number*.

El atributo ciudad con *City*.

Los atributos nombre y tipo con *Buzzword*.

El atributo aforo con *Number*, teniendo un mínimo de 100 y un máximo de 10000.

El CSV fue después convertido a SQL con una herramienta externa.

Relación con llave primaria y llave foránea.

Evento

Se generaron 500 eventos en el período oficial de los juegos olímpicos de *Los Angeles 2028*.

Se generaron 100 eventos en el período que comprende el 1 de enero de 2024 hasta el 20 de octubre de 2024.

El atributo idEvento fue generado con el tipo de dato *Row Number*.

Los atributos idLocalidad y nombreDisciplina fueron generados con el tipo de dato *Dataset Column* de manera secuencial.

El atributo fecha fue generado con el tipo de dato *Datetime* con los períodos antes mencionados.

El atributo horaInicio fue generado con el tipo de dato *Time* en formato 24 horas.

El atributo duracionMaxima fue generado con el tipo de dato *Number* con un mínimo de 1 y un máximo de 5.

Relación con llave primaria y llave foránea.

Referencias

[IOC, 2024] IOC (2024). Olympics sports list.