THE UNIVERSITY of MISSISSIPPI

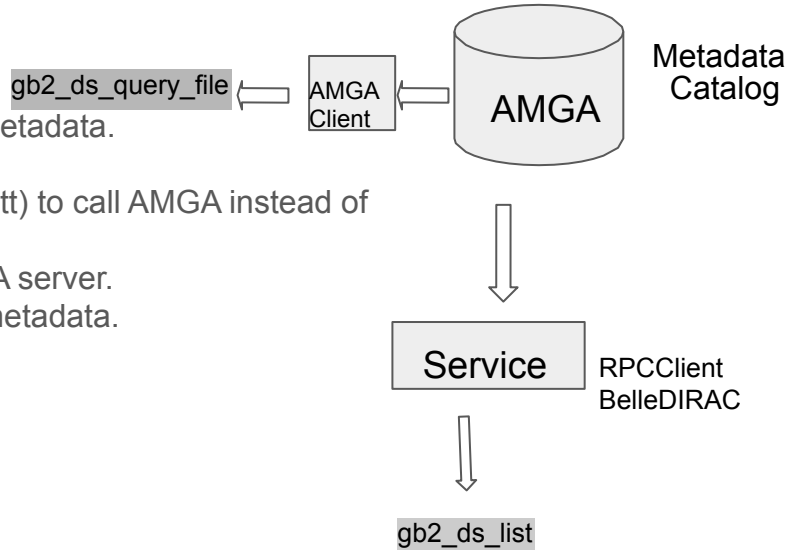
BelleDIRAC Development
gb2_ds_list [BIIDCD-902]

Branch: feature/gb2_ds_list_dev

Anil Panta, Michel Villanueva

Goal & Idea:

- Goal is to make gb2_ds_list to list files having ‘good’ status..
- Currently , gb2_ds_list just lists the LFN of the file/directory.
- On the other hand, gb2 _ds tools use AMGA client to retrieve metadata.
- Our idea is to use DIRAC service(implemented by David Dossett) to call AMGA instead of client.
- RPC helps to avoid the details in the communication with AMGA server.
- Then we will validate which files contains ‘good’ status on the metadata.



AMGA: ARDA Metadata Grid Application
LFN: Logical File Name
RPC : Remote Procedure Call

Implementation:

1) If the <dataset> is LPN that contains files or <dataset> is LFN itself.

Old flag

a) gb2_ds_list <dataset> : List the files with status ‘good’
b) gb2_ds_list <dataset> -l : List LFN of files with site name , siteLFN and size
c) gb2_ds_list <dataset> -l -g : List LFN of files with the site name and size

New flag

d) gb2_ds_list <dataset> -status <status_name> : List the LFN which have status given by user
e) gb2_ds_list <dataset> -a : List all the files (doesn’t check for status)(No AMGA call)

2) If the <dataset> is LPN that contains another directory then no AMGA call. It does the same thing as it did before this feature.

Files where code has been changed to implement this:

1) BelleDIRAC/Client/controller/datasetCLController.py
2) BelleDIRAC/Client/bin/gb2_ds_list
3) BelleDIRAC/Client/helpers/optutil.py

Note: Refer to branch for more detail.

1.a gb2_ds_list <dataset> : List the files with status ‘good’

Code for this :

```
from DIRAC.Core.DISET.RPCClient import RPCClient
am = RPCClient('DataManagement/AMGAManager')
if not (args.long or args.list_all_files or  args.list_specific_files):
    res = am.getAttributesValues(args.dataset,['status'])
    if res['OK']:
       for key,value in res['Value'].items():
          if value[0] == 'good':
             print os.path.join(args.dataset,key)
       return ''
    else:
        return os.linesep.join(sorted(lfns['Value'], key=natural))
```

Output:

```
[apanta@cw10 ~]$ gb2_ds_list /belle/user/michmx/test_badFiles/sub01
1
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-71-of-100.mdst
2
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-38-of-100.mdst
3
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-29-of-100.mdst
4
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-74-of-100.mdst
5
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-54-of-100.mdst
77
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-73-of-100.mdst
78
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-59-of-100.mdst
79
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-11-of-100.mdst
80
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-92-of-100.mdst
```

Output:

```
[apanta@cw10 ~]$ gb2_ds_list /belle/user/michmx/test_badFiles
/belle/user/michmx/test_badFiles/sub00
/belle/user/michmx/test_badFiles/sub01
/belle/user/michmx/test_badFiles/sub02
```

Note: Validation is done if the LPN contains files or another directory.
Test : Michel created dataset containing 100 files,(80 good and 20 bad status). (Thanks to Michel)
Note: The number on output is for testing purpose. Final version does not print number.

1.b gb2_ds_list <dataset> -l -g (-l :long , -g : group by SE)

Output:

```
[apanta@cw10 ~]$ gb2_ds_list /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub00 -l -g
KEK2-TMP-SE:
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub00/mdst_000001_prod00005678_task00000001.root
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub00/mdst_000002_prod00005678_task00003920.root
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub00/mdst_000003_prod00005678_task00000003.root
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub00/mdst_000004_prod00005678_task00000004.root
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub00/mdst_000005_prod00005678_task00000005.root
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub00/mdst_000006_prod00005678_task00003921.root
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub00/mdst_000007_prod00005678_task00000008.root
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub00/mdst_000008_prod00005678_task00000009.root

KEK-DISK-TMP-SE:
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub00/mdst_000001_prod00005678_task00000001.root
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub00/mdst_000002_prod00005678_task00003920.root
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub00/mdst_000003_prod00005678_task00000003.root
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub00/mdst_000997_prod00005678_task00000997.root
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub00/mdst_000998_prod00005678_task00000998.root
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub00/mdst_000999_prod00005678_task00000999.root
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub00/mdst_001000_prod00005678_task00001000.root

total 1000 good Files ,2.0 TB
```

Cont.. gb2_ds_list <dataset> -l -g : List LFN of files with the site name and size

```
lapanta@cw10 ~]$ gb2_ds_list /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst -l -g
KEK2-TMP-SE: (4/4), 7.6 TB
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub00
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub01
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub02
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub03

KEK-DISK-TMP-SE: (4/4), 7.6 TB
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub00
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub01
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub02
        /belle/MC/release-02-00-01/DB00000411/MC11/prod00005678/s00/e0000/4S/r00000/mixed/mdst/sub03

total 4 files, 15.2 TB
```

Note: Notice the difference on the output in here and in slide 6.

1.c gb2_ds_list <dataset> -l : List LFN of files with site name , siteLFN and size

```
[apanta@cw10 ~]$ gb2_ds_list /belle/user/michmx/test_badFiles/sub01 -l
1
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-1-of-100.mdst
3.8 MB
        KEK-TMP-SE - srm://kek2-se02.cc.kek.jp:8444/srm/managerv2?SFN=/belle/TMP/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-1-of-100.mdst

2
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-2-of-100.mdst
3.9 MB
        KEK-TMP-SE - srm://kek2-se02.cc.kek.jp:8444/srm/managerv2?SFN=/belle/TMP/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-2-of-100.mdst
```

```
80
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-99-of-100.mdst
3.9 MB
        KEK-TMP-SE - srm://kek2-se02.cc.kek.jp:8444/srm/managerv2?SFN=/belle/TMP/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-99-of-100.mdst
```

```
[apanta@cw10 ~]$ gb2_ds_list /belle/user/michmx/test_badFiles -l
/belle/user/michmx/test_badFiles/sub00
4.9 GB

/belle/user/michmx/test_badFiles/sub01
386.9 MB

/belle/user/michmx/test_badFiles/sub02
31.2 MB
```

1.d gb2_ds_list <dataset> -status <status_name> : List the LFN which have status given by user

Usages: gb2_ds_list <dataset> -status good
gb2_ds_list <dataset> -status bad

```
[apanta@cw13 ~]$ gb2_ds_list /belle/user/michmx/test_badFiles/sub00 -status bad
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_58.mdst.root (bad)
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_60.mdst.root (bad)
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_47.mdst.root (bad)
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_34.mdst.root (bad)
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_71.mdst.root (bad)
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_89.mdst.root (bad)
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_23.mdst.root (bad)
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_94.mdst.root (bad)
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_31.mdst.root (bad)
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_100.mdst.root (bad)
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_36.mdst.root (bad)
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_61.mdst.root (bad)
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_45.mdst.root (bad)
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_25.mdst.root (bad)
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_80.mdst.root (bad)
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_6.mdst.root (bad)
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_79.mdst.root (bad)
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_51.mdst.root (bad)
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_75.mdst.root (bad)
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_68.mdst.root (bad)
/belle/user/michmx/test_badFiles/sub00/taupair_genLevel_66.mdst.root (bad)
```

```
[apanta@cw09 BelleDIRAC]$ gb2_ds_list /belle/user/michmx/test_badFiles -status bad
error: connot get the status cause LPN  has subdirectory
```

1.e gb2_ds_list <dataset> -a : List all the files (doesn't check for status)(No AMGA call)

```
[apanta@cw10 ~]$ gb2_ds_list /belle/user/michmx/test_badFiles/sub01 -a
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-1-of-100.mdst
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-2-of-100.mdst
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-3-of-100.mdst
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-4-of-100.mdst
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-5-of-100.mdst
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-6-of-100.mdst
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-7-of-100.mdst
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-8-of-100.mdst
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-9-of-100.mdst
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-10-of-100.mdst
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-11-of-100.mdst
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-12-of-100.mdst
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-13-of-100.mdst
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-14-of-100.mdst
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-15-of-100.mdst
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-16-of-100.mdst
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-17-of-100.mdst
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-18-of-100.mdst
/belle/user/michmx/test_badFiles/sub01/gen_ee_4pi_BGx1-19-of-100.mdst
```

Performance timing:

- There is no significant difference in timing before and after AMGA call.
- So timing is not an issue while using this feature.

gb2_ds_list <dataset> : 1000 files
Before (No AMGA call) :
Real 0m 2.092s
user 0m 0.822s
sys 0m 0.227s

After AMGA(using RPCClient) call:
Real 0m 2.334s
User 0m 0.734s
sys 0m 0.211s

real -> wall clock time
user -> CPU time
Sys -> Amount of CPU time spent in the kernel

Next project:

Michel knows it .

I don’t know it yet but i know there is one.