**Dataproc:**

**Target 1: Move data into**

**Steps 1: Data ingestion**

**Part A: get data via API and save in GCS bucket**

* candidate\_master
* candidate\_committee\_linkage
* committee\_master
* pac\_summary
* contribution\_from\_committe\_to\_candiate
* any\_trans\_from\_one\_committee\_to\_other
* operating\_expenditure
* electronically\_filed\_report
* paper\_filed\_report

How to ingest data in GCS bucket?

Requirement:

GCS account:

Creds to connect via VS code:

create folder for each file name under raw and within each file name folder create two sub folder src\_zip\_files and src\_text\_files.

**Part A: get data into HDFS**

* individual\_contribution

Part c: get data via filess.io and save in gcs bucket

* all\_candidates
* senate\_current\_campaigns

How to ingest data in HDFS?