Speech-to-Text API: Qwik Start

Table of Contents

1. Introduction

This lab demonstrates the **Google Cloud Speech-to-Text API**, which converts spoken audio into written text. Key features explored:

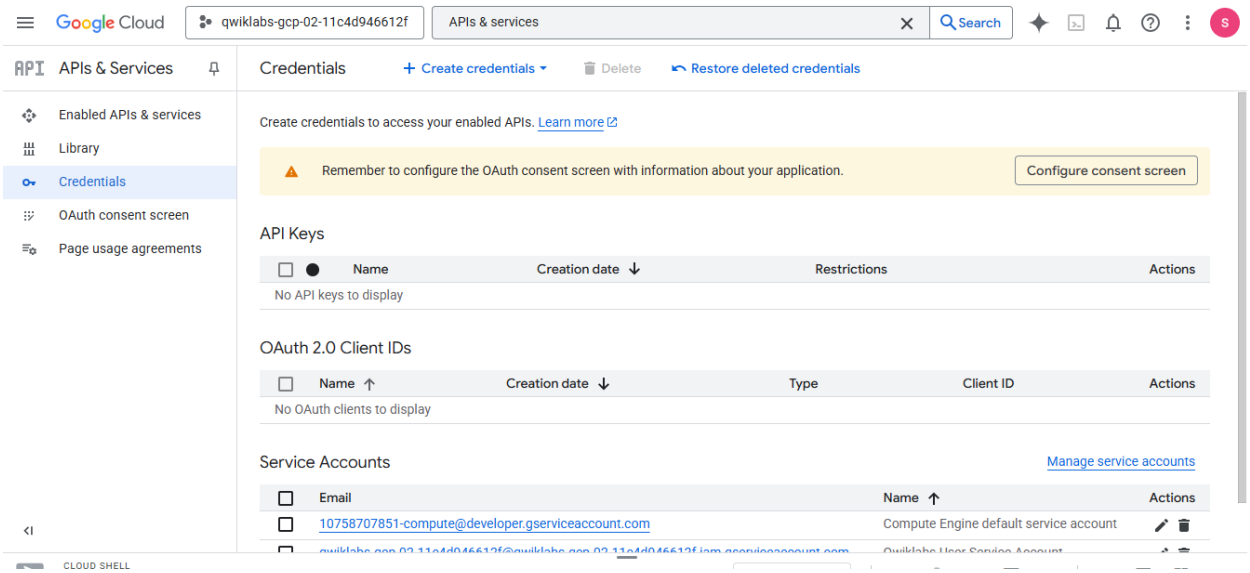
- **Synchronous transcription** of pre-recorded audio files.
- **Language detection** (English, in this case).
- **Confidence scoring** for accuracy assessment.

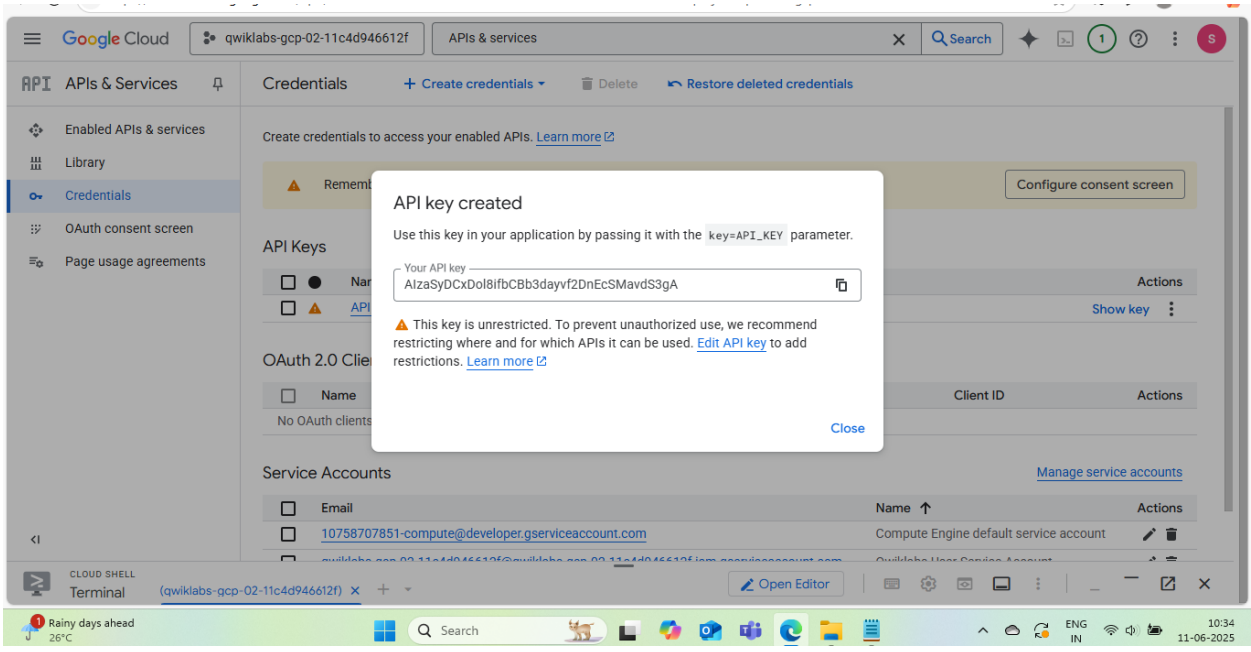
2. Task 1: Creating an API Key

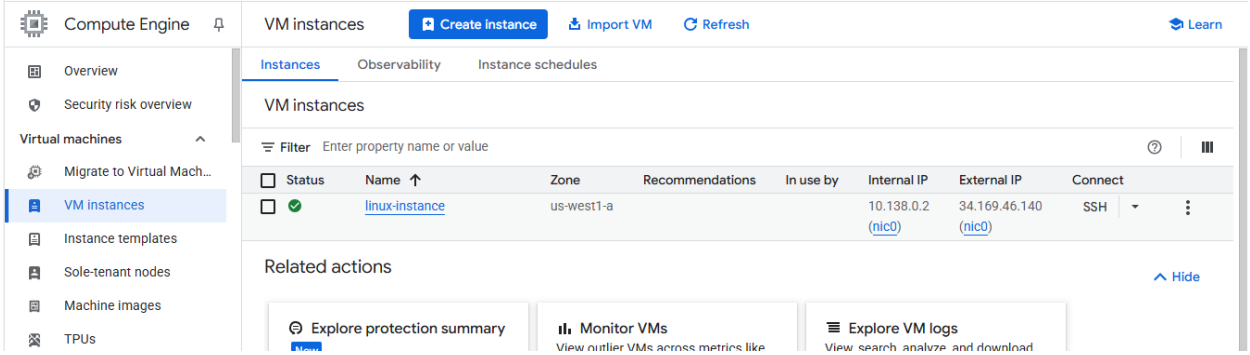
Steps Executed:

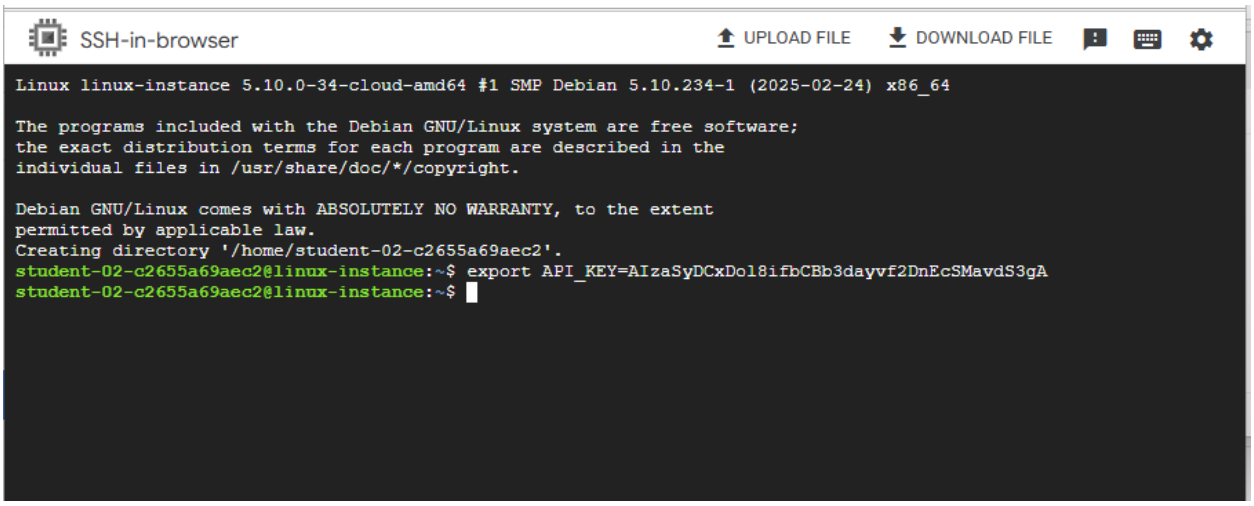
1. Navigated to **APIs & Services > Credentials** in the Cloud Console.
2. Generated a new **API key** for authentication.
3. Saved the key as an environment variable in the SSH session:

```
export API_KEY="YOUR_API_KEY"
```


Google Cloud
qwiklabs-gcp-02-11c4d946612f
APIs & services
Search
APIs & Services
Credentials
+ Create credentials
Delete
Restore deleted credentials
Enabled APIs & services
Library
Credentials
OAuth consent screen
Page usage agreements
Create credentials to access your enabled APIs. Learn more
Remember to configure the OAuth consent screen with information about your application.
Configure consent screen
API Keys
Name
Creation date
Restrictions
Actions
No API keys to display
OAuth 2.0 Client IDs
Name
Creation date
Type
Client ID
Actions
No OAuth clients to display
Service Accounts
Manage service accounts
Email
Name
Actions
10758707851-compute@developer.gserviceaccount.com
Compute Engine default service account
CLOUD SHELL


Google Cloud
qwiklabs-gcp-02-11c4d946612f
APIs & services
Search
APIs & Services
Credentials
+ Create credentials
Delete
Restore deleted credentials
Enabled APIs & services
Library
Credentials
OAuth consent screen
Page usage agreements
Create credentials to access your enabled APIs. Learn more
Configure consent screen
API key created
Use this key in your application by passing it with the key=API_KEY parameter.
Your API key
AIzaSyDCxDol8ifbCBb3dayvf2DnEcSMavdS3gA
This key is unrestricted. To prevent unauthorized use, we recommend restricting where and for which APIs it can be used. Edit API key to add restrictions. Learn more
Close
API Keys
Show key
OAuth 2.0 Clients
Service Accounts
Manage service accounts
Email
Name
Actions
10758707851-compute@developer.gserviceaccount.com
Compute Engine default service account
CLOUD SHELL
Terminal
(qwiklabs-gcp-02-11c4d946612f)
Open Editor
Rainy days ahead
26°C
Search
ENG
IN
10:34
11-06-2025


Compute Engine
VM instances
Create instance
Import VM
Refresh
Learn
Overview
Security risk overview
Virtual machines
Migrate to Virtual Mach...
VM instances
Instance templates
Sole-tenant nodes
Machine images
TPUs
Instances
Observability
Instance schedules
VM instances
Filter
Enter property name or value
Status
Name
Zone
Recommendations
In use by
Internal IP
External IP
Connect
linux-instance
us-west1-a
10.138.0.2 (nic0)
34.169.46.140 (nic0)
SSH
Related actions
Hide
Explore protection summary
Monitor VMs
Explore VM logs



Purpose:

- Securely authenticate requests to the Speech-to-Text API.

3. Task 2: Preparing the API Request

Audio File Used:

- **Source**: gs://cloud-samples-tests/speech/brooklyn.flac
- **Content**: *"How old is the Brooklyn Bridge?"*

Request Configuration:

Created request.json with:

{

"config": {

"encoding": "FLAC",

"languageCode": "en-US"

},

"audio": {

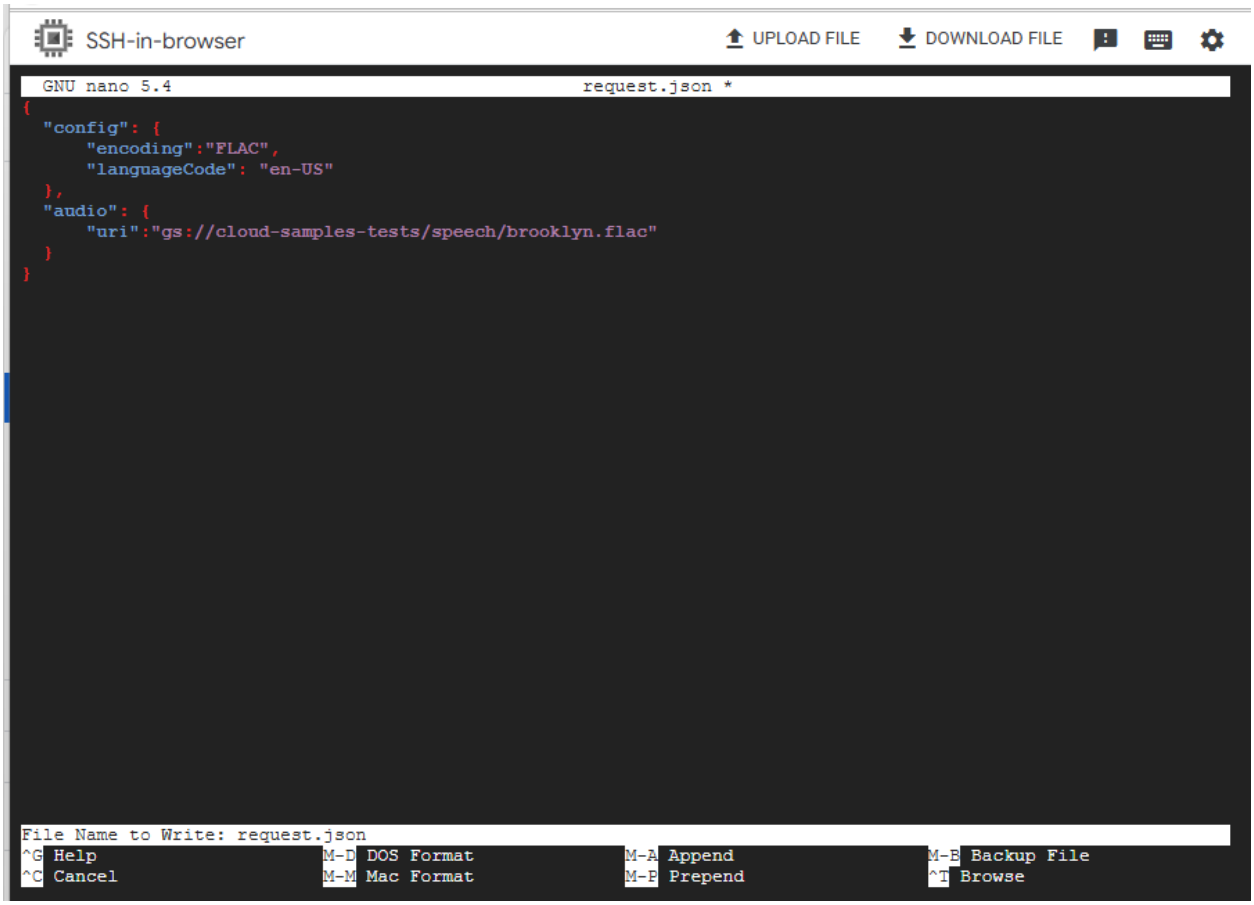
"uri": "gs://cloud-samples-tests/speech/brooklyn.flac"

}

}

Key Metrics:

Field	Value	Significance
`transcript`	"how old is the Brooklyn Bridge"	Accurate transcription of audio.
`confidence`	`0.982` (98.27%)	API is 98.27% confident in the result.



```
student-02-c2655a69aec2@linux-instance:~$ curl -s -X POST -H "Content-Type: application/json" --data-binary @req
uest.json \
"https://speech.googleapis.com/v1/speech:recognize?key=${API_KEY}"
{
  "results": [
    {
      "alternatives": [
        {
          "transcript": "how old is the Brooklyn Bridge",
          "confidence": 0.9311077
        }
      ],
      "resultEndTime": "1.770s",
      "languageCode": "en-us"
    }
  ],
  "totalBilledTime": "2s",
  "requestId": "5652133940788343918"
}
student-02-c2655a69aec2@linux-instance:~$
```

```
student-02-c2655a69aec2@linux-instance:~$ curl -s -X POST -H "Content-Type: application/json" --data-binary @req
uest.json \
"https://speech.googleapis.com/v1/speech:recognize?key=${API_KEY}" > result.json
student-02-c2655a69aec2@linux-instance:~$
```

5. Results & Analysis

✅ **Accuracy**: The API correctly transcribed the audio with **high confidence (98.27%)**.

✅ **Latency**: Synchronous response was near-instantaneous.

✅ **Use Case**: Suitable for applications like voice assistants, call center analytics, and real-time captioning.

6. Key Learnings

1. **API Setup**:
    a. API keys authenticate requests without exposing credentials.
    b. Environment variables enhance security.
2. **Audio Requirements**:
    a. Supported formats include FLAC, WAV, and LINEAR16.
    b. Language codes must be specified (e.g., en-US).
3. **Confidence Scores**:
    a. Values close to 1.0 indicate high accuracy.

7. Conclusion

This lab successfully demonstrated:

- **Speech-to-Text transcription** using Google Cloud's AI.
- **API integration** via curl commands.
- **Practical insights** into audio processing and confidence metrics.

Future Applications:

- **Real-time transcription** for live events.
- **Multilingual support** (e.g., es-ES, fr-FR).