

GAME SPEEDRUN RECORDS - GRUPO 40

Hito $N^{Q}3$

Grupo 40: Leonardo Rikhardsson

Alejandro Mori Ignacio Humire Adolfo Arenas

Profesor: Matías Toro I.

Fecha de entrega: 6 de Noviembre de 2023

Santiago, Chile

Resumen Hito 1

1. Resumen Hito 0

En el hito anterior se estableció la fuente de datos Game Speedrun Records como base de datos para desarrollar el proyecto del curso. Con respecto a esta base de datos, se identificaron los distintos atributos que posee y se describieron los mismos. Adicionalmente, se escogió como problema el desarrollar una aplicación web para visualizar y analizar las estadísticas speedruns en distintos videojuegos, para ayudar al usuario a escoger un videojuego que speedrunnear en base a sus gustos, tiempo máximo que está dispuesto a dedicar, género de interés, etc.

2. Resumen Hito 1

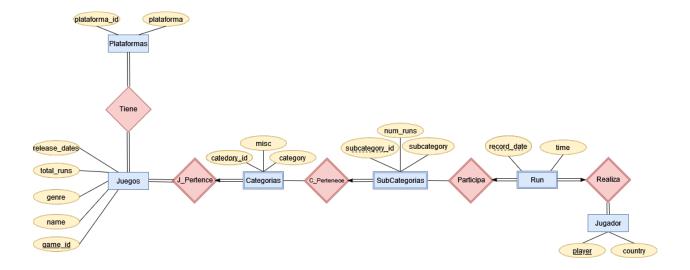


Figura 1: Modelo Entidad Relación

Por cómo está construida la base de datos, las categorías no pueden ser distinguidas unas de otras de forma única mediante sus id's, por ejemplo, la categoría «100 %» se encuentra en la mayoría de los juegos presentes, pero no habría como diferenciar el «100 %» de un juego A con el «100 %» de un juego B solamente con la id de la categoría. Para solucionar esto último, se ha optado por definir a las categorías como entidades débiles, usando las id's de estas (llaves parciales) en conjunto a las id's de los juegos (llaves primarias) para poder diferenciarlas entre ellas. El mismo argumento es válido para justificar a las subcategorías como entidades débiles. Por ejemplo, una Run se puede comparar con otra solamente si y solamente si la otra Run pertenece a la misma tupla (game_id, category_id, subcategory_id). Por otra parte, la idea de relacionar Jugador con Juegos no la consideramos necesaria para nuestro objetivo ya que nuestro foco central está en las Runs. Sin embargo para facilitar la búsqueda de información decidimos hacer para este hito que plataforma sea una entidad y relacionarla con juegos.

Un jugador puede realizar una o más runs para un solo juego u otros. Para diferenciar estas runs de otras en distintos juegos, se construirán como entidades débiles que dependerán de las

Resumen Hito 1 2

id's de una subcategoría, que por transitividad dependerá finalmente de un juego. También, para identificar las runs de otros jugadores, se tomarán los nombres de los jugadores (nicknames).

2.1. Modelo Relacional

En base al modelo Entidad Relación presentado previamente, construimos el modelo relacional como sigue:

```
Juegos(game_id: Int, name: String, genre: String, total_runs: Int, release_date: Date)
Categorias (category id: Int, ju game id: Int, category: String, misc: Bool)
   (ju game id) REF Juegos(game id)
SubCategorias(subcategory_id: Int, c_category_id: Int, c_ju_game_id: Int, subcategory: String,
num runs: Int)
   (c_category_id, c_ju_game_id) REF Categorias(category_id, ju_game_id)
Run(record_date: Date, s_subategory_id: Int, s_c_category_id: Int, s_c_ju_game_id: Int,
j player: String, time: Double precision)
  (s_subategory_id, s_c_category_id, s_c_ju_game_id, j_player) REF
  SubCategorias(subcategory id, c category id, c ju game id),
  j_player REF Jugador(player)
Jugador(player: String, country: String)
Plataformas(plataforma_id: Int, plataforma: String)
Juegos_plataformas(ju_game_id: Int, p_plataforma_id: Int)
(ju_game_id) REF Juegos(game_id),
(p_plataforma_id) REF Plataformas(plataforma_id)
```

3. Hito 2

3.1. Implementación relacional

A continuación se muestran las Queries realizadas para crear las tablas de las entidades en la base de datos. Notar que no fue necesario crear tablas para las relaciones puesto que al realizar las tablas para las entidades débiles estas ya relacionan llaves primarias entre tablas.

```
create table juegos (
      game_id bigint,
      name varchar(255) not null,
      genre varchar(255) not null,
      total_runs bigint not null,
      release_date date not null,
      primary key (game_id)
  );
8
  create table categorias (
      category_id bigint not null,
      ju_game_id bigint not null,
12
13
      category VARCHAR(255) not null,
      misc boolean not null,
14
      primary key (category_id, ju_game_id),
      foreign key (ju_game_id) references juegos(game_id)
16
  );
17
  create table subcategorias (
19
      subcategory_id bigint not null,
20
      c_category_id bigint not null,
21
      c_ju_game_id bigint not null,
      subcategory VARCHAR(255) not null,
23
      num_runs bigint,
      primary key (subcategory_id, c_category_id, c_ju_game_id),
      foreign key (c_category_id , c_ju_game_id) references categorias(
     category_id , ju_game_id)
  );
27
2.8
  create table run(
      record_date timestamp,
30
      s_subategory_id bigint not null,
31
      s_c_category_id bigint not null,
      s_c_ju_game_id bigint not null,
33
      j_player varchar(255),
      time double precision,
35
      primary key (record_date, s_subategory_id, s_c_category_id,
     s_c_ju_game_id, j_player),
      foreign key (s_subategory_id, s_c_category_id, s_c_ju_game_id)
     references subcategorias(subcategory_id, c_category_id, c_ju_game_id),
      foreign key (j_player) references jugador(player)
39 );
```

```
41 create table jugador (
      player VARCHAR(255) not null,
     country VARCHAR(255) not null,
      primary key (player)
47 create table plataformas (
      plataforma_id serial,
      plataforma varchar(255) not null,
49
      primary key (plataforma_id)
51);
 create table juegos_plataformas(
53
     ju_game_id bigint,
     p_plataforma_id bigint ,
      primary key (ju_game_id, p_plataforma_id),
      foreign key (ju_game_id) references juegos(game_id),
      foreign key (p_plataforma_id) references plataformas(plataforma_id)
59);
```

Código 1: Queries usadas para crear las tablas en la DB

Hito 2 5

3.2. Carga de datos

3.2.1. Código

A continuación se muestra el código que se escribió para cargar los datos correctamente a la base de datos. Cabe destacar que nuestro caso el .csv que escogimos tenía pocos errores por lo que fue innecesario hacer parse a los datos. Además, no separamos los jugadores cuando las runs son grupales pues no lo consideramos relevante para la creación de la DB.

```
1 from binascii import a2b_base64
2 import psycopg2
3 import psycopg2.extras
4 import csv
5 import re
  conn = psycopg2.connect(
      host="cc3201.dcc.uchile.cl",
      database="cc3201",
9
      user="cc3201",
10
      password="a",
      port="5540"
12
13
15 cur = conn.cursor()
  def findOrInsert(name):
17
      cur.execute("select plataforma_id from plataformas where plataforma=%s
      limit 1", [name])
      r = cur.fetchone()
19
      if(r):
20
          return r[0]
      else:
22
          cur.execute("insert into plataformas (plataforma) values (%s)
23
     returning plataforma_id", [name])
          return cur.fetchone()[0]
24
25
  with open('speedrun.csv', encoding="utf8") as csvfile:
26
      reader = csv.reader(csvfile, delimiter=',', quotechar='"')
27
      i = 0
28
      for row in reader:
29
          i +=1
30
          if i ==1:
               continue
32
          #tablas: juegos, categorias, juego_categoria, subcategorias,
34
     categorias_subcategorias, run, subcategory-run, jugador, run-jugador
35
          #columnas: Game_Id, Category_Id, Subcategory_Id, Name, Genre, Platforms
      , Total_Runs, Release_Date, Misc, Category, Num_Runs, Subcategory, Time_0,
     Country_0, Players_0, Record_Date_0
37
          #datos para insertar en "juegos"
```

```
game_id = row[0]
          name = row[3]
          genre = row[4]
          total runs = row[6]
          release = row[7]
44
          #buscamos si ya está el juego en la tabla
45
          cur.execute("select game_id from juegos where game_id=%s limit 1",
46
      [game_id])
          r = cur.fetchone()
47
          if(not r): #si no existe
48
              cur.execute("insert into juegos (game_id, name, genre,
49
     total_runs, release_date) values (%s, %s, %s, %s, %s) ", [game_id, name,
      genre, total_runs, release])
50
          #separamos las plataformas
          platforms = [m.strip() for m in row[5].split(',')]
53
          #buscamos o creamos sus id's
          for p in platforms:
              findOrInsert(p)
57
58
59
          #datos para crear un "jugador"
60
          player = row[14]
61
          country = row[13]
62
          position = 1
          #buscamos si ya está el player en la tabla
64
          cur.execute("select player from jugador where player=%s limit 1",
     [player])
          r = cur.fetchone()
          if(not r):
67
              cur.execute("insert into jugador (player, country) values (%s,
      %s)", [player, country])
      conn.commit()
70 conn.close()
72 conn = psycopg2.connect(
      host="cc3201.dcc.uchile.cl",
      database="cc3201",
74
      user="cc3201",
75
      password="a",
      port="5540"
77
80 cur = conn.cursor()
with open('speedrun.csv', encoding="utf8") as csvfile:
      reader = csv.reader(csvfile, delimiter=',', quotechar='"')
83
      for row in reader:
          i+=1
```

```
if i == 1:
86
               continue
88
           #empezamos llenando la tabla de "juegos plataformas"
89
           game_id = row[0] #volvemos a sacar el game_id
90
91
           #separamos las plataformas
92
           platforms = [m.strip() for m in row[5].split(',')]
93
           #buscamos o creamos sus id's
95
           platforms_ids = []
96
           for p in platforms:
97
               a = findOrInsert(p)
               platforms_ids.append(a)
99
100
               #buscamos si ya está el par en la tabla
               cur.execute("select (ju_game_id, p_plataforma_id) from
      juegos_plataformas where (ju_game_id, p_plataforma_id)=(%s, %s) limit 1
      ", [game_id, a])
               r = cur.fetchone()
104
               if(not r):
                    cur.execute("insert into juegos_plataformas (ju_game_id,
      p_plataforma_id) values (%s, %s)", [game_id, a])
106
           #datos para insertar en "categorias"
107
           category_id = row[1]
108
           category = row[9]
109
           misc = row[8]
           player = row[14] #volvemos a sacar al player
           #buscamos si ya está la categoría en la tabla
113
           cur.execute("select (category_id, ju_game_id) from categorias
114
      where (category_id, ju_game_id)=(%s, %s) limit 1", [category_id,
      game_id])
           r = cur.fetchone()
           if(not r):
               cur.execute("insert into categorias (category_id, ju_game_id,
117
      category, misc) values (%s, %s, %s, %s)", [category_id, game_id,
      category, misc])
118
           #datos para insertar en "subcategorias"
119
           sub_id = row[2]
120
           sub_category = row[11]
           num_runs = row[10]
123
           #buscamos si ya está la subcategoria en la tabla
124
           cur.execute("select (subcategory_id, c_category_id, c_ju_game_id)
      from subcategorias where (subcategory_id , c_category_id , c_ju_game_id)
      =(%s, %s, %s) limit 1", [sub_id, category_id, game_id])
           r = cur.fetchone()
126
           if(not r):
```

```
cur.execute("insert into subcategorias (subcategory_id,
128
      c_category_id, c_ju_game_id, subcategory, num_runs) values (%s, %s, %s,
       %s, %s)", [sub_id, category_id, game_id, sub_category, num_runs])
129
           #datos para crear "run"
130
           runtime = row[12]
           recordate = row[15]
132
133
           #buscamos si ya está la run en la tabla y checkeando si está vacío
           if(recordate != '' and recordate != None):
136
                cur.execute("select (record_date, s_subategory_id,
      {\tt s\_c\_category\_id}\;,\;\; {\tt s\_c\_ju\_game\_id})\;\; from\;\; run\;\; where\;\; ({\tt record\_date}\;,
      s_subategory_id, s_c_category_id, s_c_ju_game_id)=(%s, %s, %s, %s)
      limit 1", [recordate, sub_id, category_id, game_id])
               r = cur.fetchone()
137
                if(not r):
138
                    cur.execute("insert into run (record_date, s_subategory_id
       s_c_category_id, s_c_ju_game_id, j_player, time) values (%s, %s, %s,
      %s, %s, %s)", [recordate, sub_id, category_id, game_id, player, runtime
      ])
       conn.commit()
141 conn.close()
```

Código 2: Código de Python para cargar datos en las tablas de la DB

3.2.2. Evidencia DB

game_id	name	genre	total_runs release_date
0	Super Mario 64	3D Platformer	9174 1996-06-23
1	Minecraft: Java Edition	Sandbox	6869 2011-11-18
2	Super Mario Odyssey	3D Platformer	4284 2017-10-27
3	Celeste	2D Platformer	3815 2018-01-25
4	Mario Kart 8 Deluxe	Racing	7660 2017-04-28
5	Seterra	Misc	7763 1997-01-01
6	Portal	Puzzle	3676 2007-10-09
7	Super Metroid	Action	2826 1994-03-19
8	Getting Over It With Bennett Foddy	2D Platformer	2561 2017-10-06
9	Super Mario Sunshine"Platformer	3D Platformer	2575 2002-07-19
10	ROBLOX: Speed Run 4	Parkour	5086 2013-01-01
11	Super Mario World	2D Platformer	3150 1990-11-21
12	Super Mario Bros.	2D Platformer	2277 1985-09-13
13	The Legend of Zelda: Ocarina of Time	Action-Adventure	3065 1998-11-21
14	The Legend of Zelda: A Link to the Past	Action-Adventure	2891 1991-11-21
15	Minecraft: Bedrock Edition	Sandbox	2147 2011-08-16
16	SpongeBob SquarePants: Battle for Bikini Bottom	3D Platformer	1084 2003-10-31
17	Resident Evil 2 (2019)	Horror	2880 2019-01-25
18	Pou	Misc	3930 2013-02-14
19	ROBLOX: Piggy	Puzzle	4081 2020-01-23

Figura 2: Tabla de juegos



Figura 3: Tabla de categorías

subcategory_id	c_category_id	c_ju_game_id	subcategory	num_runs
0	0	0	N64	456
1	0	0	VC	135
2	0	0	EMU	302
0	1	0	N64	928
1	1	0	VC	291
2	1	0	EMU	907
0	2	0	N64	1511
1	2	0	VC	588
2	2	0	EMU	3318
0	3	0	N64	299
1	3	0	VC	55
2	3	0	EMU	222
0	4	0	N64	105
1	4	0	VC	10
2	4	0	EMU	47
0	0	1	Set Seed,Pre 1.9	218
1	0	1	Set Seed,1.9-1.15	149
2	0	1	Set Seed,1.16+	989
3	0	1	Random Seed, Pre 1.9	299
4	0	1	Random Seed,1.9-1.15	648
5	0	1	Random Seed,1.16+	3174
0	1	1	Set Seed, Pre 1.9	31
6	1	1	Set Seed,1.9-1.13	16
7 3	1 1] 1 1	Set Seed,1.14+	11 31
8	1	1 1	Random Seed,Pre 1.9 Random Seed,1.9-1.13	9
9	1	1	Random Seed,1.14+	10
10	2	1	1.0-1.6,SS	6
11	2	1	1.0-1.6,SSG	4
12	2	1	1.0-1.6,RS	3
13	2	1	1.0-1.6,RSG	5
14	2	1	1.8-1.11,55	11
15	2	1	1.8-1.11,SSG	3
16	2	1	1.8-1.11,RS	3
17	2	1		3
18	3	1	1.12,55	1
19	3	_	1.12,SSG	1
20	3		1.12,RS	ō
21	3	ī	1.12,RSG	4
22	3		1.13,SS	ø
23	3	i - 1	1.13,SSG	o o
24	3	. ī	1.13,RS	ø

Figura 4: Tabla de subcategorías



Figura 5: Tabla de run

Figura 6: Tabla de jugador

```
plataforma_id | plataforma

1 | Nintendo 64
2 | Wii Virtual Console
3 | Wii U Virtual Console
4 | Switch
5 | PC
6 | PlayStation 4
7 | Xbox One
8 | Xbox One X
9 | Xbox One S
10 | Google Stadia
11 | PlayStation 5
12 | Xbox Series S
14 | Android
15 | iOS
16 | Web
17 | PlayStation 3
18 | Xbox 360
19 | Super Nintendo
20 | New Nintendo 3DS Virtual Console
21 | SNES Classic Mini
22 | MiSTer
23 | Analogue Super Nt
24 | GameCube
25 | Wii
26 | Macintosh
27 | Nintendo 3DS Virtual Console
28 | Nintendo 3DS Virtual Console
29 | Game Boy Advance
30 | Famicom Disk System
```

Figura 7: Tabla de plataformas

Hito 2 11

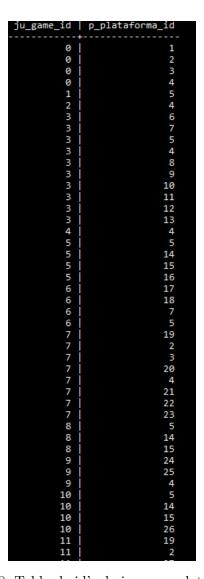


Figura 8: Tabla de id's de juegos y plataformas

4. Hito 3

A continuación se muestra el trabajo hecho para el hito tres, lo que incluye las consultas pedidas y la creación de una página web usando Apache2 más PHP.

4.1. Consultas

1. Todos los juegos, cuya categoria sea "input", tal que su runner sea del pais "input".

```
SELECT name FROM speed.juegos WHERE game_id IN

(SELECT s_c_ju_game_id FROM speed.run

WHERE (s_c_ju_game_id, s_c_category_id) IN

(SELECT ju_game_id, category_id FROM speed.categorias

WHERE category LIKE : valor1)

AND j_player IN

(SELECT player FROM speed.jugador

WHERE country = : valor2))

;
```

En :valor1 debe ir un string como input, un ejemplo de categoría es 'Any'. Por otro lado, para el :valor2 también deben ir strings como input, en este caso al ser país, un ejemplo sería: 'United States'. Es posible dejar el :valor1 vacío para elegir cualquier categoría.

2. Todos los juegos que sean de determinada plataforma y tengan sub-1 minuto

```
SELECT name FROM speed.juegos

WHERE game_id IN

((SELECT s_c_ju_game_id FROM speed.run

WHERE time < 60)

INTERSECT

(SELECT ju_game_id FROM speed.juegos_plataformas

WHERE p_plataforma_id IN

(SELECT plataforma_id FROM speed.plataformas

WHERE plataforma_id FROM speed.plataformas

WHERE plataforma_id FROM speed.plataformas

WHERE plataforma_LIKE : valor1)))
```

En :valor1 deben ir strings como input, un ejemplo de plataforma sería 'PlayStation 5'. Se puede dejar este campo en blanco para elegir un juego de cualquier plataforma.

3. Todos los juegos usando determinados inputs

```
SELECT name FROM speed.juegos

WHERE game_id IN

(SELECT DISTINCT s_c_ju_game_id FROM speed.run

WHERE time < :valor2

AND j_player IN

(SELECT player FROM speed.jugador

WHERE country = :valor3)

AND s_c_ju_game_id IN

(SELECT game_id FROM speed.juegos

WHERE name LIKE :valor1))
```

11 ;

En :valor1 deben ir strings como input, un ejemplo de franquicia sería 'Call of Duty'. Por otro lado, para el :valor2 deben ir un número entero mayor a cero como input, donde un ejemplo sería 20. Finalemente, en el :valor3 deben ir strings como input, en este caso al ser país, un ejemplo sería: 'United States'. Dejar en blanco el campo de franquicia permite ver juegos de cualquier saga.

4.2. Optimización

4.2.1. Índices

1. B+ Tree sobre time en la tabla run, ya que dentro de las consultas posibles podríamos buscar las runs cuyo tiempo sea mayor o menor a un tiempo deseado, o sea, haríamos una búsqueda por rango.

- 2. Hash sobre country en la tabla jugador, ya que acá nunca se realizarán búsquedas por rangos, solo se utilizarán igualdades para buscar los paises buscados.
- 3. B+ Tree sobre record_date en la tabla run. Utilizándo el mismo argumento para time, podríamos buscar todas las runs cuya fecha sea más reciente que la fecha especificada.

4.3. Seguridad

Utilizar PDO::prepare() junto con PDOStatement::execute() para preparar y ejecutar declaraciones que se usarán repetidamente con diversos parámetros mejora la eficiencia de la aplicación. Además, contribuye a prevenir ataques de inyección SQL al eliminar la necesidad de realizar manualmente la manipulación de comillas alrededor de los parámetros.

Figura 9: Consulta.php

4.4. Interfaz

El link de acceso para la aplicación web: https://grupo40.cc3201.dcc.uchile.cl A continuación presentaremos pantallazos para evidenciar el correcto uso de la página.

4.4.1. Consulta 1



Figura 10: Consulta 1 con ejemplo de input



Figura 11: Tabla de resultado consulta 1

Hito $n^{\varrho}3$ CC3201-1 Bases de Datos

4.4.2. Consulta 2

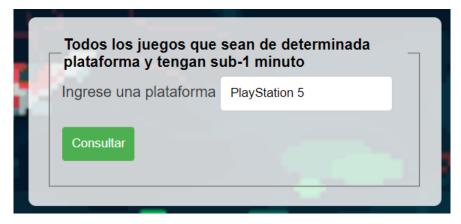


Figura 12: Consulta 2 con ejemplo de input

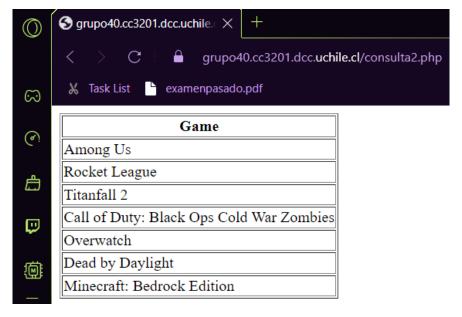


Figura 13: Tabla de resultado consulta 2

Hito 3 17

4.4.3. Consulta 3



Figura 14: Consulta 3 con ejemplo de input

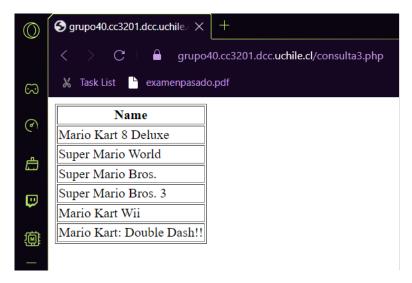


Figura 15: Tabla de resultado consulta 3