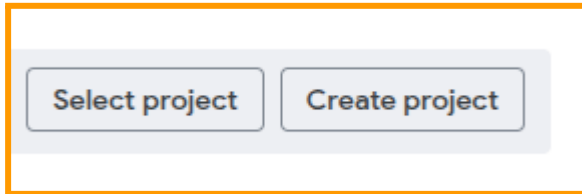
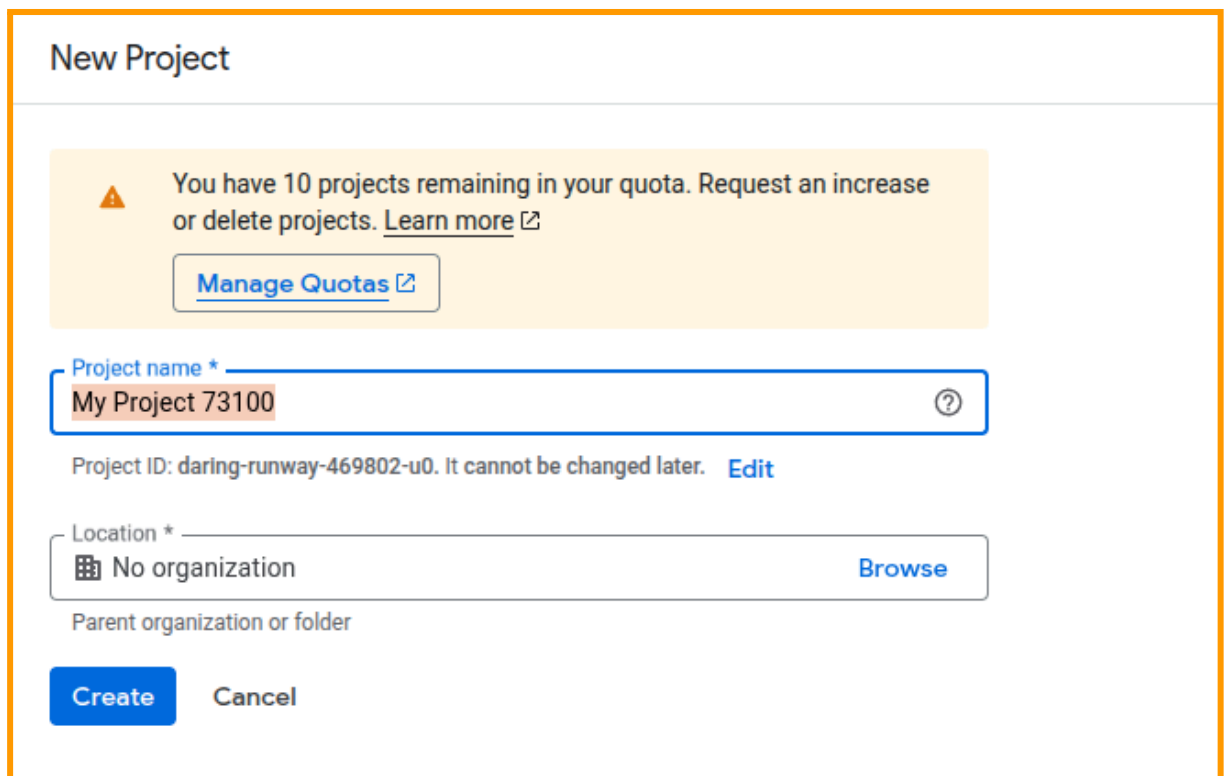


# Các bước để crawl mail

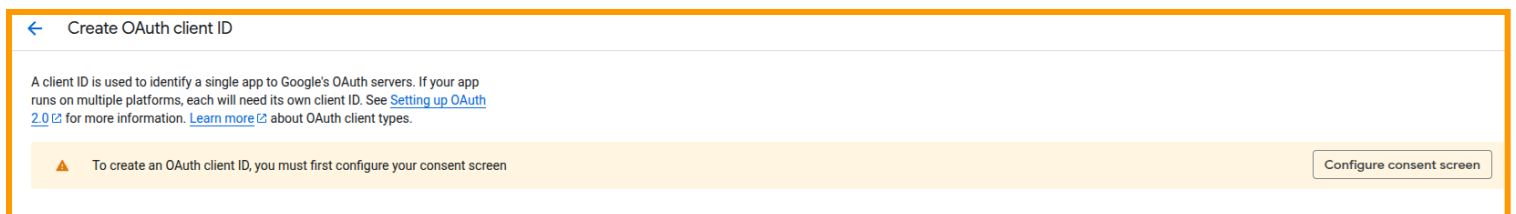
1. <https://console.cloud.google.com/apis/credentials> vào đường dẫn sử dụng chính mail muốn crawl



2. ở góc phải để tạo dự án và đặt tên rồi nhấn create để tạo dự án

A screenshot of the 'New Project' form in the Google Cloud Console. The form has a title 'New Project'. Below the title, there is a warning message: 'You have 10 projects remaining in your quota. Request an increase or delete projects. Learn more'. Below the warning, there is a 'Manage Quotas' button. The 'Project name' field is filled with 'My Project 73100'. Below the project name, the 'Project ID' is shown as 'daring-runway-469802-u0'. The 'Location' field is set to 'No organization'. At the bottom, there are 'Create' and 'Cancel' buttons.

3. Sau đó, Enable APIs and Services -> tìm GmailAPI và chọn enable
4. Credentials -> Create Credentials -> OAuth client ID
5. Sau đó sẽ hiện lên như này:


A screenshot of the 'Create OAuth client ID' page in the Google Cloud Console. The page has a title 'Create OAuth client ID'. Below the title, there is a warning message: 'To create an OAuth client ID, you must first configure your consent screen'. At the bottom right, there is a 'Configure consent screen' button.


6. Nhấn vào “configure consent screen” -> Xuất hiện “Branding” nhấn “get started” -> điền theo các mục với support mail là mail cần crawl -> Create OAuth client ID chọn desktop app . -> create -> Done


7. Sau đó bảng này sẽ xuất hiện:



### OAuth client created


The client ID can always be accessed from Clients tab under Google Auth Platform.

 OAuth access is restricted to the [test users](#) listed on your [OAuth consent screen](#)

Client ID	605648903613-vt1echqqjcsmr4iu47uv009kn1rna8.apps.googleusercontent.com	
-----------	--	---

 Starting in June 2025, you will no longer be able to view or download the client secret once you close this dialog. Make sure you have copied or downloaded the information below and securely stored it.

Client secret	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	
Creation date	August 22, 2025 at 9:38:14 AM GMT+7	
Status	 Enabled	

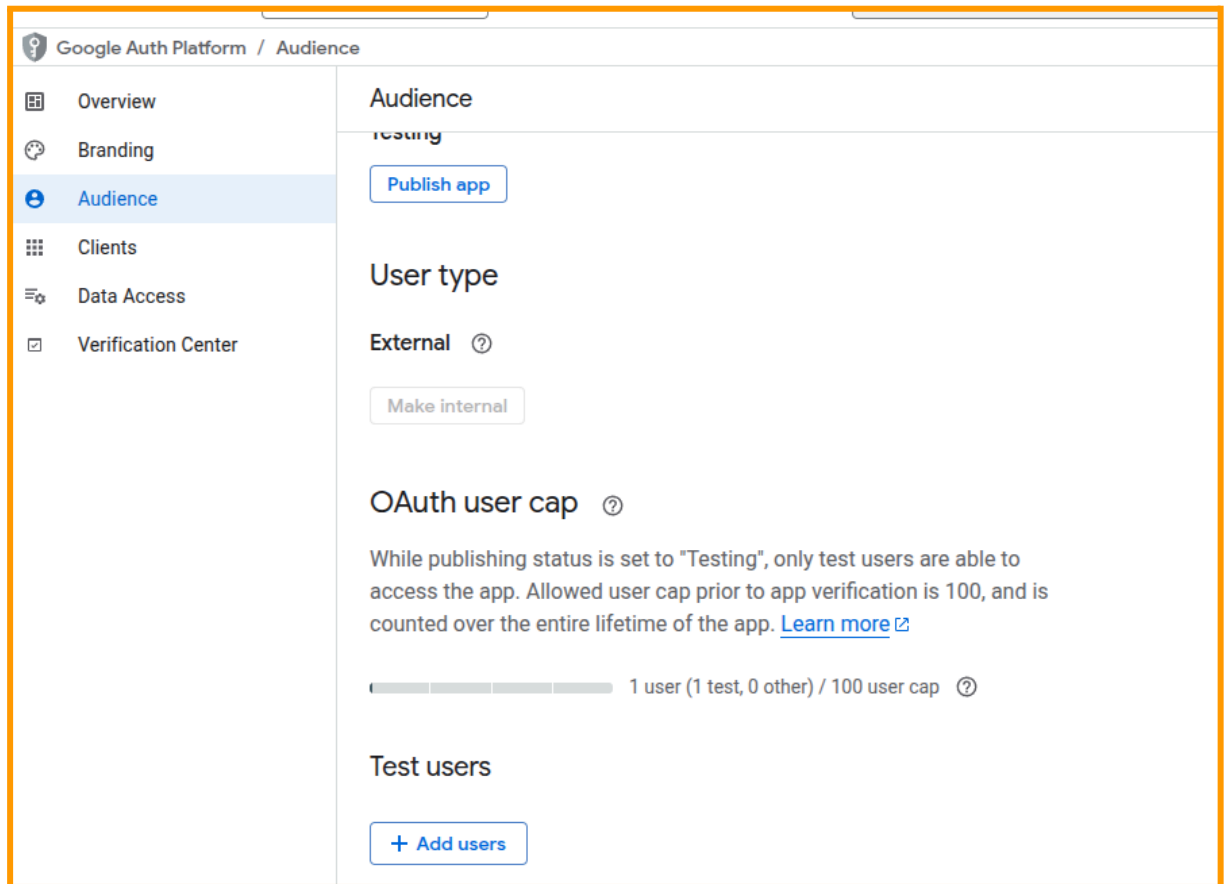
 Download JSON

OK

Nhấn download Json để có file token xác minh.

8. Sau đó đổi tên tệp thành: “client\_secret.json”

9. <https://console.cloud.google.com/auth/audience?authuser> đường dẫn này dẫn đến mục auth:



*Sau đó ở phần “Test user” nhấp add user và thêm email cần crawl vào.*

10. Lần đầu tiên dùng thì cần thì khi nhập :

```
from simplegmail import Gmail
gmail = Gmail()
```

Sẽ xuất hiện cửa sổ trình duyệt . yêu cầu chọn email cần crawl và cấp quyền. Thao tác này sau đó sẽ tự động lưu tệp “gmail-token.json”

11. Sau khi chắc chắn có file “gmail-token.json” thì có thể sử dụng liên kết với gmail cho sau này
12. Thực hiện code - cài đặt môi trường và thư viện:

+ **nếu có uv: thì chỉ cần:**

- tạo môi trường: `uv venv .venv`
- `source .venv/bin/activate`
- nhập “uv sync” để tải các thư viện cần
- Sau đó chạy: `python simple_gmail.py`

+ **Nếu ko có nhưng muốn sử dụng uv thì cần cài đặt uv trước như:**

```
# 1 - Install mise
curl -sSfL
https://github.com/jdx/mise/releases/download/v2025.1.6/install.sh |
sh
```

```
# 2 - Enable mise in your shell
echo 'eval "$(/.local/bin/mise activate bash)"' >> ~/.bashrc
source ~/.bashrc # or restart terminal
```

```
# 3 - Install required CLI tools
mise use -g uv just watchexec
```

+ **Nếu ko muốn cài uv thì có thể sử dụng pip instal với các thư viện:**

- pip install simplegmail
- pip install google-api-python-client google-auth-httpplib2 google-auth-oauthlib
- pip install openpyxl

13. ý tưởng code:

- Sử dụng simple gmail để lấy các trường thông tin của các gmail
- Đồng thời có 1 hàm sẽ lấy các giá trị của trường subject trong file .xlsx để so khớp tìm mail cần thiết
- Lưu mỗi file vào một folder chứa file json (mail) và các file tệp đính kèm của mail
- Lưu ý: cần thay đổi file excel mẫu thành file excel cần dùng. và tên cột cần lấy thông tin.
- và vị trí các file sẽ như dưới:

