BLM DQIMP QAQC BigQuery Implementation Action Timeline

# 05/07/2025

* Got new snapshot data, began extracting for upload to Google Cloud
  + Completed extract
* Loaded all Feb snapshot data into new vintage snapshot folder 20250201
  + Converted nlsdbs to parquets in format acceptable to BQ
  + Got snapshot pieces consolidated
  + Loaded feb snapshot data into BQ

# 05/05/2025

* Removed NLSDB case/case\_lands parquets from April snapshot folder. Discussed with Huy and decided spatially enabling data avoid for now.
* ~~Loaded in csv version of nlsdb into each snapshot folder~~
* March snapshot folder needs consolidating still.
* Opened cmd as admin and cd’d to schema folder
  + Sourced the load\_commands\_march.txt file for load scripts
* Confirmed google setup still correct with:
  + gcloud auth list
    - A black screen with white text

      AI-generated content may be incorrect.
* Confirmed the correct project is active:
  + gcloud config get-value project
    - A black background with white text

      AI-generated content may be incorrect.
* Ran the combine and delete of pieces for march snapshots
* Had a big learning curve on loading nlsdb. Process is to extract from FileGDB as parquet, then use pandas to load in a Notebook and drop the Shape column then upload that to cloud storage
* Converted the first SYT\_Basic test into a SQL query and saved as SYT01
  + Learned you have to run a query in same region of BigQuery project which in this case is us-central1
* Built the following tests and saved as queries:
  + SYT
    - 1,3,4,5

# 05/01/2025

* Elii setup an updated sandbox file structure for snapshots
  + Sandbox-blm-dqimp-qaqc/snapshots/2050401
* All CSV files have been moved into the subfolder
  + gsutil mv gs://sandbox-blm-seta-dqimp-qaqc/\*.csv gs://sandbox-blm-seta-dqimp-qaqc/snapshots/20250401/
* Elii created a .json file with the MC\_CASE\_ACTION schema:
* Elii used gsutil compose to combine MC\_CASE\_ACTION.csv and CR\_FULL\_CASE\_ACTION.csv into a single CSV titled CASE\_ACTION.csv
  + gsutil compose gs://sandbox-blm-seta-dqimp-qaqc/snapshots/20250401/MC\_CASE\_ACTION.csv gs://sandbox-blm-seta-dqimp-qaqc/snapshots/20250401/CR\_FULL\_CASE\_ACTION.csv gs://sandbox-blm-seta-dqimp-qaqc/snapshots/20250401/CASE\_ACTION.csv
* created a case\_action\_schema.json to load into bigquery dataset
* loaded combined CASE\_ACTION.csv with a schema as blm\_dqimp\_qaqc.CASE\_ACTION\_20250401
  + bq load --source\_format=CSV --field\_delimiter="|" --skip\_leading\_rows=0 --schema=case\_action\_schema.json blm\_dqimp\_qaqc.case\_action\_20250401 gs://sandbox-blm-seta-dqimp-qaqc/snapshots/20250401/CASE\_ACTION.csv
    - date formats non-comformant to bigquery; loading all as string to CAST later
* created schemas for all tables and loaded all tables for April snapshot
* created new snapshot subfolder for march data titled 20250301/
* loaded all CSVs to this subfolder
* created load scripts for these tables to bigquery
* ~~loaded nlsdb case and case\_land feature classes as parquet files with WKT geometry column~~
  + ~~will have to CAST to GEOMETRY column compatible with biquery later for spatial analyses~~

# 04/30/2025

* Huy and Elii met and Huy showed Elii how to load a CSV from sandbox into BigQuery dataset
* Huy showed how snapshots can be handled with low LOE via date suffix naming convention

# Setup

* Huy created a sandbox for this data: <https://console.cloud.google.com/storage/browser/sandbox-blm-seta-dqimp-qaqc;tab=objects?prefix=&forceOnObjectsSortingFiltering=false>
* Elii loaded all CSVs for one snapshot into that sandbox

# Notes

huy suggested using DataForm instead of dbt..this is under the bigquery left sidebar