# Shift for run selection, data transfer and data filtering for GSI train

MinJung Kweon

# Histories and status

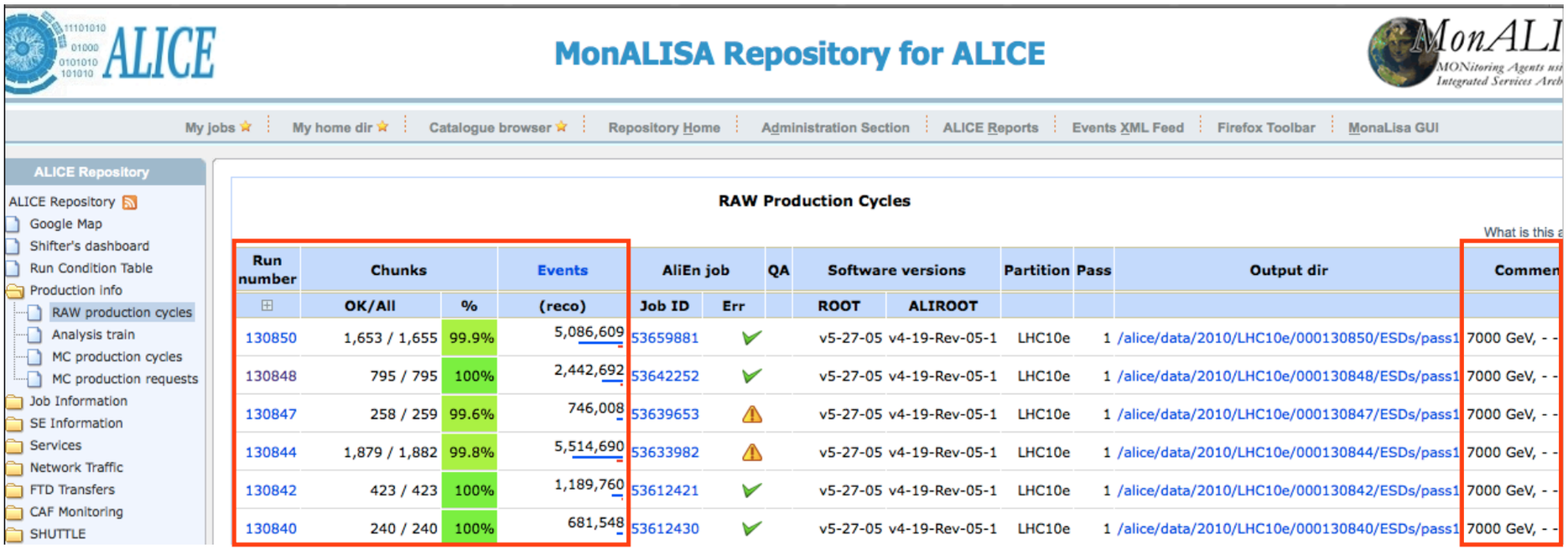
- Many of us are using GSI analysis train, have been participated on data transfer, construct train list... Now people decide to take shift per institute (~ one month) but not per single volunteer.

- September 14 ~ October 12: HD turn (volunteer for initial steps: Oliver, MinJung)

- Our main duties:
    - Run selection
    - Data quantity and quality checks in GSI and train list
    - Generate train list to be updated and data transfer list for missing runs and files
    - Hand investigation if necessary

    $\Rightarrow$ Done via “auto-scripts” developed recently except for last point
- Desired duties:
    - Follow ALICE QA meeting, share information and give feed back on our analysis list

# Run Selections



MonALISA Repository for ALICE

My jobs | My home dir | Catalogue browser | Repository Home | Administration Section | ALICE Reports | Events XML Feed | Firefox Toolbar | MonaLisa GUI

ALICE Repository

- ALICE Repository
- Google Map
- Shifter's dashboard
- Run Condition Table
- Production Info
  - RAW production cycles
  - Analysis train
  - MC production cycles
  - MC production requests
- Job Information
- SE Information
- Services
- Network Traffic
- FTD Transfers
- CAF Monitoring
- SHUTTLE

**RAW Production Cycles**

What is this a

| Run number | Chunks | | Events | AliEn job | | QA | Software versions | | Partition | Pass | Output dir | Commen |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ⊞ | OK/All | % | (reco) | Job ID | Err | | ROOT | ALIROOT | | | | |
| 130850 | 1,653 / 1,655 | 99.9% | 5,086,609 | 53659881 | ✓ | | v5-27-05 | v4-19-Rev-05-1 | LHC10e | 1 | /alice/data/2010/LHC10e/000130850/ESDs/pass1 | 7000 GeV, - - |
| 130848 | 795 / 795 | 100% | 2,442,692 | 53642252 | ✓ | | v5-27-05 | v4-19-Rev-05-1 | LHC10e | 1 | /alice/data/2010/LHC10e/000130848/ESDs/pass1 | 7000 GeV, - - |
| 130847 | 258 / 259 | 99.6% | 746,008 | 53639653 | ⚠ | | v5-27-05 | v4-19-Rev-05-1 | LHC10e | 1 | /alice/data/2010/LHC10e/000130847/ESDs/pass1 | 7000 GeV, - - |
| 130844 | 1,879 / 1,882 | 99.8% | 5,514,690 | 53633982 | ⚠ | | v5-27-05 | v4-19-Rev-05-1 | LHC10e | 1 | /alice/data/2010/LHC10e/000130844/ESDs/pass1 | 7000 GeV, - - |
| 130842 | 423 / 423 | 100% | 1,189,760 | 53612421 | ✓ | | v5-27-05 | v4-19-Rev-05-1 | LHC10e | 1 | /alice/data/2010/LHC10e/000130842/ESDs/pass1 | 7000 GeV, - - |
| 130840 | 240 / 240 | 100% | 681,548 | 53612430 | ✓ | | v5-27-05 | v4-19-Rev-05-1 | LHC10e | 1 | /alice/data/2010/LHC10e/000130840/ESDs/pass1 | 7000 GeV, - - |

- 1st retrieve **run info.** from **RAW Production Cycles**

**Run Condition Table**

| LHC10e | Beam | | | | | Triggers | | | | | | Quality | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Run# | Bunches | Scheme | Fill # | Energy | Intensity per bunch | MB Interaction | Rate (Hz) | MB Beam-Empty | MB Empty-Empty | Muon Interaction | High multiplicity trigger | Global quality | Mean vertex | Vertex quality | Comment | Field |
| 130850 | 36 | 1000ns_50b_35_14_35 | 1,309 | 3,500 | 0.90 | 4,093,491 | 417.70 | 88,743 | 1,990 | 469,617 | 235,101 | 2 | | | One partition in C00... | 1 |
| 130848 | 36 | 1000ns_50b_35_14_35 | 1,309 | 3,500 | 0.90 | 1,954,669 | 413.25 | 41,456 | 868 | 231,437 | 116,100 | | | | TPC trip | 1 |
| 130847 | 36 | 1000ns_50b_35_14_35 | 1,309 | 3,500 | 0.90 | 598,528 | 410.51 | 12,474 | 283 | 70,386 | 34,719 | | | | TPC trip | 1 |
| 130844 | 36 | 1000ns_50b_35_14_35 | 1,309 | 3,500 | 0.90 | 4,394,998 | 414.58 | 93,030 | 2,038 | 535,839 | 270,926 | | | | TPC trip | 1 |
| 130842 | 36 | 1000ns_50b_35_14_35 | 1,309 | 3,500 | 0.90 | 937,961 | 416.32 | 19,624 | 418 | 121,789 | 62,204 | | | | stop because Luminos... | 1 |
| 130840 | 36 | 1000ns_50b_35_14_35 | 1,309 | 3,500 | 0.90 | 543,185 | 415.28 | 11,795 | 212 | 65,675 | 33,297 | | | | | 1 |
| 130834 | 36 | 1000ns_50b_35_14_35 | 1,309 | 3,500 | 0.90 | 1,035,763 | 415.64 | 23,849 | 438 | 122,337 | 62,094 | | | | | 1 |
| 130833 | 36 | 1000ns_50b_35_14_35 | 1,309 | 3,500 | 0.90 | 21,325 | 73.28 | 509 | 11 | 6,892 | 3,466 | 5 | | | | 1 |
| 130831 | 36 | 1000ns_50b_35_14_35 | 1,309 | 3,500 | 0.90 | 204,999 | 351.63 | 5,007 | 113 | 37,457 | 20,755 | 4 | | | lumi scan | 1 |
| 130804 | 36 | 1000ns_50b_35_14_35 | 1,308 | 3,500 | 0.90 | 361,686 | 392.28 | 7,714 | 172 | 39,225 | 19,468 | 2 | | | TPC: some data is mi... | 1 |
| 130803 | 36 | 1000ns_50b_35_14_35 | 1,308 | 3,500 | 0.90 | 2,519,354 | 417.39 | 52,846 | 1,215 | 293,482 | 144,991 | 2 | | | TPC: some data is mi... | 1 |
| 130802 | 36 | 1000ns_50b_35_14_35 | 1,308 | 3,500 | 0.90 | 959,457 | 330.62 | 19,613 | 422 | 116,620 | 57,758 | 2 | | | TPC: some data is mi... | 1 |
| 130799 | 36 | 1000ns_50b_35_14_35 | 1,308 | 3,500 | 0.90 | 1,409,469 | 420.36 | 34,583 | 1,098 | 176,850 | 87,316 | | | | | 1 |
| 130798 | 36 | 1000ns_50b_35_14_35 | 1,308 | 3,500 | 0.90 | 372,009 | 282.47 | 9,980 | 267 | 52,204 | 25,945 | | | | | 1 |
| 130795 | 36 | 1000ns_50b_35_14_35 | 1,308 | 3,500 | 0.90 | 9,386,341 | 415.88 | 318,020 | 10,424 | 1,356,273 | 680,756 | | | | TPC trip | 1 |

- 2nd select runs based on **Global quality** flag from **Run Condition Table**, and retrieve other infos.
  - GOOD (flag " ", 1)
  - INV (flag 2, 3)

# Data quantity/quality check in GSI and Train lists

- Make GSI ESD files list and check ESD files sanity based on Marian's toolkit (run via batch mode)
  - Categorize to good chunks and bad chunks, then count
- Count number of chunks in train list(from Train SVN)
- Compare # of OK chunks in GSI and # of chunks in train list, then generate train list to be updated if these two number are different
- Generate summary table

```
==============================
 Good Runs for LHC10e pass1
==============================
```

| runNo | Quality | chunks all[ML] | chunks rec[ML] | chunks %rec[ML] | chunks ok[GSI] | chunks %ok[GSI] | chunks bad[GSI] | chunks train | chunks %train | train list | energy | field | events rec[ML] | min-bias trigger |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 127719 | GOOD | 600 | 598 | 99.7 | 598 | 100.0 | 0 | 597 | 99.8 | 7TeV | 7000 | 0.5 | 2,843,225 | 2,690,482 |
| 127724 | GOOD | 600 | 597 | 99.5 | 597 | 100.0 | 0 | 597 | 100.0 | 7TeV | 7000 | 0.5 | 2,581,870 | 2,500,913 |
| 127729 | GOOD | 1,125 | 1,117 | 99.3 | 1117 | 100.0 | 0 | 1115 | 99.8 | 7TeV | 7000 | 0.5 | 4,442,891 | 4,294,480 |
| 127730 | GOOD | 250 | 245 | 98.0 | 245 | 100.0 | 0 | 245 | 100.0 | 7TeV | 7000 | 0.5 | 980,896 | 966,046 |
| 127814 | GOOD | 490 | 469 | 95.7 | 469 | 100.0 | 0 | 469 | 100.0 | 7TeV | 7000 | 0.5 | 1,345,544 | 1,107,977 |
| 127815 | GOOD | 1,298 | 1,270 | 97.8 | 1269 | 99.9 | 1 | 1269 | 100.0 | 7TeV | 7000 | 0.5 | 4,108,607 | 3,392,854 |
| 127817 | GOOD | 1,374 | 1,327 | 96.6 | 1327 | 100.0 | 0 | 1327 | 100.0 | 7TeV | 7000 | 0.5 | 4,198,355 | 3,497,988 |
| 127819 | GOOD | 125 | 125 | 100.0 | 125 | 100.0 | 0 | | | | 7000 | 0.5 | 320,254 | 253,785 |
| 127822 | GOOD | 525 | 521 | 99.2 | 521 | 100.0 | 0 | 521 | 100.0 | 7TeV | 7000 | 0.5 | 1,663,086 | 1,323,117 |

# Data Transfer List

- Generate missing runs and file list(check done by comparing alien file lists)
  - Runs for run-mode transfer
  - Missing files for list-mode transfer
- Check BAD ESDs(failures of Marian's toolkit) to categorize them to transfer failure and central production itself problem(check done by comparing file size in alien)
  - Transfer failure files for list-mode transfer
  - Write down "central production problem" candidate files

# Cleaning List

- List of empty folders
- Runs not in the "selected run list": luminosity scan, standalone and short runs.

## Hand Investigation

- Runs with “INV” flag(Global quality flag 2, 3)
- Cleaning list(once by hands before deleting them)
- Single detector problems written in the run condition table → should be propagated into analysis based train output merging step

## Outlook and Discussion

- Scripts are in train SVN  (STEER/datatransfer)
- Make version for MC data
- What we can share
  - Works requiring “Hand Investigation”(First part for LHC10d,e shared by Yvonne already)
  - Full procedure once we are familiar with all the steps
  - Following ALICE QA meeting, sharing information and give feed back on our analysis list



**PS. We in HD will copy only ESDs for good runs. Currently LHC10b pass2 copy ongoing**