

## DESAFÍO DE ANALISTA DE DATOS DE TALENTPITCH

Para este desafío se utilizaron diferentes herramientas para la solución de cada uno de los puntos, entre ellas tenemos Python, MySQL Workbench y Power BI, explorando cada uno de los puntos, tenemos que:

### 1. API

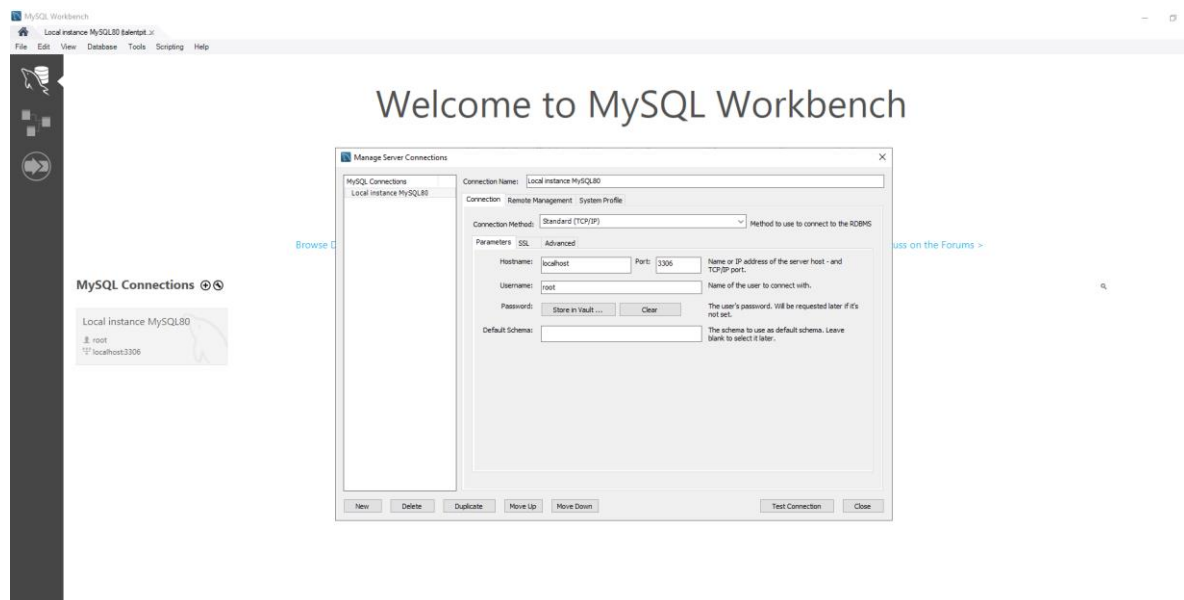
Se presentaron algunos inconvenientes con la creación de esta API, sin embargo, utilizando Python como herramienta lo gramos cargar los archivos, realizar algunos pequeños cambios en cuanto a limpieza de datos y posteriormente, se cargaron todos los archivos en una base de datos a la que se llamó “talentpitch”, se crearon 4 tablas y a estas se les dejó el mismo nombre que contiene el archivo, para prueba de ello dejo el script utilizado e imagen de la BD creada:

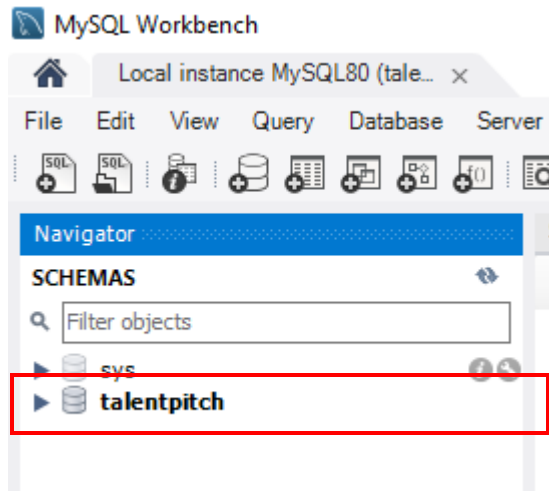
Script:



ETL.html

BD:





## 2. Tarea SQL:

Después de realizada la carga de la data a la respectiva base de datos creada, en las tablas correspondientes, se realizaron los Querys según los retos propuestos:

- Query 1:

The screenshot shows the MySQL Workbench interface with an SQL query editor and a result grid. The query is as follows:

```

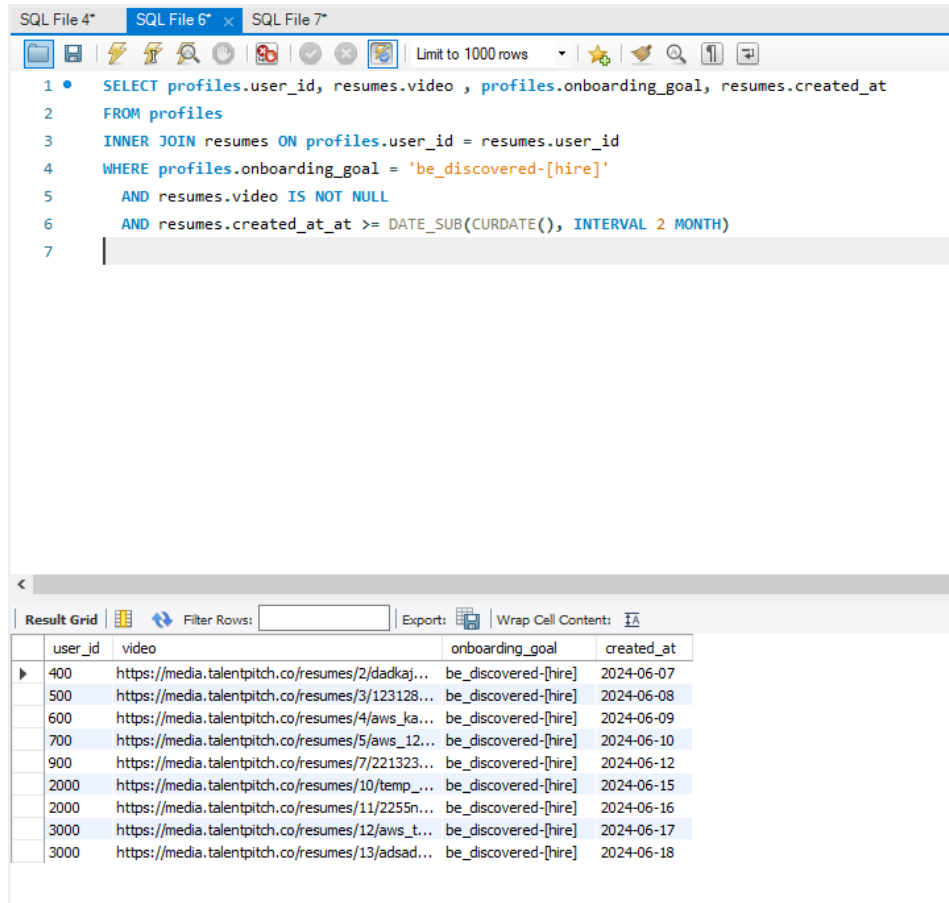
1 • SELECT opencall_objective, status, created_at
2 FROM challenges
3 WHERE status = 'published'
4 AND created_at >= DATE_SUB(CURDATE(), INTERVAL 3 MONTH)
5 ORDER BY created_at DESC;

```

The result grid shows the following data:

	opencall_objective	status	created_at
▶	Artistas	published	2024-09-21
	Innovador	published	2024-09-20

- Query 2:



The screenshot shows a SQL IDE with three tabs: 'SQL File 4\*', 'SQL File 6\*', and 'SQL File 7\*'. The active tab is 'SQL File 6\*', which contains the following SQL query:

```

1 • SELECT profiles.user_id, resumes.video , profiles.onboarding_goal, resumes.created_at
2 FROM profiles
3 INNER JOIN resumes ON profiles.user_id = resumes.user_id
4 WHERE profiles.onboarding_goal = 'be_discovered-[hire]'
5 AND resumes.video IS NOT NULL
6 AND resumes.created_at_at >= DATE_SUB(CURDATE(), INTERVAL 2 MONTH)
7

```

Below the query editor, the 'Result Grid' is displayed, showing the results of the query. The grid has four columns: 'user\_id', 'video', 'onboarding\_goal', and 'created\_at'. The results are as follows:

user_id	video	onboarding_goal	created_at
400	https://media.talentspith.co/resumes/2/dadkaj...	be_discovered-[hire]	2024-06-07
500	https://media.talentspith.co/resumes/3/123128...	be_discovered-[hire]	2024-06-08
600	https://media.talentspith.co/resumes/4/aws_ka...	be_discovered-[hire]	2024-06-09
700	https://media.talentspith.co/resumes/5/aws_12...	be_discovered-[hire]	2024-06-10
900	https://media.talentspith.co/resumes/7/221323...	be_discovered-[hire]	2024-06-12
2000	https://media.talentspith.co/resumes/10/temp_...	be_discovered-[hire]	2024-06-15
2000	https://media.talentspith.co/resumes/11/2255n...	be_discovered-[hire]	2024-06-16
3000	https://media.talentspith.co/resumes/12/aws_t...	be_discovered-[hire]	2024-06-17
3000	https://media.talentspith.co/resumes/13/adsad...	be_discovered-[hire]	2024-06-18

Para este ejercicio, no se pudieron tomar los 2 meses anteriores ya que la base de datos contiene información reducida.

- Query 3

SQL File 4*			
SQL File 6*			
SQL File 7*			
Limit to 1000 rows			
<pre> 1 • SELECT users.id, users.name, profiles.views, profiles.updated_at 2 FROM users 3 INNER JOIN profiles ON users.id = profiles.user_id 4 ORDER BY profiles.views DESC 5 LIMIT 5; </pre>			
Result Grid			
Filter Rows:			
Export:   Wrap Cell Content:   Fetch rows:			
id	name	views	updated_at
900	Ana beltran	200	2024-04-15
800	Juan jose	190	2024-04-22
3000	Clara benavidez	70	2024-04-19
1000	Luisa perez	59	2024-04-12
600	Pedro jimenez	40	2024-04-08

### 3. Visualizador

Para este punto, se utilizó el visualizador Power BI, desde este aplicativo se realizó la conexión a la base de datos creada, y se trajeron las tablas cargadas, allí se realizó el respectivo modelo relacional:



Informe Power  
BI.pdf

