

COVID-19 ANALYSIS ON GCP

Data Science for Management – LM91



Università
di Catania



Introduction

The COVID-19 pandemic of 2019-2020 is the current pandemic of the so-called “new coronavirus disease”, caused by the SARS-CoV-2 virus.

Patients experience flu-like symptoms such as dermatitis, fever, dry cough, tiredness, difficulty breathing. In the most serious cases, often found in subjects already burdened by previous pathologies, pneumonia develops, acute renal failure, up to even death.

Patients also have leukopenia and lymphocytopenia. A test to diagnose SARS-CoV-2 infection has been available since mid-January 2020, but specific treatments and vaccines are currently being researched and tested.

The healings are spontaneous and the treatments are mainly aimed at managing symptoms and supporting vital functions even if some antiviral drugs already used to contrast other infections have been tested.

The aim of this project is to analyze data about COVID-19 using Google Cloud Platform.

The dataset

Data are stored in two GitHub repositories:

- [Dipartimento di Protezione Civile](#) ([Region](#) and [Province](#)) that stores data about:

Region

Fields	Type	Description
data	datetime	Date of update
nuovi_casi	integer	New positive cases
totale_casi	integer	Total number of cases
totale_positivi	integer	Total number of positive cases
tamponi	integer	Total number of tests
tamponi_test_antigenico_rapido	integer	Number of rapid tests
totale_positivi_test_antigenico_rapido	integer	Number of positive related to rapid tests
tamponi_test_molecolare	integer	Number of PCR tests
totale_positivi_test_molecolare	integer	Number of positive related to PCR tests
ingressi_terapia_intensiva	integer	Admission to intensive care
terapia_intensiva	integer	Number of cases in intensive care
totale_ospedalizzati	integer	Total number of hospitalized cases
ricoverati_con_sintomi	integer	Total number of hospitalized with symptoms

Province

Fields	Type	Description
data	datetime	Date of update
denominazione_regione	string	Region name
denominazione_provincia	string	Province name
totale_casi	integer	Total number of cases

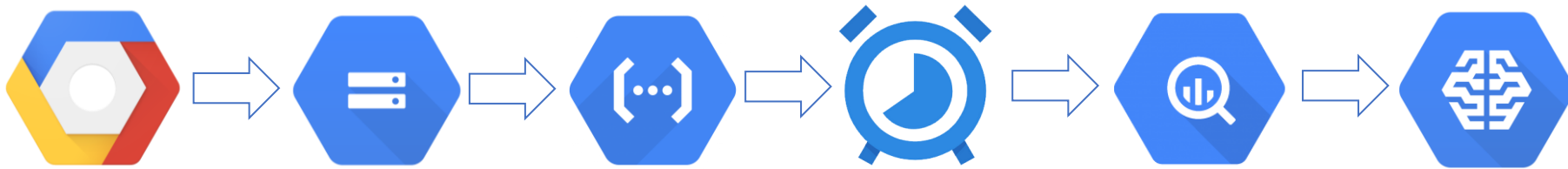
The dataset

- [Commissario straordinario per l'emergenza Covid-19](#) (about [vaccines](#)) that stores all the data related to vaccinations:

Fields	Type	Description
fornitore	string	Full name of the vaccine supplier
data_somministrazione	datetime	Date of administration
fascia_anagrafica	string	Age range of the subjects to which the vaccine was administered
Sesso_maschile	integer	Total of male subjects to which the vaccine has been administered by day, region and age group
Sesso_femminile	integer	Total of female subjects to which the vaccine has been administered by day, region and age group
prima_dose	integer	Number of first shots
seconda_dose	integer	Number of second shots
nome_regione	string	Standard area name (where a bilingual naming required)

The workflow

Here we describe the cloud workflow to download and update the data automatically from GitHub repository, the analysis using Notebooks and the data visualization



The workflow – Google Cloud Platform



New Project



You have 20 projects remaining in your quota. Request an increase or delete projects. [Learn more](#)

[MANAGE QUOTAS](#)

Project name *

covid-19-analysis



Project ID: covid-19-analysis-318015. It cannot be changed later. [EDIT](#)

Location *



No organization

[BROWSE](#)

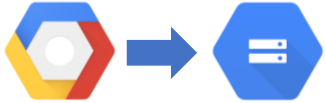
Parent organization or folder

CREATE

CANCEL

In Google Cloud create a new project naming it "covid-19-analysis"

The workflow – Google Cloud Storage



← Create a bucket

- **Name your bucket**
Pick a globally unique, permanent name. [Naming guidelines](#)

Tip: Don't include any sensitive information
CONTINUE
- Choose where to store your data
- Choose a default storage class for your data
- Choose how to control access to objects
- Advanced settings (optional)

CREATE [CANCEL](#)

1. In Google Cloud Storage create a new bucket and name it "downloaded_data"
2. Leave everything default and click on CREATE to create the bucket

The workflow – Google Cloud Functions



Cloud Functions [←](#) Create function

1 Configuration — 2 Code

Basics

Function name *
download_vaccini ?

Region
us-central1 ▼ ?

Trigger

⚙️ Cloud Pub/Sub

Trigger type
Cloud Pub/Sub ▼

Select a Cloud Pub/Sub topic * ▼

☐ Retry on failure ?

SAVE CANCEL

1. In Google Cloud Functions we create a new functions and name it as "download_vaccini"
2. Select Cloud Pub/Sub as trigger

The workflow – Google Cloud Functions



The screenshot shows the 'Create function' wizard in the Google Cloud console. The 'Configuration' tab is active, showing the 'Basics' section with 'Function name' set to 'download_vaccini' and 'Region' set to 'us-central1'. The 'Trigger' section is expanded, showing 'Cloud Pub/Sub' as the trigger type. A dropdown menu is open for 'Select a Cloud Pub/Sub topic *', displaying a list of topics including 'covid-19-analysis-318015'. At the bottom of the dropdown, there is a message 'There are no available topics in this project' and a 'CREATE A TOPIC' button. The 'CANCEL' and 'OK' buttons are also visible at the bottom of the dropdown.

1. In Google Cloud Functions we create a new functions and name it as "download_vaccini"
2. Select Cloud Pub/Sub as trigger
3. We need to select the topic, in our case we need to create a new one clicking on CREATE A TOPIC

The workflow – Google Cloud Functions



Cloud Functions Create function

1 Configuration — 2 Code

Basics

Function name *
download_vaccini

Create a topic

Topic ID *
vaccini

Topic name: projects/covid-19-analysis-318015/topics/vaccini

☐ Use a schema

☐ Use a customer-managed encryption key (CMEK)

CANCEL CREATE TOPIC

Filter type to filter topics

covid-19-analysis-318015

There are no available topics in this project

CREATE A TOPIC CANCEL OK

1. In Google Cloud Functions we create a new functions and name it as "download_vaccini"
2. Select Cloud Pub/Sub as trigger
3. We need to select the topic, in our case we need to create a new one clicking on CREATE A TOPIC
4. We name the Topic ID as "vaccini" and CREATE TOPIC

The workflow – Google Cloud Functions



Cloud Functions [←](#) Create function

1 Configuration — 2 Code

Basics

Function name *
download_vaccini ?

Region
us-central1 ?

Trigger

⚡ Cloud Pub/Sub

Trigger type
Cloud Pub/Sub ▼

Select a Cloud Pub/Sub topic *
projects/covid-19-analysis-318015/topics/vaccini ▼

☒ Retry on failure ?
Warning: your function will retry on failure for up to seven days even if it is due to a bug in your code.

SAVE CANCEL

1. In Google Cloud Functions we create a new functions and name it as "download_vaccini"
2. Select Cloud Pub/Sub as trigger
3. We need to select the topic, in our case we need to create a new one clicking on CREATE A TOPIC
4. We name the Topic ID as "vaccini" and CREATE TOPIC
5. We want also to retry on failure

The workflow – Google Cloud Functions



Runtime environment variables ?

Name *	Value
URL	https://github.com/italia/covid19-op
FILE_NAME	somministrazioni-vaccini-latest.csv
BUCKET	downloaded_data

[+ ADD VARIABLE](#)

[NEXT](#) [CANCEL](#)

1. In Google Cloud Functions we create a new functions and name it as "download_vaccini"
2. Select Cloud Pub/Sub as trigger
3. We need to select the topic, in our case we need to create a new one clicking on CREATE A TOPIC
4. We name the Topic ID as "vaccini" and CREATE TOPIC
5. We want also to retry on failure
6. Add three environment variables
 - URL
 - FILE_NAME
 - BUCKETClick on NEXT

The workflow – Google Cloud Functions



(...) Cloud Functions ← Create function

✓ Configuration — 2 Code

Runtime

Node.js 12

Node.js 10

PHP 7.4 **BETA**

Python 3.9

Python 3.8

Python 3.7

Ruby 2.7

Ruby 2.6

1. Select Python 3.7 as runtime

The workflow – Google Cloud Functions



Cloud Functions [← Create function](#)

Configuration — **2** Code

Runtime: Python 3.7

Source code: [Inline Editor](#)

- main.py
- requirements.txt

```
1 import os
2 import wget
3
4 from google.cloud import storage
5
6 url = os.environ['URL']
7 bucket_name = os.environ['BUCKET'] #without gs://
8 file_name = os.environ['FILE_NAME']
9
10 cf_path = '/tmp/{}'.format(file_name)
11
12 def import_file(event, context):
13     # set storage client
14     client = storage.Client()
15
16     # get bucket
17     bucket = client.get_bucket(bucket_name)
18
19     # download the file to Cloud Function's tmp directory
20     wget.download(url, cf_path)
21
22     # set Blob
23     blob = storage.Blob(file_name, bucket)
24
25     # upload the file to GCS
26     blob.upload_from_filename(cf_path)
27
28     print("This Function was triggered by messageId {} published at {}".format(context.event_id, context.timestamp))
```

Entry point: [import_file](#)

[PREVIOUS](#) [DEPLOY](#) [CANCEL](#)

1. Select Python 3.7 as runtime
2. In the main.py script we code the function to download the file from the [GitHub repository](#)
3. Set the entrypoint as the name of the python function

The workflow – Google Cloud Functions



(...) Cloud Functions ← Create function

✓ Configuration — 2 Code

Runtime
Python 3.7

Entry point
import_file

Source code
Inline Editor

main.py

requirements.txt

```
1 # Function dependencies, for example:
2 # package>=version
3
4 google-cloud-storage
5 wget
```

PREVIOUS DEPLOY CANCEL

1. Select Python 3.7 as runtime
2. In the main.py script we code the function to download the file from the [GitHub repository](#)
3. Set the entrypoint as the name of the python function
4. Insert the function dependencies
5. DEPLOY the function

The workflow – Google Cloud Functions



Cloud Functions

Functions

[+ CREATE FUNCTION](#)

[REFRESH](#)

Filter Filter functions

<input type="checkbox"/>	<input type="radio"/>	Name ↑	Region	Trigger	Runtime	Memory allocated	Executed function	Last deployed	Authentication ?	Actions
<input type="checkbox"/>	<input checked="" type="radio"/>	download_vaccini	us-central1	Topic: vaccini	Python 3.7	256 MiB	import_file	Jun 29, 2021, 9:01:31 PM		<div><div>Copy function</div><div>Test function</div><div>View logs</div><div>Delete</div></div>

The function has been correctly created but not deployed yet.
Just click on Test function to firstly deploy the function

The workflow – Google Cloud Functions



Cloud Functions

Function details

EDIT DELETE COPY

download_vaccini

Version
Version 1, deployed at Jun 29, 2021, 9:01:31 P...

METRICS DETAILS SOURCE VARIABLES TRIGGER PERMISSIONS LOGS TESTING

Triggering event ?

1
2

TEST THE FUNCTION

Testing in the Cloud Console has a 60s timeout. Note that this is different from the limit set in the function configuration.

1. Leave everything default just click on TEST THE FUNCTION

The workflow – Google Cloud Functions



Cloud Functions

Function details

EDIT DELETE COPY

download_vaccini

Version 1, deployed at Jun 29, 2021, 9:01:31 P...

METRICS DETAILS SOURCE VARIABLES TRIGGER PERMISSIONS LOGS TESTING

Output Complete

\$ OK

Logs

Scanned up to 6/11/21, 9:59 AM. Scanned 1.4 KB.

2021-06-29T19:02:21.920375383Z download_vaccini w06dyhqcp60l Function execution started

2021-06-29T19:02:28.610Z download_vaccini w06dyhqcp60l This Function was triggered by messageId 1940277171 published at 2021-06-29T19:02:19.184Z

2021-06-29T19:02:28.611736442Z download_vaccini w06dyhqcp60l Function execution took 6692 ms, finished with status: 'ok'

1. Leave everything default just click on TEST THE FUNCTION
2. The function execution started
3. When the function is finished return the status "ok"

The workflow – Google Cloud Functions



Cloud Functions

Functions

CREATE FUNCTION

REFRESH

Filter

Filter functions

<input type="checkbox"/>	<input checked="" type="checkbox"/>	Name ↑	Region	Trigger	Runtime	Memory allocated	Executed function	Last deployed	Authentication ?	Action
<input type="checkbox"/>	<input checked="" type="checkbox"/>	download_province	us-central1	Topic: province	Python 3.7	256 MiB	import_file	Jun 29, 2021, 9:05:28 PM		<div>⋮</div>
<input type="checkbox"/>	<input checked="" type="checkbox"/>	download_regioni	us-central1	Topic: regioni	Python 3.7	256 MiB	import_file	Jun 29, 2021, 9:07:36 PM		<div>⋮</div>
<input type="checkbox"/>	<input checked="" type="checkbox"/>	download_vaccini	us-central1	Topic: vaccini	Python 3.7	256 MiB	import_file	Jun 29, 2021, 9:01:31 PM		<div>⋮</div>

1. Leave everything default just click on TEST THE FUNCTION
2. The function execution started
3. When the function is finished return the status "ok"
4. Replicate the previous steps for the other files

The workflow – Google Cloud Scheduler



Cloud Scheduler | Jobs [+ CREATE JOB](#) [REFRESH](#) [EDIT](#) [PAUSE](#)

[SCHEDULER JOBS](#) [APP ENGINE CRON JOBS ?](#)

[Filter](#) Filter jobs

! Sorry, the server was not able to fulfill your request.

<input type="checkbox"/>	Name	State	Description	Frequency	Target	Last run	Result	Logs	Run
No rows to display									

1. Click on CREATE JOB to create a new schedule

The workflow – Google Cloud Scheduler



Cloud Scheduler

Create job

- 1 Select a location — 2 Create a job

Where would you like your Cloud Scheduler jobs to live?

Select a region for your Cloud Scheduler jobs. Note: you cannot change the region for this project later.



Select a region *


europa-west6

NEXT

1. Click on CREATE JOB to create a new schedule
2. Choose where the Cloud Scheduler jobs live

The workflow – Google Cloud Scheduler



 Cloud Scheduler [← Create a job](#)

- Define the job

Name *

update_regioni

Must be unique across all jobs in this project

Description

Update the GitHub repository of Covid19 regional data

Frequency *

0 20 ***

Schedules are specified using unix-cron format. E.g. every minute: `* * * * *`, every 3 hours: `"0 */3 * * *"`, every monday at 9:00: `"0 9 * * 1"`. [Learn more](#)

Timezone *

Central European Summer Time (CEST) ▼

[CONTINUE](#)

- Configure the job's target

- Configure advanced settings (optional)

1. Click on CREATE JOB to create a new schedule
2. Choose where the Cloud Scheduler jobs live
3. Put name, description, frequency and timezone (in our case it runs every day at 20 PM)

[CREATE](#)

[CANCEL](#)

The workflow – Google Cloud Scheduler



Cloud Scheduler [← Create a job](#)

- ✓ Define the job
 - Configure the job's target
 - Target type *
Pub/Sub
 - Select a Cloud Pub/Sub topic *
projects/covid-19-analysis-318015/topics/region1
 - Message body *
Updated!
Message size cannot exceed 10MB.
 - Message attributes
[+ ADD AN ATTRIBUTE](#)
 - [CONTINUE](#)
 - Configure advanced settings (optional)

[CREATE](#) [CANCEL](#)

1. Click on CREATE JOB to create a new schedule
2. Choose where the Cloud Scheduler jobs live
3. Put name, description, frequency and timezone (in our case it runs every day at 20 PM)
4. Select Pub/Sub as target
5. Select the topic previously created
6. Insert the Message Body

The workflow – Google Cloud Scheduler



 Cloud Scheduler [← Create a job](#)

- ✓ Define the job
- ✓ Configure the job's target
- Configure advanced settings (optional)

Retry config

If a job does not complete successfully, it is retried, with exponential backoff, according to the settings in retry config. [Learn more](#)

Max retry attempts

Maximum number of retry attempts for a failed job

Max retry duration *

Time limit for retrying a failed job, 0s means unlimited

Min backoff duration * Max backoff duration *

Minimum time to wait before retrying a job after it fails

Maximum time to wait before retrying a job after it fails

Max doublings *

The time between retries will double max doublings times

CREATE

CANCEL

1. Click on CREATE JOB to create a new schedule
2. Choose where the Cloud Scheduler jobs live
3. Put name, description, frequency and timezone (in our case it runs every day at 20 PM)
4. Select Pub/Sub as target
5. Select the topic previously created
6. Insert the Message Body
7. Insert how many attempts and max retry durations
8. CREATE the schedule

The workflow – Google Cloud Scheduler



<div>Cloud Scheduler</div> <div>Jobs</div> <div>+ CREATE JOB</div> <div>↻ REFRESH</div> <div>✎ EDIT</div> <div>⏸ PAUSE</div> <div>▶ RESUME</div> <div>🗑 DELETE</div>									
<div>SCHEDULER JOBS</div> <div>APP ENGINE CRON JOBS ?</div>									
<div>≡ Filter</div> <div>Filter jobs</div>									
<input type="checkbox"/>	Name ↑	State	Description	Frequency	Target	Last run	Result	Logs	Run
<input type="checkbox"/>	update_regioni	Enabled	Update the GitHub repository of Covid19 regional data	0 20 * * * (Europe/Rome)	Topic : projects/covid-19-analysis-318015/topics/regioni		Has not run yet	View	RUN NOW

The job has been scheduled but not run yet.
Click on RUN NOW to execute the job!

The workflow – Google Cloud Scheduler



Cloud Scheduler

Jobs

CREATE JOB

REFRESH

EDIT

PAUSE

RESUME

DELETE

SCHEDULER JOBS

APP ENGINE CRON JOBS ?

Filter

Filter jobs

<input type="checkbox"/>	Name	State	Description	Frequency	Target	Last run	Result	Logs	Run
<input type="checkbox"/>	update_regioni	Enabled	Update the GitHub repository of Covid19 regional data	0 20 * * * (Europe/Rome)	Topic : projects/covid-19-analysis-318015/topics/regioni	Jun 29, 2021, 9:12:33 PM	Success	View	RUN NOW

The job has been scheduled correctly and executed successfully!

The workflow – Google Cloud Scheduler



Cloud Scheduler

Jobs

CREATE JOB

REFRESH

EDIT

PAUSE

RESUME

DELETE

SCHEDULER JOBS

APP ENGINE CRON JOBS

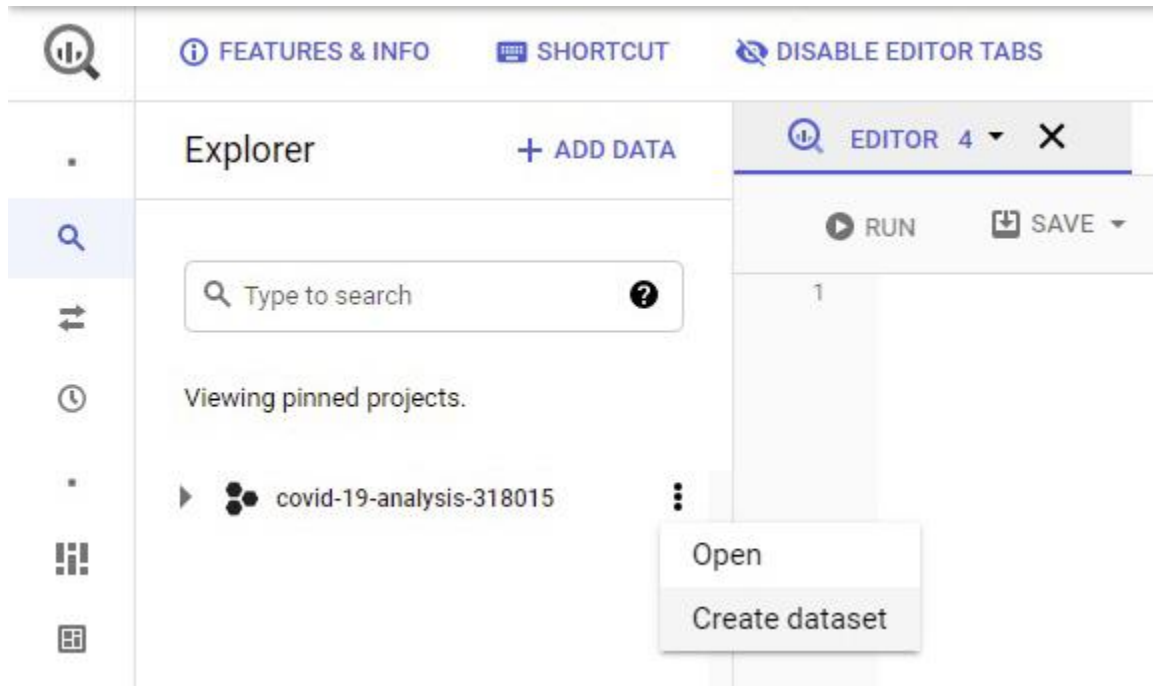
Filter

Filter jobs

<input type="checkbox"/>	Name	State	Description	Frequency	Target	Last run	Result	Logs	Run
<input type="checkbox"/>	update_province	Enabled	Update the GitHub repository of Covid19 province data	0 20 * * * (Europe/Rome)	Topic : projects/covid-19-analysis-318015/topics/province	Jun 29, 2021, 9:13:43 PM	Success	View	RUN NOW
<input type="checkbox"/>	update_regioni	Enabled	Update the GitHub repository of Covid19 regional data	0 20 * * * (Europe/Rome)	Topic : projects/covid-19-analysis-318015/topics/regioni	Jun 29, 2021, 9:12:33 PM	Success	View	RUN NOW
<input type="checkbox"/>	update_vaccini	Enabled	Update the GitHub repository of Covid19 vaccines data	0 20 * * * (Europe/Rome)	Topic : projects/covid-19-analysis-318015/topics/vaccini	Jun 29, 2021, 9:15:01 PM	Success	View	RUN NOW

Repeat for the other schedules!

The workflow – Google BigQuery



1. In Google BigQuery we select the project and Create dataset

The workflow – Google BigQuery



Create dataset

Dataset ID *

Letters, numbers, and underscores allowed

Data location

Default table expiration

☐ Enable table expiration

Default maximum table age

Encryption

- ☒ Google-managed encryption key
No configuration required
- ☐ Customer-managed encryption key (CMEK)
Manage via Google Cloud Key Management Service

1. In Google BigQuery we select the project and Create dataset
2. Insert the id for Dataset ID and we choose "dataset"
3. Click on CREATE DATASET to create the dataset

CREATE DATASET

CANCEL

The workflow – Google BigQuery



Explorer + ADD DATA

EDITOR 4 X DATASET X

covid-19-analysis-318015:dataset + CREATE TABLE

Description None Labels None

Dataset info

Dataset ID	covid-19-analysis-318015:dataset
Created	Jun 26, 2021, 5:09:29 PM
Default table expiration	Never
Last modified	Jun 26, 2021, 5:09:29 PM
Data location	US

1. CREATE TABLE to create a table in the dataset

The workflow – Google BigQuery



Create table

Source

Create table from:

Empty table

- Empty table
- Google Cloud Storage
- Upload
- Drive
- Google Cloud Bigtable

Destination

☒ Search for a project name

Project name

covid-19-analysis

Dataset name

dataset

Table type ?

Native table

Table name

Letters, numbers, and underscores allowed

1. CREATE TABLE to create a table in the dataset
2. As source we use Google Cloud Storage

The workflow – Google BigQuery



Create table

Source

Create table from:

Google Cloud Storage

Select file from GCS bucket: ?

bucket/folder/file

Browse

File format:

Avro

☐ Source Data Partitioning

Destination

☒ Search for a project

☐ Enter a project name

Project name

covid-19-analysis

Dataset name

dataset

Table type ?

Native table

Table name

Letters, numbers, and underscores allowed

1. CREATE TABLE to create a table in the dataset
2. As source we use Google Cloud Storage and Browse the file in the bucket

The workflow – Google BigQuery



Choose a file

Name

New file name

Location

← Buckets

- covid-19-analysis-318015.appspot.com >
- downloaded_data >
- gcf-sources-941570791529-us-central1 >
- staging.covid-19-analysis-318015.appsp... >
- us.artifacts.covid-19-analysis-318015.ap... >

Select

Cancel

1. CREATE TABLE to create a table in the dataset
2. As source we use Google Cloud Storage and Browse the file in the bucket
3. We select the Bucket

The workflow – Google BigQuery



Choose a file

Name

Location

← downloaded_data 🔍

- dpc-covid19-ita-province.csv
- dpc-covid19-ita-regioni.csv
- somministrazioni-vaccini-latest.csv**

Select Cancel

1. CREATE TABLE to create a table in the dataset
2. As source we use Google Cloud Storage and Browse the file in the bucket
3. We select the Bucket and the corresponding file

The workflow – Google BigQuery



Source

Create table from: Google Cloud Storage Select file from GCS bucket: downloaded_data/somministrazioni-vaccini-latest.c File format: CSV

☐ Source Data Partitioning

Destination

☒ Search for a project ☐ Enter a project name

Project name: covid-19-analysis Dataset name: dataset Table type: Native table

Table name: vaccine

Schema

Auto detect ☒ Schema and input parameters

Schema will be automatically generated.

Create table Cancel

1. CREATE TABLE to create a table in the dataset
2. As source we use Google Cloud Storage and Browse the file in the bucket
3. We select the Bucket and the corresponding file
4. Insert the table name and auto-detect the schema
5. Click on CREATE TABLE to create the table
6. Repeat the previous steps for the other files

The workflow – Google BigQuery



Explorer [+ ADD DATA](#)

🔍 Type to search ?

Viewing pinned projects.

- 📁 covid-19-analysis-318015
 - 📁 dataset
 - 📄 province
 - 📄 regioni
 - 📄 vaccine

vacchine

[SCHEMA](#) [DETAILS](#) [PREVIEW](#)

[QUERY](#) [SHARE](#) [COPY](#) [DELETE](#) [EXPORT](#) [+ COMPOSE NEW QUERY](#)

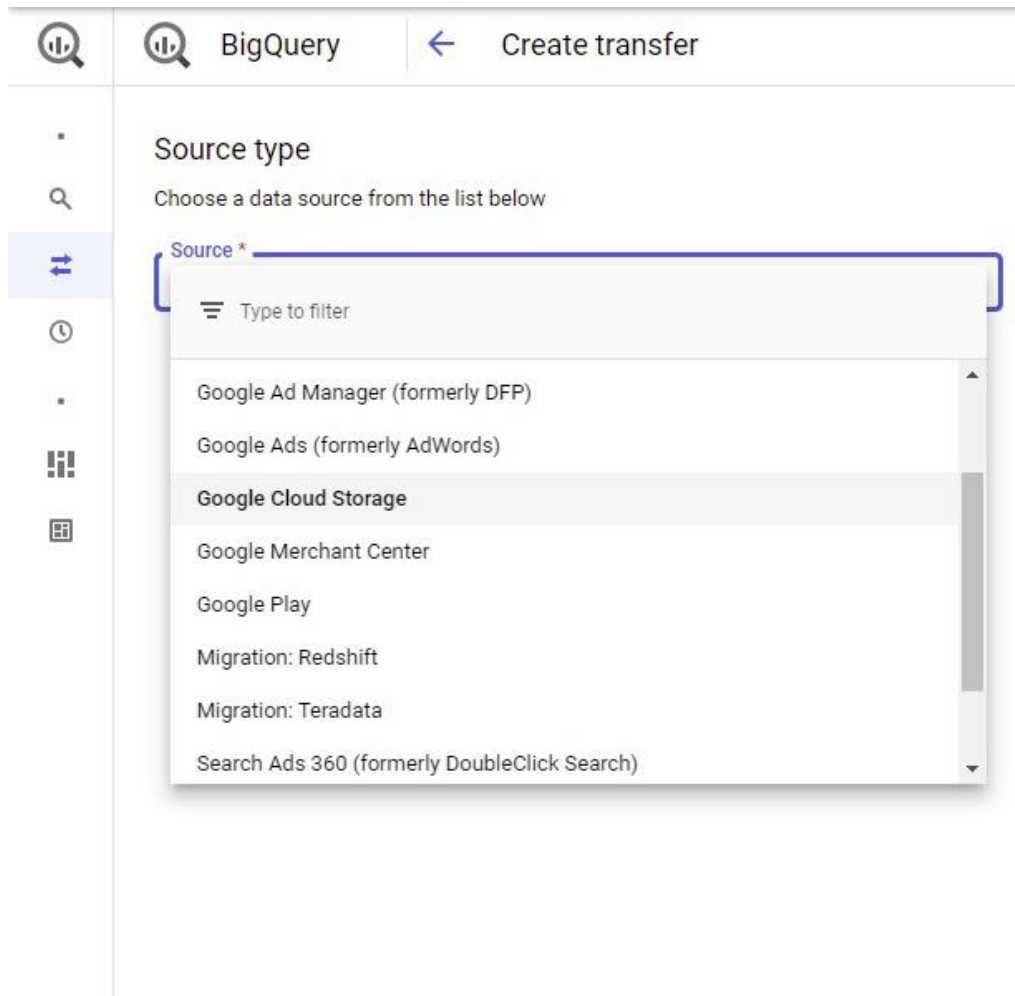
Table schema

Filter Enter property name or value ?

Field name	Type	Mode	Policy Tags ?	Description
data_somministrazione	DATE	NULLABLE		
fornitore	STRING	NULLABLE		
area	STRING	NULLABLE		
fascia_anagrafica	STRING	NULLABLE		

Tables have been created correctly!

The workflow – Google BigQuery (Data transfer)



1. Select the type of Data Transfer (in our case Google Cloud Storage)

The workflow – Google BigQuery (Data transfer)



Source type

Choose a data source from the list below

Source *
Google Cloud Storage

i This is the Google Cloud Storage configuration. [Learn more](#)

Transfer config name

Display name *
transfer_vaccine

Schedule options

☐ Start now ☒ Start at set time

Repeats *
Daily

Start date and run time *
6/30/21, 8:15 PM CEST

i Every day at 20:15:00 Europe/Paris

1. Select the type of Data Transfer (in our case Google Cloud Storage)
2. Insert the transfer name and schedule options

The workflow – Google BigQuery (Data transfer)



Destination settings

Select the destination for the transfer data

Dataset ID *

CREATE NEW DATASET
Launches a form to create a new dataset before continuing

Loaded datasets

dataset

Cloud Storage URI * BROWSE ?

Write preference
APPEND ▼ ?

☐ Delete source files after transfer ?

File format *
CSV ▼ ?

1. Select the type of Data Transfer (in our case Google Cloud Storage)
2. Insert the name and schedule options
3. Select the dataset ID

The workflow – Google BigQuery (Data transfer)



Data source details

Destination table *
data ?

Cloud Storage URI * BROWSE ?
Field Cloud Storage URI is required

Write preference
APPEND ▼ ?

☐ Delete source files after transfer ?

File format *
CSV ▼ ?

1. Select the type of Data Transfer (in our case Google Cloud Storage)
2. Insert the name and schedule options
3. Select the destination dataset ID
4. Select the destination table
5. Select the Google Cloud Storage URI

The workflow – Google BigQuery (Data transfer)



Select object

< Buckets ▾

🗑️ 🔍

- 🗑️ covid-19-analysis-318015.appspot.com >
- 🗑️ downloaded_data >
- 🗑️ gcf-sources-941570791529-us-central1 >
- 🗑️ staging.covid-19-analysis-318015.appspot.com >
- 🗑️ us.artifacts.covid-19-analysis-318015.appspot.com >

Filename

SELECT

CANCEL

1. Select the type of Data Transfer (in our case Google Cloud Storage)
2. Insert the name and schedule options
3. Select the destination dataset ID
4. Select the destination table
5. Select the Google Cloud Storage URI
6. Select the Bucket

The workflow – Google BigQuery (Data transfer)



1. Select the type of Data Transfer (in our case Google Cloud Storage)
2. Insert the name and schedule options
3. Select the destination dataset ID
4. Select the destination table
5. Select the Google Cloud Storage URI
6. Select the Bucket and the file

The workflow – Google BigQuery (Data transfer)



Data source details

Destination table *
data ?

Cloud Storage URI *
☒ downloaded_data/somministrazioni-vaccini-latest.csv BROWSE ?

Write preference
MIRROR ?

☐ Delete source files after transfer ?

File format *
CSV ?

Transfer Options

All Formats

Number of errors allowed
0 ?

Decimal target types ?

1. Select the type of Data Transfer (in our case Google Cloud Storage)
2. Insert the name and schedule options
3. Select the destination dataset ID
4. Select the destination table
5. Select the Google Cloud Storage URI
6. Select the Bucket and the file
7. To overwrite the entire table select MIRROR

The workflow – Google BigQuery (Data transfer)



Transfer Options

CSV

Field delimiter ?

Header rows to skip ?

☐ Allow quoted newlines ?

☐ Allow jagged rows ?

Notification options

☐ Email notifications
When enabled, the transfer administrator will receive e-mail notifications on transfer run failures.

1. Select the type of Data Transfer (in our case Google Cloud Storage)
2. Insert the name and schedule options
3. Select the destination dataset ID
4. Select the destination table
5. Select the Google Cloud Storage URI
6. Select the Bucket and the file
7. To overwrite the entire table select MIRROR
8. Insert the delimiter character and skip the first header row
9. Click on SAVE

The workflow – Google BigQuery (Data transfer)



BigQuery

Data transfers

[+ CREATE TRANSFER](#)

Filter

Filter transfer configs

	Display name	Source	Schedule (UTC)	Region	Destination dataset	Next scheduled	UTC+2 ▾		Actions
	transfer_vaccine	Google Cloud Storage	every day 18:15	us	dataset	June 30, 2021 at 8:15:00 PM UTC+2			

The data transfer from the file in the bucket to the table in the dataset has been created correctly and scheduled!
Repeat again the previous steps for the other data transfers!

BigQuery

Data transfers

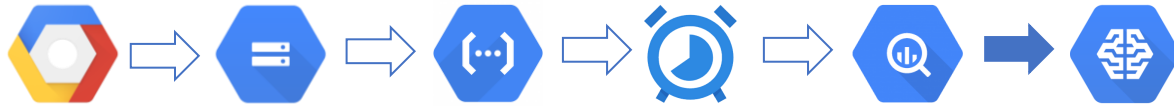
[+ CREATE TRANSFER](#)


Filter







Filter transfer configs

	Display name	Source	Schedule (UTC)	Region	Destination dataset	Next scheduled	UTC+2 ▾	↑	Actions
	transfer_vaccine	Google Cloud Storage	every day 18:15	us	dataset	June 30, 2021 at 8:15:00 PM UTC+2			
	transfer_province	Google Cloud Storage	every day 18:15	us	dataset	June 30, 2021 at 8:15:00 PM UTC+2			
	transfer_regioni	Google Cloud Storage	every day 18:15	us	dataset	June 30, 2021 at 8:15:00 PM UTC+2			


The workflow – Google AI Platform (Notebooks)



 **Notebooks** [+ NEW INSTANCE](#) [REFRESH](#) [▶ START](#) [■ STOP](#) [⏻ RESET](#) [🗑 DELETE](#)

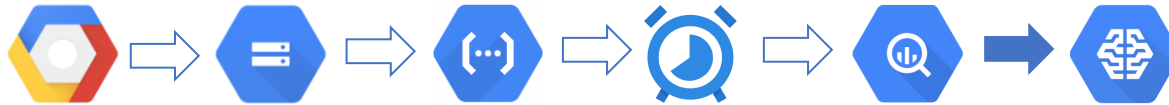
Create and use Jupyter Notebooks with a notebook instance. Notebook instances have JupyterLab pre-installed and are configured with GPU-enabled machine learning frameworks. [Learn more](#)


 **Filter** Enter property name or value







<input type="checkbox"/>	<input checked="" type="radio"/>	Instance name ↑	Zone	Environment Version	Auto-upgrade
--------------------------	----------------------------------	-----------------	------	---------------------	--------------

1. Create a NEW INSTANCE

The workflow – Google AI Platform (Notebooks)



 < Create a notebook instance

Instance name *

covid-19-analysis-vm


63-char limit with lowercase letters, digits, or '-' only. Must start with a letter. Cannot end with a '-'.

Region *

us-central1 (Iowa) ▼ ?

Zone *

us-central1-a ▼ ?



Requests to your instance from the Datalab/Jupyter interface may be routed through a different region than selected above depending on service availability.

Environment ▼

Security ▼

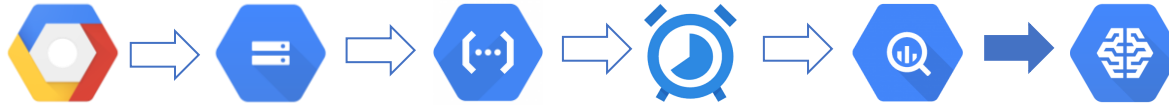
Environment upgrade ▼

CREATE

CANCEL

1. Create a NEW INSTANCE
2. Insert Instance Name
3. Leave all parameters to Default
4. CREATE

The workflow – Google AI Platform (Notebooks)



Notebooks [+ NEW INSTANCE](#) [REFRESH](#) [▶ START](#) [■ STOP](#) [⏻ RESET](#) [🗑 DELETE](#) [SHOW INFO PANEL](#)

Create and use Jupyter Notebooks with a notebook instance. Notebook instances have JupyterLab pre-installed and are configured with GPU-enabled machine learning frameworks. [Learn more](#)

[Filter](#) Enter property name or value

<input type="checkbox"/>	<input checked="" type="checkbox"/>	Instance name ↑		Zone	Environment Version	Auto-upgrade	Environment	Machine type	GPUs	Permission	Labels
<input type="checkbox"/>	<input checked="" type="checkbox"/>	covid-19-analysis-vm	OPEN JUPYTERLAB	us-central1-a	M74	—	TensorFlow:2.3	4 vCPUs, 15 GB RAM ▾	None ▾	Service account	goog-caip-notebook

The virtual machine has been created correctly and we can open JupyterLab!