MIDI Analysis Script and Database

Contents

[Code Location 1](#_Toc123636201)

[Requirements 1](#_Toc123636202)

[Building the Project 1](#_Toc123636203)

[Build Product 1](#_Toc123636204)

[Preparing the Run Environment 2](#_Toc123636205)

[Mapping to Buckets 2](#_Toc123636206)

[Run the Script Named RunProcess 2](#_Toc123636207)

[Parameters 2](#_Toc123636208)

[Load to BigQuery 2](#_Toc123636209)

## Code Location

<https://github.com/CBIIT/MIDI>

## Requirements

* Java
* Maven 3.8.5

## Building the Project

mvn clean compile assembly:single

### Build Product

A single jar is created in the target directory. This jar includes all dependencies. This jar is later renamed to *midi-analysis.jar*.

Directory of C:\MIDI2\target

12/28/2022 05:44 AM <DIR> .

12/28/2022 05:44 AM <DIR> ..

12/28/2022 05:44 AM <DIR> archive-tmp

12/28/2022 05:44 AM <DIR> classes

12/28/2022 05:44 AM <DIR> generated-sources

12/28/2022 05:44 AM <DIR> maven-status

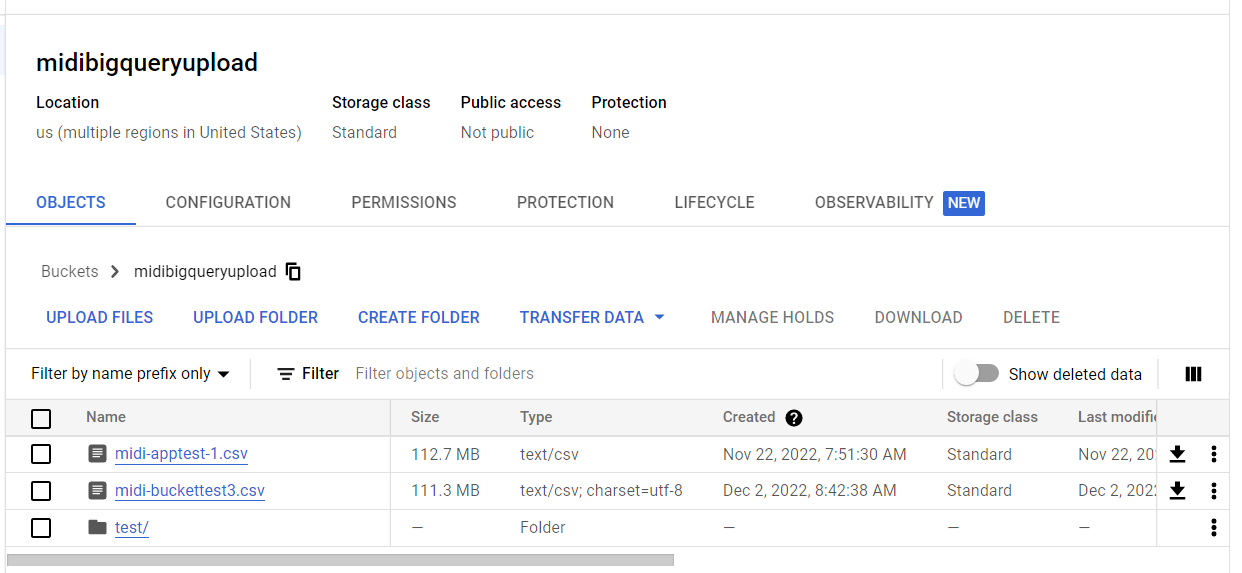
12/28/2022 05:44 AM 7,323,796 midi-analysis-1.0-SNAPSHOT-jar-with-dependencies.jar

## Preparing the Run Environment

Use instance *midi-linux-vm-analysis -1*.

### Mapping to Buckets

A single output bucket will hold the output csv files.



Create the directory you want to use for the output and map it.

cd midi\_app/

mkdir output

gcsfuse midibigqueryupload /home/gustafss/midi\_app/output

Create the directory you want to use for input and map it to the bucket where the DICOM files are using the switch --implicit-dirs to discover existing directories.

mkdir input

cd input

mkdir 21\_patient

gcsfuse --implicit-dirs 21patients\_10-19 /home/gustafss/midi\_app/input/21\_patient

## Run the Script Named RunProcess

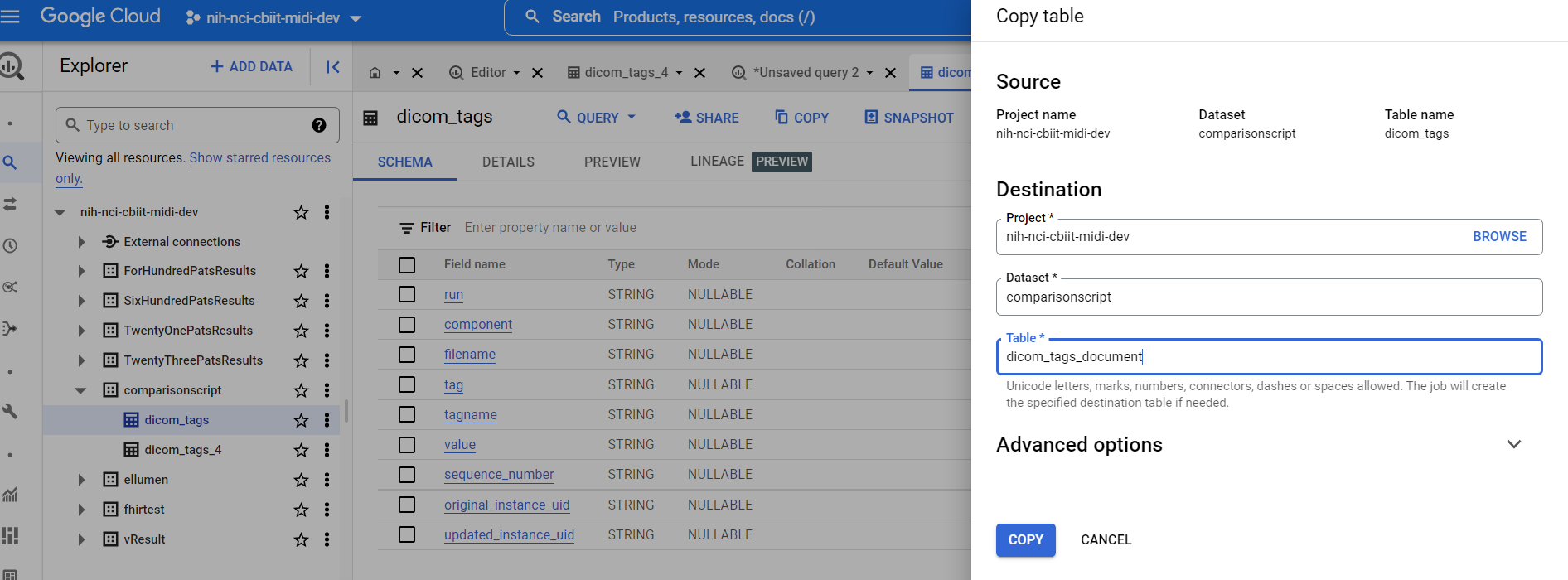
java -classpath "midi-analysis.jar" gov.nih.nci.midi.RunProcess buckettest3 midi /home/gustafss/input/21\_patient none /home/gustafss/midi\_app/output

### Parameters

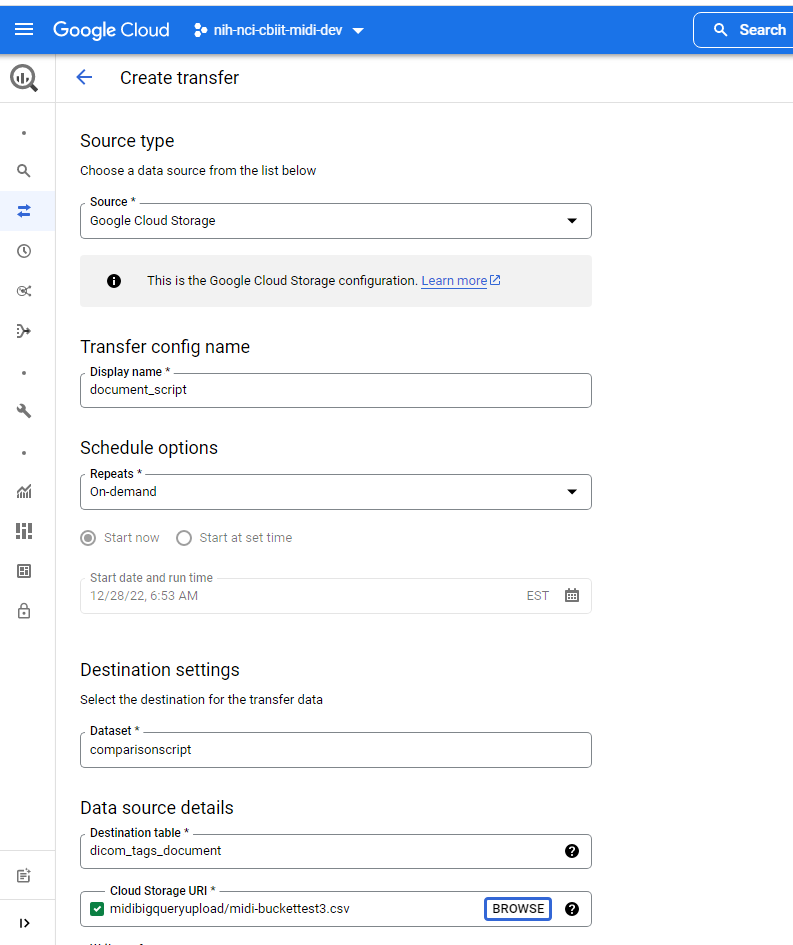
* Name of the run – this will be used to name the output csv
* Component being tested (midi, tcia, ctp)
* Path to the input directory
* Path to the crosswalk file – none means no crosswalk
* Path to the output directory

## Load to BigQuery

Create a new table for the load based on dicom\_tags.



Load the data from the output file in the bucket by creating and running a data transfer.



You can now use SQL to compare runs by joining the tables on the tags. Once you complete the analysis, simply delete the table.