

Dataset-Searcher tools to gb2 commands.

Anil Panta

Dr. Michel H. Villanueva

Olemiss Belle II Meeting

Date: Month 25,2020

Dataset-Searcher:

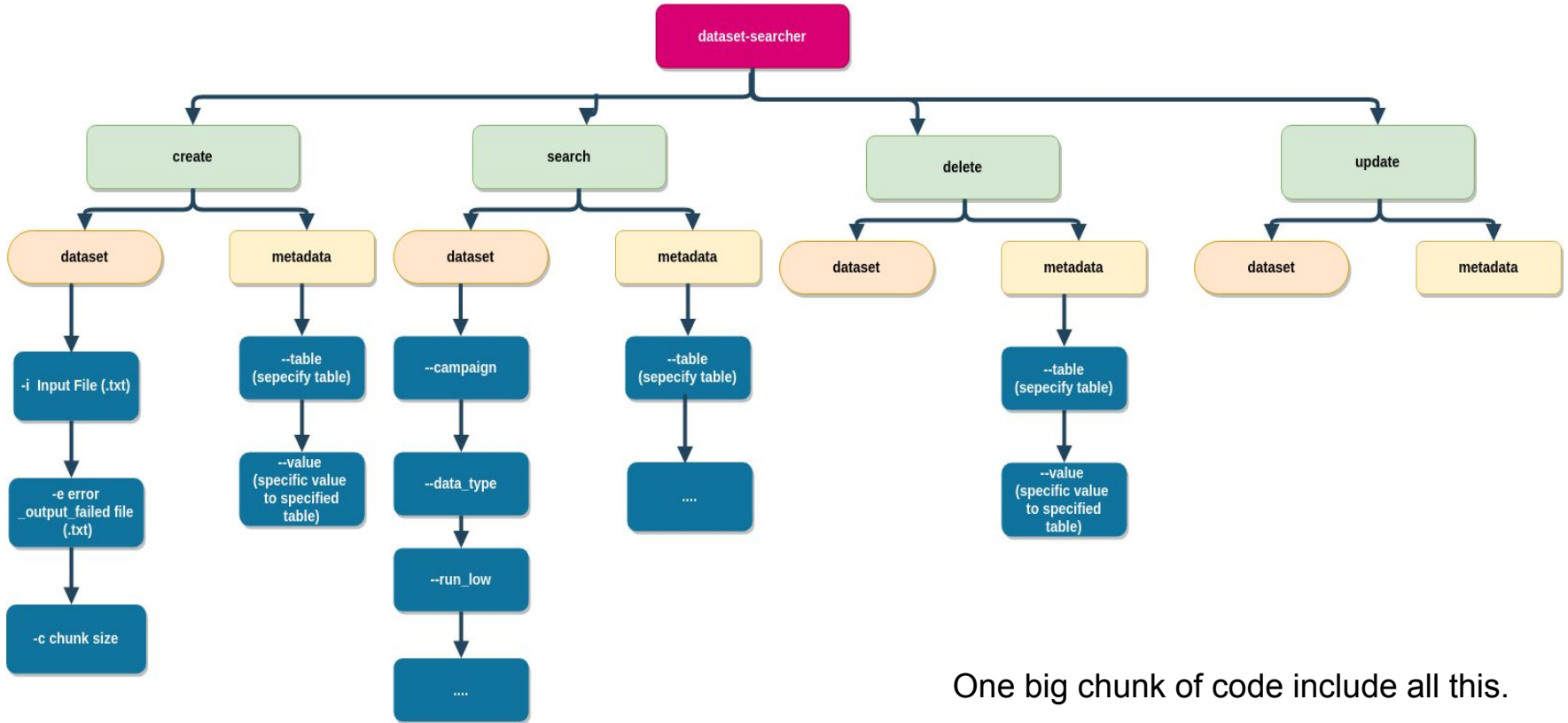
- Currently, Dataset-Searcher tools interacts with B2SystemLookUp in BelleDIRAC
- But the commands for using this tools are separate. Like,

```
python dataset-searching/dataset-searcher.py create dataset -i metadata_mc12b.txt -e failed_insert_datasets_mc12b.txt
```

- This is not the standard way to interact with (Belle)DIRAC.

Goal :

- Create the bunch of tool `gb2_tools` analogous with recent tool for consistency and easiness.



One big chunk of code include all this.

Idea:

- For separate tools:
 - a. gb2_dataset_searcher_search
 - b. gb2_dataset_searcher_create
 - c. gb2_dataset_searcher_delete
 - d. gb2_dataset_searcher_update

Each has option : gb2_dataset_searcher_<...> metadata [-- all_flags]

gb2_dataset_searcher_<...> dataset [-- all_flags]

Workflow:

- Move parser into `Client/helpers/optutil.py`
- Add four commands to `Client/bin/....`
- Add command functionality to `Client/controllers/datasetCLController.py`
- RPC call to `gbasf2/lib/ds/manager.py`
 - I choose `datasetcontroller.py` cause this is related to dataset.
[Question: Is this logical?]
 - Due to this RPC call has to be placed in `gbasf2/lib/ds/manager.py`

RPC call to be moved to gbasf2/lib/ds/manager.py:

1. Search:

```
searcher = get_searcher()
result = searcher.getRefTableValues(table_name)
```

2. Delete:

```
searcher = get_searcher()
result = getattr(searcher, 'delete'+table_name)(to_delete)['Value']
```

3. Create:

```
result = getattr(searcher, 'add'+table_name)(to_create)['Value']
dbCall = searcher.addDatasets(add_metadata, True)
```

4. Update:

```
result = getattr(searcher, method_name)(args.old_value, args.new_value)
result = searcher.updateDataset(args.lpn, fields, new_values)
result = searcher.getDatasets(['lpn', 'experimentLow', 'experimentHigh'], {'lpn':args.lpn}, {}, {})
```

- 4 diff function added to manager.py

a. gb2_dataset_searcher_search

```
Singularity sl_6.sif:~/belleUMclient> gb2_dataset_searcher_search --help
usage: gb2_dataset_searcher_search [-h] [-v] [--usage] {metadata,dataset} ...
```

positional arguments:

| | |
|--------------------|--|
| {metadata,dataset} | What kind of search is this? |
| metadata | Searches for dataset LPNs matching the given metadata constraints. |
| dataset | Returns available metadata values. |

optional arguments:

| | |
|---------------|---------------------------------|
| -h, --help | show this help message and exit |
| -v, --verbose | increase verbosity (up to -vv) |
| --usage | show detailed usage |

```
usage: gb2_dataset_searcher_search metadata --help
usage: gb2_dataset_searcher_search metadata [-h] --table
                                     {Releases,GlobalTags,MCEventTypes,GeneralSkimNames,SkimDecayModes,Campaigns,DataTypes,DataLevels,BeamEnergies}
optional arguments:
  -h, --help            show this help message and exit
  --table {Releases,GlobalTags,MCEventTypes,GeneralSkimNames,SkimDecayModes,Campaigns,DataTypes,DataLevels,BeamEnergies}
                        Specify a table to search values from .
```

```
Singularity sl_6.sif:~/belleUMclient> gb2_dataset_searcher_search dataset --help
usage: gb2_dataset_searcher_search dataset [-h] [-o OUTPUT_FILE]
                                           [--campaign CAMPAIGN]
                                           [--data_type DATA_TYPE]
                                           [--data_level DATA_LEVEL]
                                           [--run_high RUN_HIGH]
                                           [--exp_high EXP_HIGH]
                                           [--run_low RUN_LOW]
                                           [--exp_low EXP_LOW]
                                           [--mc_event MC_EVENT]
                                           [--skim_decay SKIM_DECAY]
                                           [--general_skim GENERAL_SKIM]
                                           [--beam_energy BEAM_ENERGY]
                                           [--global_tag GLOBAL_TAG]
                                           [--release RELEASE]
```


b. gb2_dataset_searcher_create

```
[apanta@cw13 BelleDIRAC]$ gb2_dataset_searcher_create --help
usage: gb2_dataset_searcher_create [-h] [-v] [--usage] {metadata,dataset} ...
```

positional arguments:

```
{metadata,dataset}  Adds new datasets/metadata options to the database so
                    they can be searched.
                    metadata  creates new metadata values in specific tables.
                    dataset   Creates new datasets using given metadata
```

optional arguments:

```
-h, --help          show this help message and exit
-v, --verbose       increase verbosity (up to -vv)
--usage            show detailed usage
```

```
[apanta@cw13 BelleDIRAC]$ gb2_dataset_searcher_create metadata --help
```

```
usage: gb2_dataset_searcher_create metadata [-h] --table
                                             {Releases,GlobalTags,MCEventTypes,GeneralSkimNames,SkimDecayModes,Campaigns,DataTypes,DataLevels,BeamEnergies}
                                             [--value VALUE]
```

optional arguments:

```
-h, --help          show this help message and exit
--table {Releases,GlobalTags,MCEventTypes,GeneralSkimNames,SkimDecayModes,Campaigns,DataTypes,DataLevels,BeamEnergies}
                    Specify a table to create new value in.
--value VALUE       Specify a value to create in a table.
```

```
[apanta@cw13 BelleDIRAC]$ gb2_dataset_searcher_create dataset --help
```

```
usage: gb2_dataset_searcher_create dataset [-h] -i INPUT_FILE
                                           [-e OUTPUT_FAILED] [-c CHUNK_SIZE]
```

optional arguments:

```
-h, --help          show this help message and exit
-i INPUT_FILE, --input-file INPUT_FILE
                    Specify a text file containing metadata of datasets.
-e OUTPUT_FAILED, --output-failed OUTPUT_FAILED
                    Output a text file containing all dataset entries that
                    failed to be uploaded.
-c CHUNK_SIZE, --chunk-size CHUNK_SIZE
                    Datasets will be uploaded in batches of this size.
```

C. gb2_dataset_searcher_delete

```
[apanta@cw07 Client]$ gb2_dataset_searcher_delete --help
usage: gb2_dataset_searcher_delete [-h] [-v] [--usage]
                                   (-i INPUT_FILE | --value VALUE) --table
                                   {Releases,GlobalTags,MCEventTypes,GeneralSkimNames,SkimDecayModes,Campaigns,DataTypes,DataLevels,BeamEnergies,Datasets}

optional arguments:
  -h, --help            show this help message and exit
  -v, --verbose         increase verbosity (up to -vv)
  --usage              show detailed usage
  -i INPUT_FILE, --input-file INPUT_FILE
                       Specify a text file containing values to delete.
  --value VALUE        Specify a value to delete from a table.
  --table {Releases,GlobalTags,MCEventTypes,GeneralSkimNames,SkimDecayModes,Campaigns,DataTypes,DataLevels,BeamEnergies,Datasets}
                       Specify a table to delete values from.
```

```
[apanta@cw13 BelleDIRAC]$ gb2_dataset_searcher update dataset --help
usage: gb2_dataset_searcher_update dataset [-h] --lfn LPN
                                           [--campaign CAMPAIGN]
                                           [--data_type DATA_TYPE]
                                           [--data_level DATA_LEVEL]
                                           [--run_high RUN_HIGH]
                                           [--exp_high EXP_HIGH]
                                           [--run_low RUN_LOW]
                                           [--exp_low EXP_LOW]
                                           [--mc_event MC_EVENT]
                                           [--skim_decay SKIM_DECAY]
                                           [--general_skim GENERAL_SKIM]
                                           [--beam_energy BEAM_ENERGY]
                                           [--global_tag GLOBAL_TAG]
                                           [--release RELEASE]
```

optional arguments:

```
-h, --help          show this help message and exit
--lfn LPN           The current dataset whose metadata want to update.
--campaign CAMPAIGN The MC or Data production campaign name.
--data_type DATA_TYPE
                   MC or Data
--data_level DATA_LEVEL
                   udst, mdst, etc
--run_high RUN_HIGH
--exp_high EXP_HIGH
--run_low RUN_LOW
--exp_low EXP_LOW
--mc_event MC_EVENT The MC event type ("uubar", "1110043100", etc) used
                   for
--skim_decay SKIM_DECAY
                   The skim type used to reconstruct and select events.
--general_skim GENERAL_SKIM
                   The general skim name (not in use currently!)
--beam_energy BEAM_ENERGY
                   4S, 5S, etc
--global_tag GLOBAL_TAG
                   The global tag used to create the dataset.
--release RELEASE  The basf2 release used to create the dataset.
```

D. gb2_dataset_searcher_update

```
[apanta@cw13 BelleDIRAC]$ gb2_dataset_searcher update --help
usage: gb2_dataset_searcher_update [-h] [-v] [--usage] {metadata,dataset} ...
```

positional arguments:

```
{metadata,dataset} Are we updating a Dataset or Metadata value
metadata           Updates the current value of a metadata table value
dataset            Updates a datasets metadata
```

optional arguments:

```
-h, --help          show this help message and exit
-v, --verbose       increase verbosity (up to -vv)
--usage             show detailed usage
```

```
[apanta@cw13 BelleDIRAC]$ gb2_dataset_searcher update metadata --help
usage: gb2_dataset_searcher_update metadata [-h] --table
                                           {Releases,GlobalTags,MCEventTypes,GeneralSkimNames,SkimDecayModes,Campaigns,DataTypes,DataLevels,BeamEnergies}
                                           --old-value OLD_VALUE --new-value
                                           NEW_VALUE
```

optional arguments:

```
-h, --help          show this help message and exit
--table {Releases,GlobalTags,MCEventTypes,GeneralSkimNames,SkimDecayModes,Campaigns,DataTypes,DataLevels,BeamEnergies}
                   Specify a table to delete value from.
--old-value OLD_VALUE
                   The currently held value we want to alter.
--new-value NEW_VALUE
                   The new value we want
```

Testing the code:

- I have tested the code at development server at Olemiss via UMClient.

Big thanks to Michel.

I have tested the code manually for each of four.

`All_of_what_i_tested['OK']`

Do we have to check all the flag?

If yes then need some efficient way of testing. [May be UnitTest] or bash script. ???

Authorization:

- Currently Authorization is given to `__auth_admin` + `__auth_devel` + `__auth_prod` + `__auth_user`
- I believe authorization for create, delete, update should be for `__auth_prod`
- And search for all .

Suggestion needed ?

```
__auth_user = ['belle', 'belle_monitor', 'belle_beast',  
              'belle_skim', 'test']  
__auth_prod = ['belle_mcprod', 'belle_prod', 'belle_dataprod']  
__auth_devel = ['belle_devel']  
__auth_admin = ['dirac_admin']
```

Next step:

1. Put check point for rpc and other call and return the call messages if fail.
2. Finally, polish the code according rules.
3. Documentation.

Target for completion: **By end of this week.**

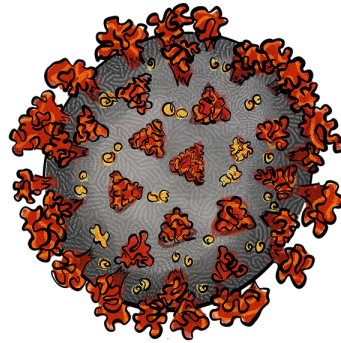
Next Next Plan for dataset-searcher:

Query doesn't happen in the safest way when there are potentially 10s of 1000s of available LPNs. Should probably fail without any query, or without sufficient criteria. Should also return things in "chunks" of 1000 LPNs. Then the searchers can query repeatedly until it returns an empty result. Get query to return records that are offset (e.g. offset by 0 returns first 1000, offset by 1000 returns next 1000, etc.)

- I have idea on how to solve this.
- I know how to put offset on database query.
- I will get back to this after making pull request for the gb2_tools for dataset-searcher is done.
- The development Environment for this is made ready by [Michel](#).

[Suggestion , comments ?](#)

Stay Safe.



COVID-19

Take Prevention.