Data Loss Prevention: Qwik Start - JSON

30 minutesFree

**GSP107**



The Data Loss Prevention API provides programmatic access to a powerful detection engine for personally identifiable information (PII) and other privacy-sensitive data in unstructured data streams.

The DLP API provides fast, scalable classification and optional redaction for sensitive data elements like credit card numbers, names, social security numbers, passport numbers, and phone numbers. The API supports text and images – just send data to the API or specify data stored on your Cloud Storage, BigQuery, and Cloud Datastore instances.

In this lab, you set up a JSON file to analyze, send it to the Data Loss Prevention API, to inspect a string of data for sensitive information, then redact any sensitive information that was found.

What you'll learn

* Use the DLP API to inspect a string for sensitive information.
* Use the DLP API to redact sensitive data from text content.

**Setup and requirements**

Before you click the Start Lab button

Read these instructions. Labs are timed and you cannot pause them. The timer, which starts when you click **Start Lab**, shows how long Google Cloud resources will be made available to you.

This hands-on lab lets you do the lab activities yourself in a real cloud environment, not in a simulation or demo environment. It does so by giving you new, temporary credentials that you use to sign in and access Google Cloud for the duration of the lab.

To complete this lab, you need:

* Access to a standard internet browser (Chrome browser recommended).

**Note:** Use an Incognito or private browser window to run this lab. This prevents any conflicts between your personal account and the Student account, which may cause extra charges incurred to your personal account.

* Time to complete the lab---remember, once you start, you cannot pause a lab.

**Note:** If you already have your own personal Google Cloud account or project, do not use it for this lab to avoid extra charges to your account.

How to start your lab and sign in to the Google Cloud Console

1. Click the **Start Lab** button. If you need to pay for the lab, a pop-up opens for you to select your payment method. On the left is the **Lab Details** panel with the following:
   * The **Open Google Console** button
   * Time remaining
   * The temporary credentials that you must use for this lab
   * Other information, if needed, to step through this lab
2. Click **Open Google Console**. The lab spins up resources, and then opens another tab that shows the **Sign in** page.

***Tip:*** Arrange the tabs in separate windows, side-by-side.

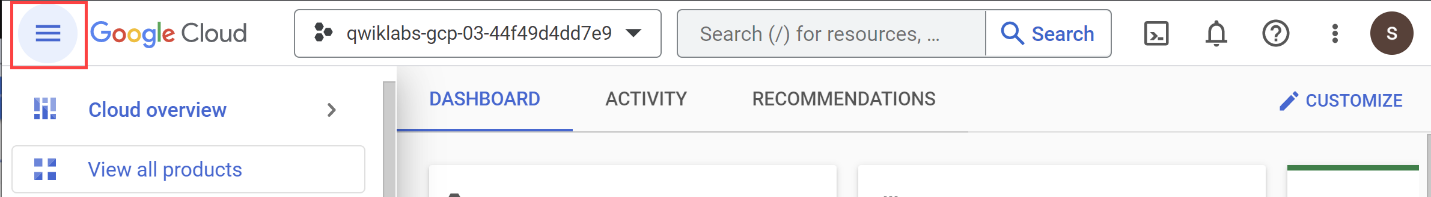
**Note:**If you see the **Choose an account** dialog, click **Use Another Account**.

1. If necessary, copy the **Username** from the **Lab Details** panel and paste it into the **Sign in** dialog. Click **Next**.
2. Copy the **Password** from the **Lab Details** panel and paste it into the **Welcome** dialog. Click **Next**.

**Important:**You must use the credentials from the left panel. Do not use your Google Cloud Skills Boost credentials.**Note:**Using your own Google Cloud account for this lab may incur extra charges.

1. Click through the subsequent pages:
   * Accept the terms and conditions.
   * Do not add recovery options or two-factor authentication (because this is a temporary account).
   * Do not sign up for free trials.

After a few moments, the Cloud Console opens in this tab.

**Note:** You can view the menu with a list of Google Cloud Products and Services by clicking the **Navigation menu** at the top-left. 

Activate Cloud Shell

Cloud Shell is a virtual machine that is loaded with development tools. It offers a persistent 5GB home directory and runs on the Google Cloud. Cloud Shell provides command-line access to your Google Cloud resources.

1. Click **Activate Cloud Shell** Activate Cloud Shell icon at the top of the Google Cloud console.

When you are connected, you are already authenticated, and the project is set to your **PROJECT\_ID**. The output contains a line that declares the **PROJECT\_ID** for this session:

Your Cloud Platform project in this session is set to YOUR\_PROJECT\_ID

gcloud is the command-line tool for Google Cloud. It comes pre-installed on Cloud Shell and supports tab-completion.

1. (Optional) You can list the active account name with this command:

gcloud auth list

Copied!

content\_copy

1. Click **Authorize**.
2. Your output should now look like this:

**Output:**

ACTIVE: \*

ACCOUNT: student-01-xxxxxxxxxxxx@qwiklabs.net

To set the active account, run:

$ gcloud config set account `ACCOUNT`

1. (Optional) You can list the project ID with this command:

gcloud config list project

Copied!

content\_copy

**Output:**

[core]

project = <project\_ID>

**Example output:**

[core]

project = qwiklabs-gcp-44776a13dea667a6

**Note:**For full documentation of gcloud, in Google Cloud, refer to [the gcloud CLI overview guide](https://cloud.google.com/sdk/gcloud).

Set an environmental variable for your project ID

* In Cloud Shell, run the following command to set an environment variable for your project ID:

export PROJECT\_ID=$DEVSHELL\_PROJECT\_ID

Copied!

content\_copy

**Task 1. Inspect a string for sensitive information**

This section shows you how to ask the service to scan sample text using the [projects.content.inspect](https://cloud.google.com/dlp/docs/reference/rest/v2beta2/projects.content/inspect) REST method. The JSON file you create contains an [InspectConfig](https://cloud.google.com/dlp/docs/reference/rest/v2beta2/InspectConfig) and a [ContentItem](https://cloud.google.com/dlp/docs/reference/rest/v2beta2/ContentItem) object.

1. Using your preferred editor (nano, vim, etc.) or Cloud Shell, create a JSON request file with the following text, and save it as inspect-request.json:

{

"item":{

"value":"My phone number is (206) 555-0123."

},

"inspectConfig":{

"infoTypes":[

{

"name":"PHONE\_NUMBER"

},

{

"name":"US\_TOLLFREE\_PHONE\_NUMBER"

}

],

"minLikelihood":"POSSIBLE",

"limits":{

"maxFindingsPerItem":0

},

"includeQuote":true

}

}

Copied!

content\_copy

1. Obtain an authorization token using your account:

gcloud auth print-access-token

Copied!

content\_copy

A huge string is returned. You need this token for the next step.

If you receive an error that no service account is being used, wait a few minutes and run the command again.

1. Use curl to make a content:inspect request, replacing ACCESS\_TOKEN with the string that was returned in the previous step:

curl -s \

-H "Authorization: Bearer ACCESS\_TOKEN" \

-H "Content-Type: application/json" \

https://dlp.googleapis.com/v2/projects/$PROJECT\_ID/content:inspect \

-d @inspect-request.json -o inspect-output.txt

Copied!

content\_copy

**Note: Here's what's going on**

To pass a filename to curl you use the -d option (for "data") and precede the filename with an @ sign. This file should be in the same directory in which you execute the curl command.

It saves the curl response in inspect-output.txt file. Check the output using below command:

cat inspect-output.txt

Copied!

content\_copy

You should see a response similar to the following:

{

"result": {

"findings": [

{

"quote": "(206) 555-0123",

"infoType": {

"name": "PHONE\_NUMBER"

},

"likelihood": "LIKELY",

"location": {

"byteRange": {

"start": "19",

"end": "33"

},

"codepointRange": {

"start": "19",

"end": "33"

}

},

"createTime": "2018-07-03T02:20:26.043Z"

}

]

}

}

Upload output to Cloud Storage

Run the following command to upload the curl response on Cloud Storage for activity tracking validation:

gsutil cp inspect-output.txt gs://qwiklabs-gcp-00-096b7a4815a4-bucket

Copied!

content\_copy

Inspect a string for sensitive information

Check my progress

**Task 2. Redacting sensitive data from text content**

The Data Loss Prevention API can automatically redact sensitive data from text files instead of giving you a list of findings.

Try sending the API JSON file using [deidentifyConfig](https://cloud.google.com/dlp/docs/reference/rest/v2beta2/projects.deidentifyTemplates) object, so sensitive information is redacted from the output.

1. Create a new JSON file (called new-inspect-file.json) that includes the following:

{

"item": {

"value":"My email is test@gmail.com",

},

"deidentifyConfig": {

"infoTypeTransformations":{

"transformations": [

{

"primitiveTransformation": {

"replaceWithInfoTypeConfig": {}

}

}

]

}

},

"inspectConfig": {

"infoTypes": {

"name": "EMAIL\_ADDRESS"

}

}

}

Copied!

content\_copy

1. Use curl to make a content:deidentify request (ACCESS\_TOKEN has been replaced with a command to print the access token):

curl -s \

-H "Authorization: Bearer $(gcloud auth print-access-token)" \

-H "Content-Type: application/json" \

https://dlp.googleapis.com/v2/projects/$PROJECT\_ID/content:deidentify \

-d @new-inspect-file.json -o redact-output.txt

Copied!

content\_copy

It saves the curl response in redact-output.txt file. Check the output using below command:

cat redact-output.txt

Copied!

content\_copy

You should see a response similar to the following:

{

"item": {

"value": "My email is [EMAIL\_ADDRESS]"

},

"overview": {

"transformedBytes": "14",

"transformationSummaries": [

{

"infoType": {

"name": "EMAIL\_ADDRESS"

},

"transformation": {

"replaceWithInfoTypeConfig": {}

},

"results": [

{

"count": "1",

"code": "SUCCESS"

}

],

"transformedBytes": "14"

}

]

}

}

You've sent your first request to the Data Loss Prevention API and redacted sensitive information from output!

Upload output to Cloud Storage

Run the following command to upload the curl response on Cloud Storage for activity tracking validation:

gsutil cp redact-output.txt gs://qwiklabs-gcp-00-096b7a4815a4-bucket

Copied!

content\_copy

Redacting sensitive data from text content

Check my progress

**Congratulations!**

Finish your quest

This self-paced lab is part of the [Baseline: Deploy & Develop](https://www.cloudskillsboost.google/quests/37) and [Security & Identity Fundamentals](https://www.cloudskillsboost.google/quests/40) quests. A quest is a series of related labs that form a learning path. Completing a quest earns you a badge to recognize your achievement. You can make your badge or badges public and link to them in your online resume or social media account. Enroll in any quest that contains this lab and get immediate completion credit. Refer to the [catalog](http://cloudskillsboost.google/catalog) for all available quests.

Take your next lab

Continue your quest with these suggestions:

* [Cloud Function: Qwik Start - Command Line](https://www.cloudskillsboost.google/catalog_lab/924)
* [Getting Started with Cloud KMS](https://www.cloudskillsboost.google/catalog_lab/368)
* [Setting up a Private Kubernetes Cluster](https://www.cloudskillsboost.google/catalog_lab/908)
* [Video Intelligence: Qwik Start](https://www.cloudskillsboost.google/catalog_lab/804)

Next steps / learn more

This lab is also part of a series of labs called Qwik Starts. These labs are designed to give you a little taste of the many features available with Google Cloud. Search for "Qwik Starts" in the [lab catalog](https://google.qwiklabs.com/catalog) to find the next lab you'd like to take!

Google Cloud training and certification

...helps you make the most of Google Cloud technologies. [Our classes](https://cloud.google.com/training/courses) include technical skills and best practices to help you get up to speed quickly and continue your learning journey. We offer fundamental to advanced level training, with on-demand, live, and virtual options to suit your busy schedule. [Certifications](https://cloud.google.com/certification/) help you validate and prove your skill and expertise in Google Cloud technologies.

**Manual Last Updated: May 31, 2023**

**Lab Last Tested: May 31, 2023**

Copyright 2023 Google LLC All rights reserved. Google and the Google logo are trademarks of Google LLC. All other company and product names may be trademarks of the respective companies with which they are associated.