

Skim Production Manual

Skim Task Manager Version 2007

G. Castelli
Rutherford Appleton Laboratory

March 17, 2008

Contents

1	Introduction	2
2	It works so	2
2.1	Import	2
2.2	Environment and configuration files	2
2.2.1	BbkEditTask -c	2
2.2.2	BbkEditTask -w	3
2.3	Skim	3
2.3.1	Conventions	3
2.3.2	BbkCreateTask	3
2.3.3	BbkTaskInfo	3
2.3.4	BbkCreateSkims -n 1	3
2.3.5	BbkCreateSkims -n 10	3
2.3.6	BbkSubmitSkims	3
2.3.7	BbkSubmitSkims -t	4
2.3.8	BbkShowSkimsJobs	4
2.3.9	BbkCheckSkims	4
2.3.10	BbkCheckSkims -pend -status 4	4
2.3.11	BbkCheckSkims -pend -status 1	4
2.4	Merge	4
2.4.1	BbkCreateMerges	4
2.4.2	BbkCreateMerges -force	4
2.4.3	BbkSubmitMerges	5
2.4.4	BbkShowMergeJobs	5
2.4.5	BbkCheckMerges	5

2.4.6	BbkCheckMerges -pend -status 4	5
2.4.7	BbkCheckMerges -pend -status 1	5
2.5	Bookkeeping	5
2.6	Export	5
2.6.1	BbkTMExport -list	5
2.6.2	BbkUser ...	5
2.6.3	BbkTMExport -noexport	6
2.6.4	BbkTMExport	6
2.7	Database commands	6
2.7.1	rundir with status 1	6
2.7.2	min and status 1	6
2.7.3	max and status 1	6
2.7.4	All infos	6
2.7.5	rundir for a particular job in a particular status	7
2.7.6	min	7
2.7.7	mj_rundir	7
2.7.8	sdse_name	7
2.8	Non-Grid monitor	7
2.9	Grid Monitor	7
2.10	Grid retrieve	8
2.11	To delete wrong merged files	8
2.12	To delete skimmed and merged successful .root files	8

1 Introduction

This short manual is intended to be a fast but almost complete guide to the Task Manager version 2.

The right environment as well as the right configuration files are needed and supposed to be already correctly set up.

The two needed configuration files will be called: `configFileTask.txt` and `configFileTaskTemplate.txt`.

2 It works so

2.1 Import

2.2 Environment and configuration files

2.2.1 BbkEditTask -c

To change the configuration of the task (<taskname> is defined in `configFileTaskTemplate.txt`):

```
BbkEditTask -c <config-update-file> <taskname>
```

2.2.2 BbkEditTask -w

To write in <current-config-file> the configuration of the task:

```
BbkEditTask -w <current-config-file> <taskname>
```

2.3 Skim

2.3.1 Conventions

The following names need to be set up for the dbsite, dbname and dbuser variables for the different cases (i.e. ral, slac, man, ...):

dbsite	dbname	dbuser
ral	stm1	bbruser
slac	slacskim	slacskim
man

All the following examples are for the RAL case.

2.3.2 BbkCreateTask

To create a task:

```
BbkCreateTask --dbsite=ral --dbname=stm1 \
  --dbuser=bbruser configFileTask.txt
```

2.3.3 BbkTaskInfo

About the infos of all the tasks created:

```
BbkTaskInfo --dbsite=ral --dbname=stm1 \
  --dbuser=bbruser --listtasks
```

2.3.4 BbkCreateSkims -n 1

To create a single skim job:

```
BbkCreateSkims -n 1 <taskname>
```

2.3.5 BbkCreateSkims -n 10

To create ten skim jobs:

```
BbkCreateSkims -n 10 <taskname>
```

2.3.6 BbkSubmitSkims

To submit a skim job:

```
BbkSubmitSkims -v --njobs 1 <taskname>
```

2.3.7 **BbkSubmitSkims -t**

To try to submit a skim job without submitting it:

```
BbkSubmitSkims --dbsite=ral --dbname=stm1 \  
                --dbuser=bbruser -v --njobs 1 -t <taskname>
```

2.3.8 **BbkShowSkimsJobs**

To see the results of the skim jobs:

```
BbkShowSkimsJobs <taskname>
```

2.3.9 **BbkCheckSkims**

To check the skim:

```
BbkCheckSkims <taskname>
```

2.3.10 **BbkCheckSkims --pend --status 4**

To reprepare the failed skim jobs:

```
BbkCheckSkims --pend --status 4 <taskname>
```

2.3.11 **BbkCheckSkims --pend --status 1**

To reprepare the submitted skim jobs:

```
BbkCheckSkims --pend --status 1 <taskname>
```

2.4 Merge

The merge commands are pretty similar to the skim commands.

2.4.1 **BbkCreateMerges**

To create merge jobs from what has been skimmed:

```
BbkCreateMerges <taskname>
```

2.4.2 **BbkCreateMerges --force**

To force to create merge jobs (the 'G-run2-1-18.6.4a' or similar string comes out from the previous command if the number of events is less than what's requested):

```
BbkCreateMerges <taskname> --force 'G-run2-1-18.6.4a'
```

2.4.3 BbkSubmitMerges

To submit a merge job:

```
BbkSubmitMerges -v --njobs 1 <taskname>
```

2.4.4 BbkShowMergeJobs

To see the results of the merge jobs:

```
BbkShowMergeJobs <taskname>
```

2.4.5 BbkCheckMerges

To check the merge:

```
BbkCheckMerges <taskname>
```

2.4.6 BbkCheckMerges --pend --status 4

To reprepare the failed merge jobs:

```
BbkCheckMerges --pend --status 4 <taskname>
```

2.4.7 BbkCheckMerges --pend --status 1

To reprepare the submitted merge jobs:

```
BbkCheckMerges --pend --status 1 <taskname>
```

2.5 Bookkeeping

This command is run at RAL to load the collection in the bookkeeping database:

```
BbkTMLoad --dbname=stm1 --dbuser=bbruser --bbkpubuser=bbkr18 \  
          --skimcycle=R18dTestUK -n 1 \  
          <taskname>
```

2.6 Export

2.6.1 BbkTMExport --list

To list the .root files that could be exported:

```
BbkTMExport --list --taskname <taskname>
```

2.6.2 BbkUser ...

To see the UK collection exported at SLAC:

```
BbkUser --site=slac --skim_cycle=R18dTestUK collection
```

2.6.3 BbkTMExport --noexport

To try to export without exporting:

```
BbkTMExport --number 1 --noexport --debug --taskname <taskname>
```

2.6.4 BbkTMExport

To export:

```
BbkTMExport --number 1 --debug --taskname <taskname>
```

2.7 Database commands

There are a lot of queries that can be submitted to the database, in the following some examples.

2.7.1 rundir with status 1

To see the run directories of the jobs with status 1:

```
BbkTMUser --dbname=stm1 --dbsite=ral --dbuser=bbruser \
  --taskname <taskname> \
  --sjob_status 1 sjob_id rundir
```

2.7.2 min and status 1

To see the minimum job identification number with status 1:

```
BbkTMUser --dbsite=ral --dbname=stm1 --dbuser=bbruser \
  --taskname <taskname> \
  'MIN(sjob_id)' --sjob_status 1 \
  -q --style=list
```

2.7.3 max and status 1

To see the maximum job identification number with status 1:

```
BbkTMUser --dbsite=ral --dbname=stm1 --dbuser=bbruser \
  --taskname <taskname> \
  'MAX(sjob_id)' --sjob_status 1 \
  -q --style=list
```

2.7.4 All infos

To list the infos in the database:

```
BbkTMUser --list
```

2.7.5 rundir for a particular job in a particular status

```
BbkTMUser --dbname=stm1 --dbsite=ral --dbuser=bbruser \
-q --style=list \
--taskname <taskname> \
--sjob_id <jobid> sjob_status rundir
```

2.7.6 min

To see the minimum job identification number:

```
BbkTMUser --dbsite=ral --dbname=stm1 --dbuser=bbruser \
--taskname <taskname> 'MIN(sjob_id)' \
-q --style=list
```

2.7.7 mj_rundir

To look for the mj_rundir value:

```
BbkTMUser --dbname=stm1 --dbsite=ral --dbuser=bbruser \
-q --style=list --mjob_id <jobid> mj_rundir
```

2.7.8 sdse_name

To look for the sdse_name value:

```
BbkTMUser --dbname=stm1 --dbsite=ral --dbuser=bbruser \
-q --style=list --mjob_id <jobid> sdse_name
```

2.8 Non-Grid monitor

2.9 Grid Monitor

To run the Grid job monitor as a daemon:

```
BbkTMGridSkimJobMonitor --verbose --logfile=<logfile> --task <taskname>
```

where <logfile> is generated by submitting jobs with:

```
BbkSubmitSkims -n 15 <taskname> >> <logfile>
```

That should produce a rolling list of job status changes. To get a snapshot of the current statuses you can run it like:

```
BbkTMGridSkimJobMonitor --task <taskname> --print (--check)
```

CE Name	Sched	Running	Success	Aborted	Cleared	Other
heplnx206.pp.rl.ac.uk	4	2	0	0	0	0

RB Name	Sched	Running	Success	Aborted	Cleared	Other
lcgrb02.gridpp.rl.ac.uk:9000	1	0	0	0	0	0
lcgrb01.gridpp.rl.ac.uk:9000	3	2	0	0	0	0

If it isn't running as a daemon –check to call `edg-job-status` on all runs should be added.

2.10 Grid retrieve

To retrieve the tarball from the Storage Element (Grid case):

```
BbkTMGridSkimJobRetriever <global_path_output_directory> \
    <taskname> (<jobid_min> <jobid_max>)
```

2.11 To delete wrong merged files

To delete wrong `.root` merged files that will not be created next time because of their existence, go to the right directory (`<taskname>/00` for instance) and type the first line for check and the second one as operative command:

```
awk '/already exists/{print $5}' */*/merge_*.log | xargs -n 1 echo rm
awk '/already exists/{print $5}' */*/merge_*.log | xargs -n 1 rm
```

2.12 To delete skimmed and merged successful `.root` files

To delete skimmed and merged `.root` files already copied at SLAC and published in the database with the aim of freeing local disk space, go to the beginning of the output directory tree for the `.root` skimmed and merged files (I'm in the `/stage/bdata-data93/prod/output.skimming` directory), then, for deleting the skimmed files (first command is a check and the second is the operative one):

```
awk '/unmerged/{print $1}' \
    /stage/babar-awg1/prod/work.my/my-18.6.4a.chris/workdir/deletable.lis \
    | sed -e 's/\(.*\)/.\1.01.root/' | xargs -n 1 echo rm
awk '/unmerged/{print $1}' \
    /stage/babar-awg1/prod/work.my/my-18.6.4a.chris/workdir/deletable.lis \
    | sed -e 's/\(.*\)/.\1.01.root/' | xargs -n 1 rm
```

where the `deletable.lis` file is created with the command:

```
BbkTMUser --dbname=stm1 --mjob_status=5 --taskname=<taskname> \
    sdse_name --output=deletable.lis
```

For deleting the merged files (idem as above):

```
awk '/.root/{print $1}' \
    /stage/babar-awg1/prod/work.my/my-18.6.4a.chris/workdir/del.lis \
    | sed -e 's/\(.*\)/.\1/' | xargs -n 1 echo rm
awk '/.root/{print $1}' \
    /stage/babar-awg1/prod/work.my/my-18.6.4a.chris/workdir/del.lis \
    | sed -e 's/\(.*\)/.\1/' | xargs -n 1 rm
```

where the del.lis file is created with a command that requires the version tja070206a of BbkSqlAbstractor:

```
bcvs update -r tja070206a BbkSqlAbstractor
```

The command is this one:

```
BbkTMUser --dbname=stm1 --mdse_bbk=1 --mdse_xfer=1 --mdse_hpss=0 \
          mdse_name -q | sort -u \
| relBbkUser --site=slac --dbname=bbkr18 --skim_cycle=R18dTestUK \
          --is_local=1 -i dse=- dse -q | sort -u \
| BbkTMUser --dbname=stm1 --dbuser=bbruser -i mdse_name=- \
          mfile_name --set mdse_hpss=1 --output=del.lis
```

If there are no new files that can be deleted this will say:

```
2007/02/06 15:59:42-Exiting-Input file - for column 'mdse_name' is empty
```

The list of files will be written to del.lis.

The above command will only show newly completed exports, if you need it, you can get a complete list of the successfully exported merge files with:

```
BbkTMUser --dbname=stm1 --mdse_bbk=1 --mdse_xfer=1 --mdse_hpss=1 \
          mfile_name -q
```