

# Reprocessing

## Small reprocessing requests for 2016

<https://twiki.cern.ch/twiki/bin/view/Atlas/Special2016Reprocessing>

## Release 21 validation

<https://twiki.cern.ch/twiki/bin/view/Atlas/Release21DataReprocessingValidation>

## Reprocessing run 1 in latest releases

<https://twiki.cern.ch/twiki/bin/view/Atlas/Run1Reprocessing>

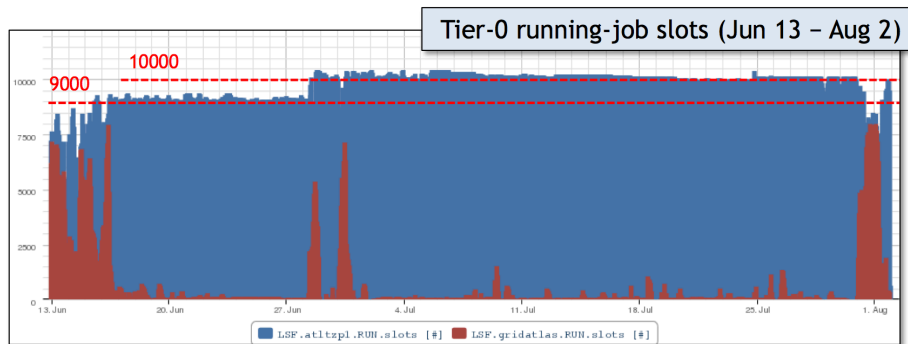
David South (DESY)  
Kostas Nikolopoulos (Birmingham)

ADC session, S&C Week  
September 27th 2016



# 2016 has been busy for data reprocessing

- > .. despite not (yet) running a full campaign on the 2016 data



- > Many more small campaigns run on the grid than usual, due to high Tier-0 occupancy

- And because we can:  
“the system works..”

TWiki > Atlas Web > DataPreparation > DataPreparationReprocessing > Special2016Reprocessing (2016-09-26, DavidSouth)

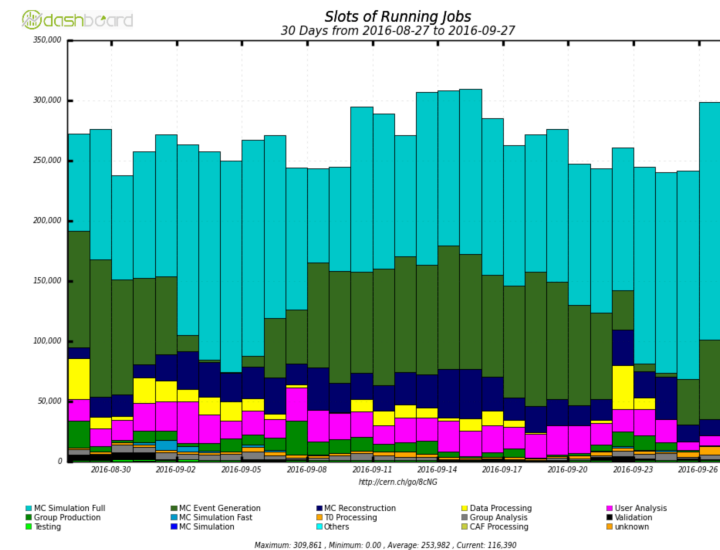
## Special2016Reprocessing

- 1) **DONE** Run production on background stream for 283270  
↓ Available datasets
- 2) **DONE** Low-pt data reconstruction request for pixel dE/dx calibration  
↓ Available datasets
- 3) **DONE** Reprocess 299184 with different pixel settings  
↓ Available datasets
- 4) **DONE** Produce missing IDTIDE derivation for runs 278880-279928  
↓ Available datasets
- 5) **DONE** Reprocess 302391 with different pixel settings  
↓ Available datasets
- 6) **DONE** Reprocessing run 302737 with online alignment constants  
↓ Available datasets
- 7) **DONE** Reprocessing run 302737 with different reconstruction setup  
↓ Available datasets
- 8) **DONE** Reprocessing of run 301973 with all pixel information  
↓ Available datasets
- 9) **DONE** Reprocessing of the VdM stream from the PbPb run 287594  
↓ Available datasets
- 10) **DONE** Production of ESD for run 302829  
↓ Available datasets
- 11) **DONE** Production of DAOD\_IDTRKVALID for run 305543  
↓ Available datasets
- 12) **DONE** Testing and validation of the new ID alignment COOL folders using run 304128  
↓ Available datasets
- 13) **DONE** Processing of 2016 data scouting stream for trigger-level analysis  
↓ Available datasets
- 14) **DONE** DRAW\_ZMUMU reprocessing with updated magnetic field maps  
↓ Available datasets
- 15) **DONE** Processing of physics\_MinBias stream of run 305359 (doMinBias, AFP)  
↓ Available datasets
- 16) **DONE** Processing of physics\_MinBias stream of run 305359 (doHIP)  
↓ Available datasets
- 17) **DONE** Reprocessing of 2015 machine development run 277458  
↓ Available datasets
- 18) **DONE** Processing of physics\_MinBias stream of run 305359 (doHIP)  
↓ Available datasets
- 19) Processing of calibration\_PixelBeam stream of VdM runs 299315, 299390 and 300287 (IBL off)
- 20) **DONE** Test automatic processing of DRAW\_RPVLL  
↓ Available datasets

# Release 21 validation

- > On top of that we are regularly running rel 21 validation with each new release
  - One run from 2015, one from 2016
  - For release 20.7, and the coming release 21, this means:  
**AOD HIST DRAW\_ZMUMU DRAW\_EGZ DRAW\_TAUMUH DRAW\_EMU DESDM\_SGLEL  
DESDM\_SLTTMU DESDM\_MCP DESDM\_CALJET DESDM\_PHOJET DESDM\_EGAMMA  
DAOD\_IDTIDE DRAW\_RPVLL DESDM\_EXOTHIP**
  - And then: **DESDM\_ZMUMU DAOD\_ZMUMU DESDM\_EGZ DAOD\_EGZ DAOD\_TAUMUH  
DESDM\_EMU DAOD\_EMU**
  - As well as testing: **Reco\_tf, AODMerge\_tf, ESDMerge\_tf, HISTMerge\_tf**

- > So all of this this creates regular data reprocessing activity visible on the dashboard



## What's still to come this year

- > We cannot currently use **RAWtoALL** (4GB/core vs 2.2GB/core), but will follow this up if the event forking issue can be resolved  
<https://its.cern.ch/jira/browse/ATEAM-341>
- > There is still a chance the fast reprocessing of a fraction of 2016 will come after the conclusion of pp data taking, towards the end of the year
- > What is certainly to come is the full bulk reprocessing of both the 2016 and 2015 data in release 21
  - This was in the very beginning planned to start essentially now with 2015, to be followed by 2016 and concluded in January
  - It now looks like we will start with 2016 only in January and 2015 will follow that
  - We have a little more time this Winter shutdown, but it still looks very tight
- > This needs to be planned in advance with ADC, for example staging the RAW inputs, of which there are many more than for 2015
- > After all that, as usual there is also the Heavy Ion data to come, and how much of this (in terms of streams) will be reprocessed is still to be clarified

# Reprocessing issues

- > To be honest, most things are fine
- > There is still the issue of error handling of m-core jobs, where much more digging down is required
  - If the 1/8 culprit job could be presented on the front page, this would help us
- > Other than this, monitoring is working well
- > There are times when a job is repeatedly assigned to the same site, with the same configuration
  - Also, at some point reporting of **bad\_allocs**, or jobs that run out of RAM, changed to “**lost heartbeats**” so this may have had an impact here?
- > We are still running with up to 15 retries, and recent campaigns have shown that this long tail is still needed
- > Should probably look at our `project_mode` again, to be sure is all needed
  - The `changeType=DESDM_EGZ:DESDM_ALLCELLS,DAOD_EGZ:AOD;` stuff can be quite cumbersome..

# Back up

# Generic data reprocessing schedule

Meeting today to try to pin this down more

|  |                                    |
|--|------------------------------------|
| Several rounds of Physics Validation   | A few months                       |
| Deadline of conditions updates for data reprocessing   |                                    |
| Frozen release, <b>launch ES1</b> (express and CosmicCalo streams) to get beam-spot derivation and DQ sign off (10 runs) |                                    |
| Express stream HISTs of selected runs available for DQ   | 1 week to get here                 |
| DQ meeting to sign off for bulk launch   |                                    |
| Express stream AODs of all runs ready for beam-spot derivation   |                                    |
| New beam-spot conditions ready and uploaded for all runs   | Another week to get here           |
| Make final conditions tag, <b>launch data reprocessing, all streams and formats</b>                                      |                                    |
| As merged AODs become available, launch derivation production  | 3-4 weeks.. or 5-6 weeks for 2016? |
| Derivations available  | 2 weeks                            |
| As HISTs become available, DQ sign off done on each run  | A few months                       |

5-8 weeks

- > Note that this is for 1 year, i.e. 2015 or 2016, although something like 2016 bulk and 2015 ES1 could run simultaneously (it's a question of resources)