MIDI Analysis Script and Database

Contents

[Code Location 1](#_Toc124874772)

[Requirements 1](#_Toc124874773)

[Building the Project 1](#_Toc124874774)

[Build Product 1](#_Toc124874775)

[Preparing the Run Environment 2](#_Toc124874776)

[Mapping to Buckets 2](#_Toc124874777)

[Run the Script Named RunProcess 2](#_Toc124874778)

[Script Parameters 3](#_Toc124874779)

[Load to BigQuery 3](#_Toc124874780)

## 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

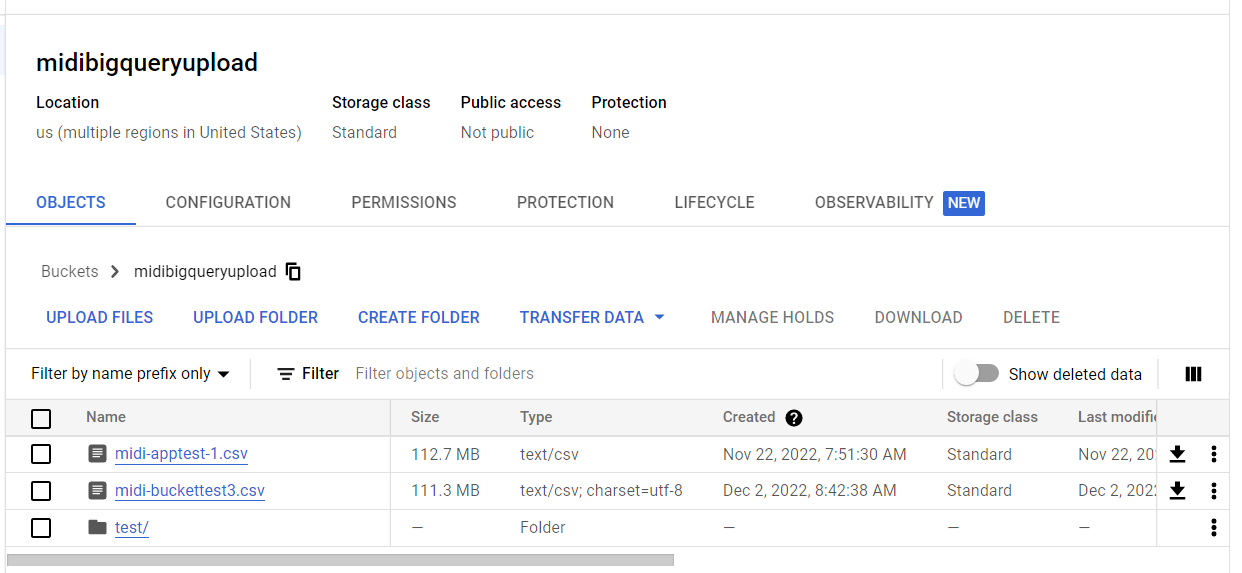
12/28/2022 05:44 AM 7,323,796 midi-analysis-1.0-SNAPSHOT-jar-with-dependencies.jar

## 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.

#### Example

cd midi\_app/

mkdir output

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

**Note:** This procedure uses the Google command *gcsfuse*. See its [Github page](https://github.com/GoogleCloudPlatform/gcsfuse) for more information.

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

#### Example

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

#### Example

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

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

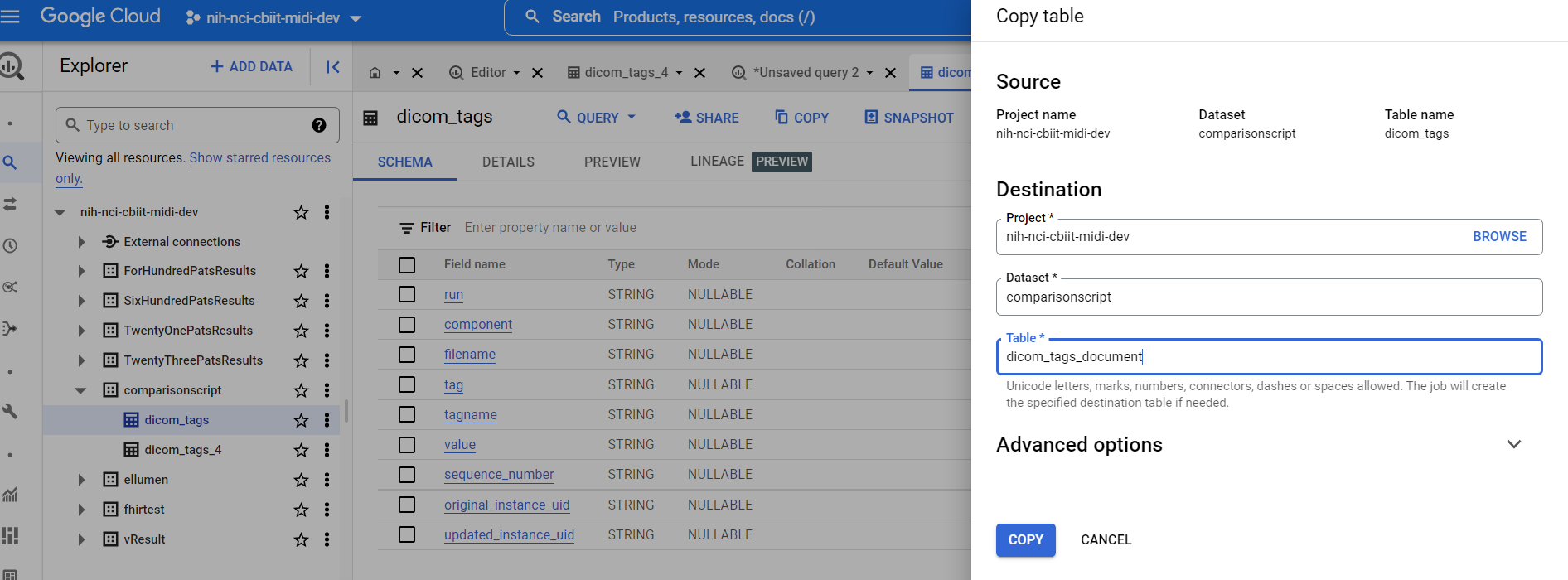
### Script Parameters

The script’s parameters are defined below.

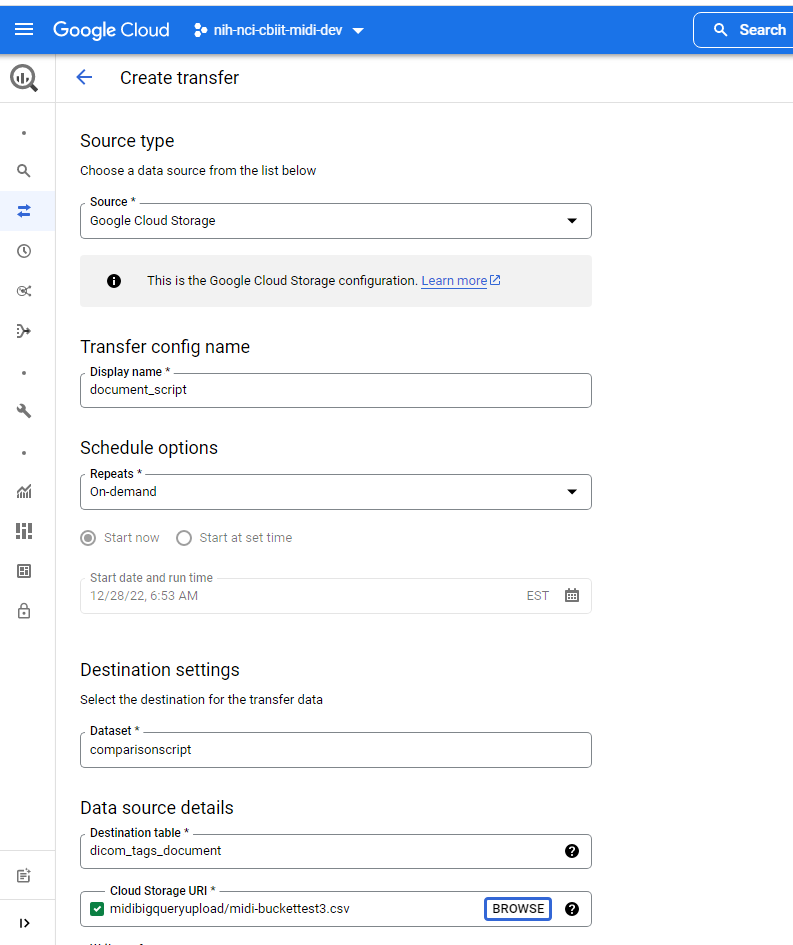
|  |  |
| --- | --- |
| **Definition** | **Example Parameter Value** |
| Name of the run – this will be used to name the output csv | buckettest3 |
| Component being tested (midi, tcia, ctp) | midi |
| Path to the input directory | /home/gustafss/input/21\_patient |
| Path to the crosswalk file – none means no crosswalk | none |
| Path to the output directory | /home/gustafss/midi\_app/output |

## Load to BigQuery

Create a new table for the load by copying the dicom\_tags table and renaming it.



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.