

Table of Contents

Closing Productions.....	1
What does it mean closing a production?.....	1
Which Production?.....	1
How to close?.....	1
Reconstruction production (FULL and EXPRESS).....	3
Merging productions.....	4
Replication transformations.....	5

Closing Productions

In this page we illustrate how and which production must be closed. As far as concerns MC productions it is worth to say that what is reported in this twiki is carried out **automatically** by an agent ad hoc. Only real data reconstruction/reprocessing/merging productions have then to be considered pertinent to this document requiring manual intervention and supervision of human being.

What does it mean closing a production?

Closing a production means that all records on all ProductionDBs will be scrapped and just the reminiscence of such production and its bare description are kept. The output produced by the production (must) remains in the catalogs and storages. As matter of fact this procedure is meant exactly to guarantee that the output produced by a given production is consistently available in the SEs and catalogs (file catalog and BKK) as per computing model or as per policy defined on per production basis. Closing a production it is just about to:

- Log into the dirac portal as production manager
- open the production monitor
- select the production
- set it to complete.

Automatically, after 15 days, the production will be archived by the *TransformationCleaningAgent*.

This operation however should be done after having checked the output as described in this page

Which Production?

In general all productions not meant to produce any longer any physically relevant data should be stopped. The Production Manager or the Grid Expert on Call should go through all active productions in the Production Monitoring system and pick up these productions.

- A first trivial selection are all productions used for validating some particular processing pass. These productions not only will be closed (set to Complete and then Archived afterward) but must be Cleaned. **Be aware** that cleaning a production it is a tough operations that implies all jobs in the system are killed and all output produced wiped from the system. We strongly suggest - before cleaning any production - to get in touch with experienced operations managers and double check with them.
- Another set of productions one can imagine to close are all productions in status *stopped* . The PM or the GEOC should go through various elog and find as much information as possible about this production and the reason has been stopped. It may be that this production was stopped because producing crappy data (wrong conditions/alignment/version of software), in that case the production could be even cleaned. It may well be that this production has been just temporarily stopped, in that case the production should be left as it is.
- The last set of productions candidate for being closed are all productions at 100%. This is usually the most important set of production that must be carefully checked before being closed.

How to close?

As said closing a production is not just matter of setting its status to complete. Before doing that one has to insure that a production producing physically relevant data had its output correctly stored in the SE and registered in the catalogs. This is why one has to run - depending on the type of production a couple of

command lines as shown later. Please note that Stripping production do not require these checks. Indeed the output of the stripping (un-merged DSTs) is usually removed being just a temporary step before the merging. Irrespectively the type of production, the PM has to:

- set temporary active the production that is going to be closed (otherwise he/she can't run the relevant command line as shown in the example). Please bear in mind that setting active a production may resurrect tasks that were still in the belly but not submitted in the WMS. For this reason we suggest to
 - verify beforehand that no more tasks were created and then that the number of submitted jobs matches the number of created tasks in the production monitoring.
 - check that the number of jobs in final status (Completed/Done/Failed) is the same as the number of submitted jobs;
 - check that there are no running/waiting/staging jobs. Indeed 100% completed in the production monitoring may not necessarily mean that all files have been processed and all jobs get to completion. A more deep investigation before proceeding is to
 - check the file status (click on the production raw on the monitoring page and select "File Status"). Ideally the totality of the files should be Processed. In real life there are often files in MaxReset status and even worse files in Assigned or Unused status. In this case, despite the production is claimed to be at 100% it is not; we suggest to first check the procedure described at this link.
 - For Merging(Stripping) production of a first processing activity you must also check that the number of files in Done status from the previous reconstruction production is exactly the same of the input files for the stripping. It might happen that some runs have not been flagged OK and then not picked up by the stripping.
-
- Last point is about completed jobs. It may well be that there are still pending requests to be honored for some of these jobs. Often, old requests are just due to a bug in the Request Management System and the operation has been carried on successfully long while ago but simply not registered properly in the RMS. This is not harmful. However a pending request is usually something one should worry about. For example a failover transfer request or a file catalog registration. On a case by case basis the PM or GEOC has to check all these pending requests ("Show jobs" --> "Select Status = Completed" --> "Get PendingRequest" for each of these selected jobs) and verify that the original requests does not affect the consistency of the output data among catalog and storages.

Reconstruction production (FULL and EXPRESS).

Once the points listed above have been done the PM can finally run the various data integrity and consistency checks. For Reconstruction production, being the output (SDST) stored only in one site, it is enough to check that all output are properly registered in the catalogs and stored on the Grid SEs. The following command must be issued.

```
dirac-production-verify-outputdata <ProdID>
```

This command will return nothing in case everything is OK. Otherwise inconsistencies are reported and these must be passed to the Data Operations Manager for being cured. Once all pathologies are fixed the PM/GEOC can set the production to Complete.

Merging productions.

As the merging productions wait until there is enough data to be merged or for the completion of the stripping of each run, it may happen that some files have not been processed by the merging production(s). In which case they can be seen as "Unused" in the "File status". If this occurs, look at the "Run status" to see which runs are not completed (Unused files) and check the same runs in the "Run status" of the stripping production. If the run(s) have been fully stripped, or cannot be fully stripped (problematic files), one can force the merging either with "Action/Flush" of the whole production, or "Flush" of individual run(s).

Replication transformations

The output of the merging is supposed to have multiple replicas around the various centers; for this reason the PM/GEOC has **also** to check also the number of copies replicated around the Grid and compare with the expectations for a given production. The default is to have 2 archive replicas and 4 other replicas (sometimes there are more due to failover mechanisms). The replication transformations are usually in request 0. As they are often very generic (by processing pass), they should not be set to Completed unless the processing pass is over (new processing pass, and all productions completed for this processing pass).

Before running the command to check the consistency of the data there is an extra command to run:

```
dirac-dms-replica-stats -p <ProdID>
```

whose example of output is given below

```
[pclhcb53] ~ $ dirac-dms-replica-stats -p 10571
67 files found in ['/lhcb/LHCb/Collision10/MINIBIAS.DST/00010571/0000/']
```

```
Replica statistics:
0 archives: 0 files
1 archives: 0 files
2 archives: 67 files
0 replicas: 0 files
1 replicas: 0 files
2 replicas: 0 files
3 replicas: 0 files
4 replicas: 64 files
5 replicas: 3 files
```

```
SE statistics:
  CERN-ARCHIVE: 67 files
  CNAF-ARCHIVE: 2 files
  IN2P3-ARCHIVE: 20 files
    PIC-ARCHIVE: 18 files
    RAL-ARCHIVE: 9 files
    SARA-ARCHIVE: 18 files
  CERN_M-DST: 67 files
  CNAF-DST: 3 files
  CNAF_M-DST: 21 files
  GRIDKA-DST: 30 files
  GRIDKA_M-DST: 16 files
  IN2P3-DST: 35 files
  IN2P3_M-DST: 21 files
    PIC-DST: 20 files
    PIC_M-DST: 9 files
    RAL-DST: 30 files
    SARA-DST: 19 files
```

```
Sites statistics:
  LCG.CERN.ch: 67 files
  LCG.CNAF.it: 24 files
  LCG.GRIDKA.de: 46 files
  LCG.IN2P3.fr: 56 files
    LCG.PIC.es: 29 files
    LCG.RAL.uk: 30 files
  LCG.SARA.nl: 19 files
```

Once both the commands give comfortable results, the production can be set to "Complete" and considered Closed.

The LHCb computing model for dst files states that each file should be in:

- 2 archives - one being at Cern
- 4 replicas - one being at Cern

-- RobertoSantinel - 07-Jun-2011

This topic: LHCb > ProductionClose

Topic revision: r6 - 2011-08-19 - PhilippeCharpentier



Copyright &© 2008-2020 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.

Ideas, requests, problems regarding TWiki? Send feedback