LHCb Grid Shifters NEWSLETTER

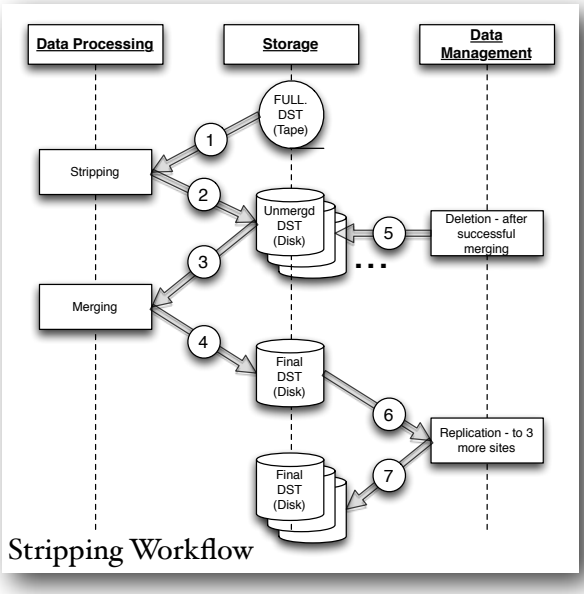
The fall 2013 Incremental Stripping campaign has just started

In this Volume
- Incremental Stripping
- Dirac CLIs explained

The incremental stripping campaign has started on October 3. We will process some 800k Reco14 FULL.DST files of both 2011 and 2012 data, starting with 2012. The campaign is expected to run for 6-8 weeks. In this volume of the newsletter I would like to concentrate on this campaign and explain how to monitor it.

Workflow

Let's start with the workflow of this stripping. The campaign will ***only run at CERN + T1 sites***. The input data for the campaign are FULL.DST files containing the reconstructed event information and will be staged from the T0/T1 tape systems. As we have only short jobs we will stage a lot of data in short time and put a ***lot of stress on the tape systems***. In the past we did experience problems in this area and this needs to be monitored carefully. 2 FULL.DST files will be processed by one Stripping job, running for ~ 6 hours, and output one "unmerged DST" file per physics stream (9 streams in total). Once enough unmerged files in a given stream and for a particular run have accumulated for a 5 GB file a Merging job (runs for ~30 minutes) will collect those files and put them together. The output of the merging job is a "merged DST" file. After the successful merging the unmerged DST files will be removed and the merged DST file will be replicated to another 3 sites.



Stripping Workflow

Monitoring Tools

A few tips on how this campaign could be monitored

Shared Shifter Plots: Those will allow you to check the staging performance on the different sites (bottom left plots in the "Shifter - Data Transfer" page) and the job success rates and WMS overall states of the Stripping and Merging productions (e.g. plots "Shifter - DataStripping - Job Execution" and "Shifter - DataStripping - WMS", same for merging).

Shifters Dashboard: On the dashboard page (https://lhcb-shifters.web.cern.ch/dashboard#Productions) the production managers will put the request numbers and production numbers of the stripping campaign. You should expect 2 productions (Stripping/Merging) per polarity and year, so 8 productions in total. Please check also for the "Hot" flag in the table which will show you whether these productions are still regarded as important to monitor or not.

Production Monitoring: Select only those requests which are marked in the shifters dashboard as "hot", e.g. https://lhcb-web-dirac.cern.ch/DIRAC/LHCb-Production/lhcb_shifter/Production/LHCbTransformationMonitor/display?requestID=16968,16969. The page will show you the how many files have been processed and the jobs in the different states (similar to the WMS monitoring above)

Overall progress page: This page (http://lhcbproject.web.cern.ch/lhcbproject/Reprocessing/stats-inc-stripping-fall13.html) will show you the overall progress of the campaign, i.e. how many files have been processed, whether we are behind or ahead of target and how many pb-1 have been merged. The page also shows the target number of files and pb-1 for the whole campaign.

How to monitor?

A few tips on what to check for

- Job success rate for both stripping and merging jobs, e.g. via the "Shifters - DataStripping - Job Execution" plots. With pattern matching you may find out if a particular site, production etc. is affected by a problem.
- Staging is the most important and critical operation. It can be followed on the "Staging" plots of the "Shifters - Data Operations" plots.
- In the WMS plots e.g. (Shifters - DataStripping - WMS) you may follow the jobs in the different states, we shall always have enough jobs in staging, waiting, running. Especially if the staging is going down (even on only a few sites) it means we probably have exhausted the number of current input files and the run range for the productions need to be extended.

Dirac CLIs explained

Following a suggestion I got after the first volume I would like to introduce this new section, i.e. explain or bring to your attention some Dirac commands that could be useful for shifts.

dirac-transformation-debug can help you debugging data processing or data management productions. There are many options to play with. Just one word of caution the "--FixIt, --FixRun and --KickRequests" will alter states in Dirac so need to be handled with care and preferably left to the GEOCs ;-).

A few examples:

- Get the files in MaxReset (or any other status) of a given production

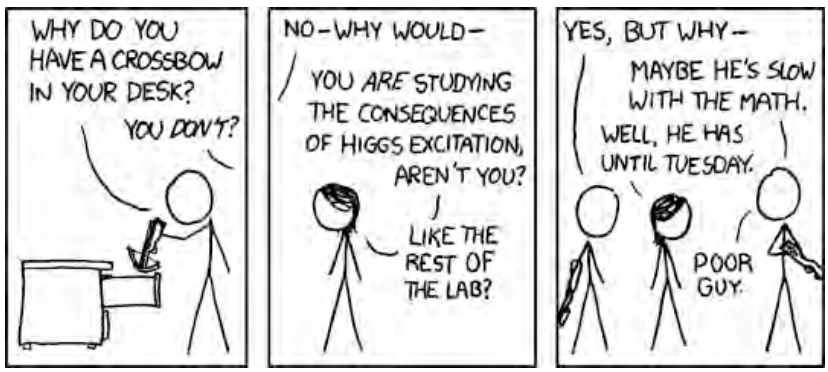
```
dirac-tranformation-debug
--Status MaxReset <production(s)>
```

- Find out which jobs have worked on a given file which is now in MaxReset

```
dirac-transformation-debug -l <lfn>
--Info jobs <production>
```

- Reset files which have been in Assigned for long again to Waiting

```
dirac-transformation-debug
--KickRequests <production>
```



http://xkcd.com (???)

In case you have any comments or suggestions for improvements please let me know.

Also if you have any issues during your shifts please don't hesitate to contact me.

Cheers and happy shifting

LHCb Computing Operations Meeting at 11:30 (CEST) on Monday, Wednesday, Friday in Room 2-R-14