

Database

It's never a good idea to save blobs in any database, rather we prefer to save only the reference of blobs in a database, same approach is taken here, the actual data will be stored in google bucket but will be handled using bigQuery database.

The data folder had the entire data set, that is already uploaded to cloud and to save space is now deleted, whereas the data_sample contains a chunk of it (80 data points per class) that can be used to quickly check the working of code.

This code is not moved to docker container as we will always upload the data from host machine not from a container. Multithreaded approach is used to reduce the time of upload.

Overview of Data, The database follows the following structure.

Class	Datapoints
bed	1713
bird	1731
cat	1733
dog	1746
down	2359
eight	2352
five	2357
four	2372
go	2372
happy	1742
house	1750
left	2353
marvin	1746
nine	2364
no	2375
off	2357
on	2367
one	2370
right	2367
seven	2377
sheila	1734
six	2369
stop	2380
three	2356
tree	1733
two	2373
up	2375
wow	1745
yes	2377
zero	2376
Total	64721

Database Schema

The database schema is simple enough, and has only 1 table named as sounds, which has following three columns:

- type : the class of data point
- name : the name of data point
- location : the complete location of data point in the actual bucket where it is stored