

Laboratorijska vežba 4

Cilj vežbe: Učitavanje podataka iz različitih izvora podataka u Google BigQuery i generisanje izveštaja upotrebom PowerBI aplikacije

U okviru ove vežbe biće prikazan primer učitavanje podataka u Google BigQuery korišćenjem aplikacije pisane u Python-u. Učitavaju se podaci koji predstavljaju informacije o platenicima: ime platenika, informacija o temperaturi u plateniku, vlažnost zemljišta i vlažnost vazduha, kao i vreme čitanja informacija sa senzora. Takođe, biće prikazan primer učitavanja podataka iz csv fajla u BigQuery, i to informacije o platenicima i mernim mestima. Nad kreiranim tabelama biće prikazan primer izvršavanja upita korišćenjem SQL-a u BigQuery-u i kreiranje izveštaja korišćenjem PowerBI aplikacije.

Podaci za vežbu

- gardens.csv – sadrži podatke o platenicima: jedinstveni identifikator, naziv, lokaciju i veličinu
- measuring_points.csv – sadrži merna mesta u platenicima u okviru kojih se meri temperatura vazduha, vlažnost vazduha i vlažnost zemljišta. Za svako merno mesto pamti se: jedinstveni identifikator, naziv i pozicija
- readings.csv – sadrži čitanja sa senzora. Svako čitanje ima podatke o jedinstvenom identifikatoru platenika, jedinstvenom identifikatoru mernog mesta, vremenu čitanja, temperaturi, vlažnosti vazduha i vlažnosti zemljišta.

Potrebni alati i instalacije

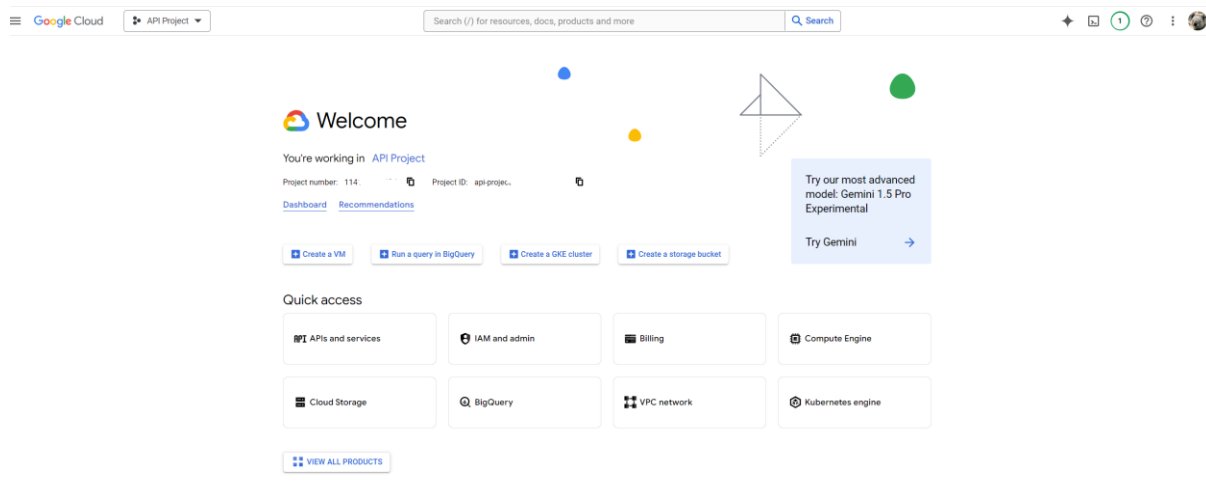
- Google Cloud sandbox
- Python
- PowerBI aplikacija

Google BigQuery

Google [BigQuery](#) je potpuno upravljani data warehouse. Osmišljen je da radi sa velikom količinom podataka, omogućava učitavanje podataka iz različitih izvora, brzu obradu podataka, pisanje SQL upita nad podacima i mnoge druge mogućnosti.

Kreiranje projekta u Google Cloud sandbox

Otvoriti stranicu <https://console.cloud.google.com/> i prijaviti se sa Google nalogom.

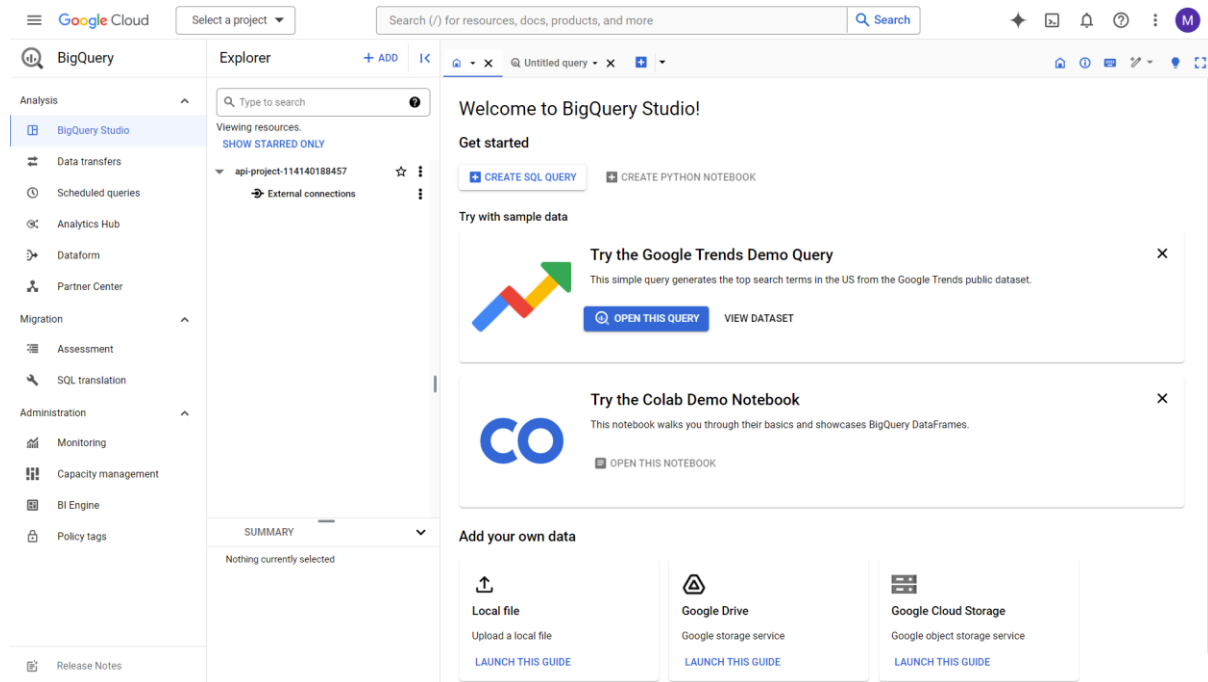


Otvoriti API Project i izabrati NEW Project. Otvoriće se forma za kreiranje novog projekta.

U polje Project name uneti ime projekta: ipz-vezba4 i odabrati Create.

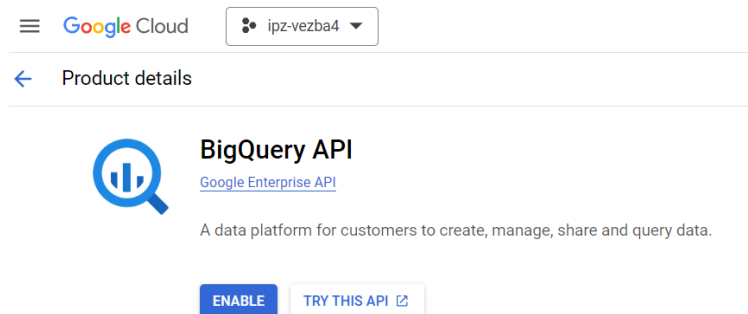
The image shows the 'New Project' form in Google Cloud. At the top, there's a warning message: 'You have 8 projects remaining in your quota. Request an increase or delete projects. Learn more'. Below this is a 'Project name' field with the value 'ipz-vezba4' and a help icon. Underneath, it shows the 'Project ID: ipz-vezba4-421000. It cannot be changed later. EDIT'. The 'Location' field is set to 'No organisation' with a 'BROWSE' button. At the bottom, there are 'CREATE' and 'CANCEL' buttons.

Odabrati projekat ipz-vezba4, čime se otvaraju sva podešavanja za ovaj projekat.



Omogućavanje upotrebe BigQuery API-a

- Otvoriti glavni meni
- Odabrati APIs & Services -> Enable APIs and services -> Enable APIs and services.
- Otvoriće se API Library u okviru koje je potrebno pronaći i omogućiti BigQuery API.



Kreiranje kredencijala za pristup BigQuery API-u

- Otvoriti glavni meni i odabrati IAM & Admin -> Service accounts -> + Create service account.
- U prvom koraku u polje *Service account name* uneti ime, npr. ipz_lab4. Na osnovu imena biće kreiran *service account ID* i *email* adresa. Odabrati *Create and continue*.

← Create service account

- Service account details**

Service account name
ipz_lab4

Display name for this service account

Service account ID *
ipz-lab4

Email address: ipz-lab4@ipz-vezba4-421011.iam.gserviceaccount.com

Service account description
Describe what this service account will do

CREATE AND CONTINUE
- Grant this service account access to project (optional)**
- Grant users access to this service account (optional)**

DONE CANCEL

- U drugom koraku dodeljuju se privilegije. Odabrati privilegije Basic -> Owner. Odabrati *Continue* i nakon toga *Done*.

- Service account details**
- Grant this service account access to project (optional)**

Grant this service account access to ipz-vezba4 so that it has permission to complete specific actions on the resources in your project. [Learn more](#)

Role: Owner

IAM condition (optional) + ADD IAM CONDITION

Full access to most Google Cloud resources. See the list of included permissions.

+ ADD ANOTHER ROLE

CONTINUE
- Grant users access to this service account (optional)**

DONE CANCEL

- U listi svih *Service account*-a za projekat ipz-vezba4 nalaziće se novi nalog.

Service accounts + CREATE SERVICE ACCOUNT DELETE MANAGE ACCESS REFRESH LEARN

Service accounts for project "ipz-vezba4"

A service account represents a Google Cloud service identity, such as code running on Compute Engine VMs, App Engine apps, or systems running outside Google. [Learn more about service accounts.](#)


Organization policies can be used to secure service accounts and block risky service account features, such as automatic IAM Grants, key creation/upload, or the creation of service accounts entirely. [Learn more about service account organization policies.](#)

Filter	Enter property name or value								
	Email	Status	Name	Description	Key ID	Key creation date	OAuth 2 Client ID		Actions
<input type="checkbox"/>	ipz-lab4@ipz-vezba4-421011.iam.gserviceaccount.com	Enabled	ipz_lab4		No keys		10102416480067136432		

Za kreirani nalog, potrebno je kreirati ključeve.

- U okviru kolone *Action* odabrati Manage Keys -> Add Key -> Create new key -> odabrati JSON i Create, nakon čega će biti preuzet json file u kome se nalaze kredencijali za pristup.




Keys

 Service account keys could pose a security risk if compromised. We recommend you avoid downloading service account keys and instead use the [Workload Identity Federation](#). You can learn more about the best way to authenticate service accounts on Google Cloud [here](#).

Add a new key pair or upload a public key certificate from an existing key pair.

Block service account key creation using [organization policies](#).
[Learn more about setting organization policies for service accounts](#)

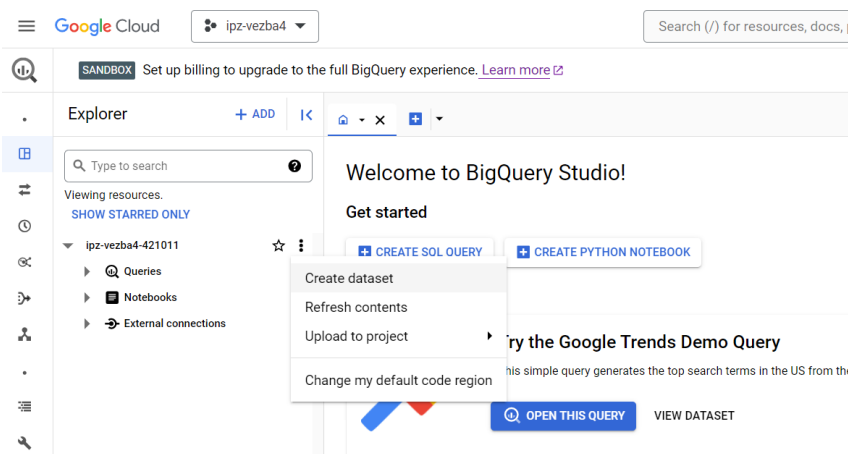
ADD KEY ▾

Type	Status	Key	Key creation date	Key expiration date	
	 Active	16d68d891b512bbe03847332fc578be6484552f6	Apr 21, 2024	Jan 1, 10000	

Učitavanje podataka u Google BigQuery iz CSV fajla

Otvoriti BigQuery Studio i kreirati novi dataset u okviru koga će biti upamćeni podaci koje želimo da učitamo.

U okviru Explorer dala, otvoriti akcije dostupne za projekat i odabrati opciju *Create dataset*.



Otvoriće se forma za kreiranje nogov seta podataka.

Ukoliko polje *Project ID* ne ukazuje na željeni projekat klikom na dugme *Change* izabrati projekat ipz_vezba4. U polje *Dataset ID* uneti ime seta podataka, npr. garden_readings. U okviru podešavanja *Default table expiration* moguće je postaviti automatski brisanje tabela nakon određenog broja dana u odnosu na trenutak njihovog kreiranja. Kliknuti na *Create dataset*.

Nakon ove akcije kreiran je set podataka u okviru koga može da se kreira više tabela.

Create dataset

Project ID *
ipz-vezba4-421011 [CHANGE](#)

Dataset ID *
garden_readings
Letters, numbers, and underscores allowed

Location type [?](#)
☐ Region
Specify a region to colocate your datasets with other Google Cloud services.
☒ Multi-region
Allow BigQuery to select a region within a group to achieve higher quota limits.

Multi-region *
US (multiple regions in United States) ▼

Default table expiration
☐ Enable table expiration [?](#)
Default maximum table age Days

Advanced options ▼

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

Dodavanje podatka o platenicima

Dodavanje nove tabele sa podacima iz CSV fajla radi se klikom na dugme **ADD**, čime se otvara forma za izbor izvora podataka. Odabrati *Local file* opciju čime se otvara forma za kreiranje tabele.

- U okviru polja **Create table from** odabrati opciju *Upload*,
- U okviru polja **Select file**, klikom na dugme **Browse** odabrati fajl gardens.csv,
- U **Destination** dalu odabrati projekat *ipz_vezba4* i kreirani set podataka *garden_readings*
- U polje **Table** uneti ime tabele, npr. *gardens*,
- Odabrati **Schema Auto detect** čime će struktura šeme biti automatski generisana na osnovu podataka,
- Kliknuti na **Create table**.

Create table [×](#)

Source

Create table from
Upload ▼

Select file *
gardens.csv [×](#) [BROWSE](#) [?](#)

File format
CSV ▼

Destination

Project *
ipz-vezba4-421011 [BROWSE](#)

Dataset *
garden_readings

Table *
gardens
Maximum name size is 1,024 UTF-8 bytes. Unicode letters, marks, numbers, connectors, dashes, and spaces are allowed.

Table type
Native table ▼ [?](#)

Schema

☒ Auto detect

[?](#) Schema will be automatically generated.

Partition and cluster settings

Partitioning
No partitioning ▼ [?](#)

Clustering order [?](#)
Clustering order determines the sort order of the data. Clustering can be used on both partitioned and non-partitioned tables.

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

Dodavanje podatka o mernim mestima

Ponoviti postupak iz prethodnog koraka i učitati fajl *measuring_points.csv* u tabelu *measuring_points*.

Učitavanje podataka u Google BigQuery iz Python aplikacije

Kreirati novu tabelu u kojoj će biti zapamćeni podaci o merenjima. U opcijama seta podataka *garden_readings* odabrati opciju **Create table**. Otvara se forma za kreiranje nove tabele:

- U okviru polja **Create table from** odabrati opciju *Empty table*,
- U okviru **Destination** podešavanja odabrati projekat *ipz-vezba4* i kreirani set podataka *garden_readings*,
- U polje **Table** uneti ime tabele, npr. *garden_readings*,
- U okviru **Schema** dela dodati kolone i njihove tipove podataka kao na slici u nastavku.

Destination

Project *	ipz-vezba4-421011	BROWSE
Dataset *	garden_readings	
Table *	garden_readings	
<small>Maximum name size is 1,024 UTF-8 bytes. Unicode letters, marks, numbers, connectors, dashes, and spaces are allowed.</small>		
Table type	Native table	

Schema

☒ Edit as text

1	Field name *	Type *	Mode	Description
	garden_id	INTEGER	NULLABLE	
2	Field name *	Type *	Mode	Description
	measuring_point_id	INTEGER	NULLABLE	
3	Field name *	Type *	Mode	Description
	timestamp	TIMESTAMP	NULLABLE	
4	Field name *	Type *	Mode	Description
	temperature	FLOAT	NULLABLE	
5	Field name *	Type *	Mode	Description
	air_humidity	INTEGER	NULLABLE	
6	Field name *	Type *	Mode	Description
	soil_humidity	INTEGER	NULLABLE	
7	+			

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

Potrebno je instalirati Google Cloud BigQuery biblioteku: `pip install google-cloud-bigquery`.

Kod za učitavanje podataka u tabelu *readings*:

```
import os

from google.cloud import bigquery
import pandas as pd

os.environ['GOOGLE_APPLICATION_CREDENTIALS'] = 'ipz-vezba4-421011-16d68d891b51.json'

client = bigquery.Client()

project_id = 'ipz-vezba4-421011'
dataset_id = 'ipz-vezba4-421011.garden_readings'
table_id = "ipz-vezba4-421011.garden_readings.readings"
```

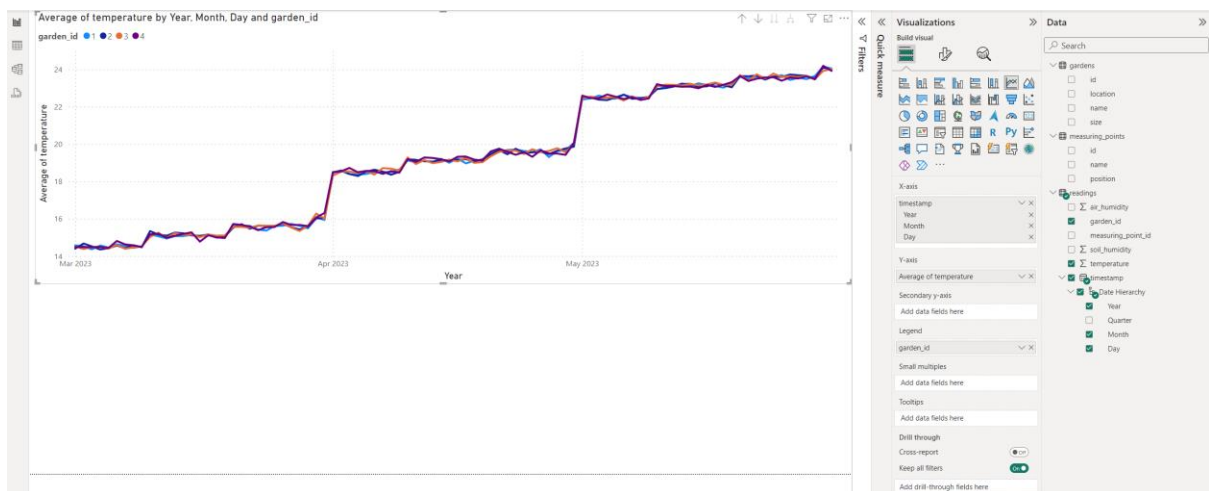
```
df = pd.read_csv('garden_readings.csv')
df['timestamp'] = pd.to_datetime(df['timestamp'], format='%Y-%m-%d %H:%M:%S')
```

```
job_config = bigquery.LoadJobConfig()
job_config.write_disposition = bigquery.WriteDisposition.WRITE_TRUNCATE
```

```
client.load_table_from_dataframe(df, table_id, job_config=job_config)
print(f"Data loaded into {project_id}.{dataset_id}.{table_id}")
```

Kreiranje izveštaja korišćenjem PowerBI aplikacije

1. Pokrenuti Power BI aplikaciju.
2. U okviru Home kartice kliknuti na dugme Get data.
3. Selektovati Google BigQuery i kliknuti na connect.
4. Odabrati Organizational account i prijaviti se.
5. Učitati sve tabele.
6. Kreirati izveštaj koji prikazuje prosečnu temperaturu po platenicima.



7. Kreirati izveštaj koji za sve lokacije platenika i nazive mernih mesta prikazuje prosečnu vlažnost zemljišta.

