

Data Preparation Group

Recent activities

ALICE Offline Week, July 19th 2017

F. Prino, C. Zampolli for the DPG coordination

Disclaimers

- The following slides are the result of the assiduous and hard work of all the DPG members
 - Coordination, A. Barbano, J. Bohm, M. Concas, M. Ivanov, I. Sputowska, C. Terrevoli, A. Toia...
- Names won't be explicitly quoted to not forget anyone - unless some material was prepared by someone in particular
- The cooperation with Offline, BTG, detectors and PWGs is a key ingredient for the success of our activities

Outline

- DPG **organization**
- Data and Monte Carlo **(re)processing**
- Anchored **pass** stored in MC
- **AOD** merging issue
- **RCT**
- Overview of **AOT**
- **Other** main activities
- **Service tasks**
- Summary and conclusions

DPG Coordination

Chiara Zampolli
Francesco Prino

DPG Coordination



PROC

Processing

Roberto Preghenella/
Michael Weber

Catalin-Lucian Ristea (*)
Chiara Zampolli



detectors

QAT

QA and Tools

Marie Germain, E. Botta
Jacek Otwinowski (*)



detectors

PWGS

AOT

Analysis Objects and Tools

Catalin-Lucian Ristea (*),
David Dobrigkeit
Chinellato, Andrea Dainese
& Andrea Rossi (*)



detectors

PWGS

Names and photos in alphabetical order
(*) Institutional responsibility

Re-Processing

- 2016 pp data reprocessing started end of February, completed at the beginning of June
 - Prediction from last Offline Week was end of May
 - N.B.: data taking in parallel from ~mid/end May
 - No major show-stoppers
 - Some issues with the grid, CASTOR, and a bit of delay in the QA/manual calibration for the last period
 - Different periods done in parallel
 - **Wonderful job of the production team, QA, and detectors**
- 2015 pp data reprocessing started beginning of June
 - Few issues with data staging
 - As data taking is ongoing, only one period at a time
- LHC15o (PbPb) low IR and LHC15n (pp @5 TeV) reprocessed with improved TOF calibration

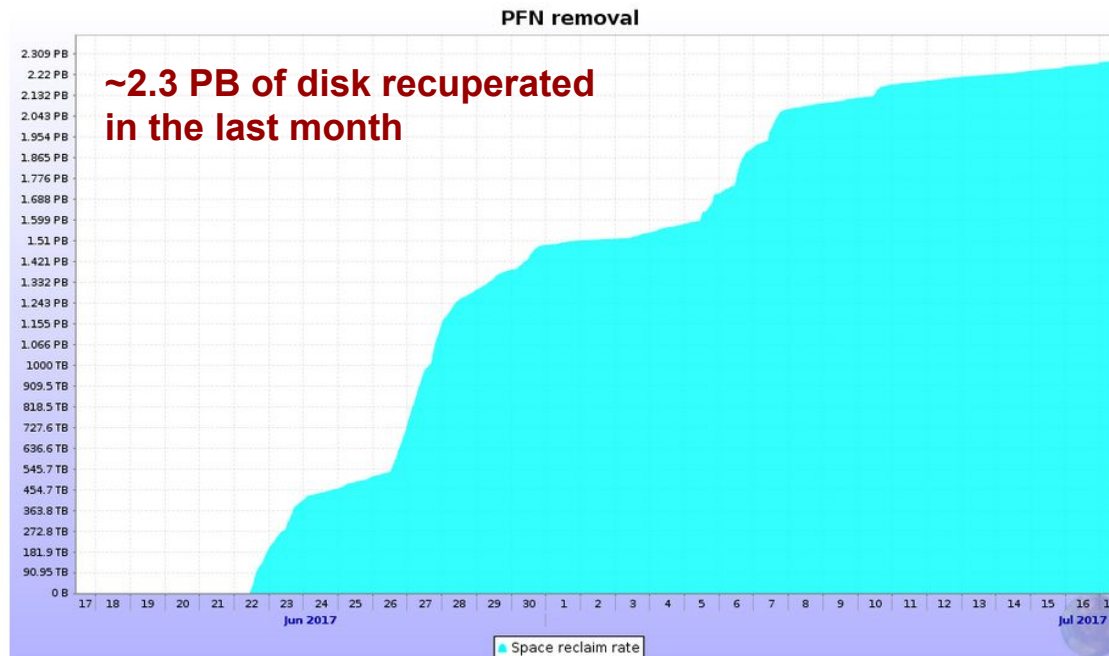
Processing

- From mid/end May, 2017 data are processed:
 - LHC17c, LHC17d (B=0 → no PPass foreseen), LHC17e, LHC17f, LHC17g (low B), LHC17h, LHC17i (ongoing)
 - **CPass0/CPass1 and muon_calor running automatically**
 - **PPass1** done for 17c, 17e, 17f, ongoing for 17g, ready to go for 17h
 - **Missing: new alignment** (work ongoing - R. Shahoyan - some issues with TRD)
 - Until new alignment, quality will be the same as for 2016 data
 - From new alignment, improvement in the tracking performance
 - Advisable to re-do 17g (low B), as it is the only period with low B, when new alignment is ready
 - LHC17h: data for new HLT CF validation (see reco session)

- AOD refiltering was requested by Physics Board (especially DQ, HF, GA) with the following goals:
 - have in the additional variables (AODs the golden chi2, number of TPC clusters and number of TPC tracks with KITSout) which are used in event and track selection
 - fix an issue with vertexing HF delta AODs
 - improve PCM delta AODs
- PbPb data (AOD186) and MC samples (AOD188) done
 - Feedback from **LF and UD so far** → **All good!**
- Pb-p 8 TeV (LHC16s) redone (AOD190)
 - Using updated AliRoot which fixes an issue with the ZDC information in the AODs (affecting centrality when using ZNA estimator)
 - **Validated!**
- Next: p-Pb 8 TeV (LHC16r) and 5 TeV (LHC16q,t) and corresponding MCs

Data and MC cleanup

- Huge effort to recuperate space both from data productions and MC simulations not used in the last year and not used for any physics results (no papers, no preliminaries)
 - In conjunction with Physics Board (PWGs) and Offline (Costin)



PFN removal				
Series	Last value	Min	Avg	Max
1. Space reclaim rate	2.278 PB/s	8.959 TB/s	1.564 PB/s	2.278 PB/s

Data and MC cleanup

- Huge effort to recuperate space both from data productions and MC simulations not used in the last year and not used for any physics results (no papers, no preliminaries)
 - In conjunction with Physics Board (PWGs)
 - ~2.3 PB of disk recuperated in the last month
 - In the process, 20 runs from LHC11d pass1 removed
 - Effort to re-process these data ongoing

Monte Carlo productions

- **Huge and successful** Monte Carlo campaign in the last months for the EPS and SQM conferences
- General purpose productions always started by the DPG after a dataset is reconstructed (25% sampling, PYTHIA8)
- Not many PWG-specific requests at the moment...
- ...but ~60 cycles foreseen for GA (under discussion in terms of optimization)

C. Ristea

07/17/2017 09:57:38

Summer productions

Production	Details				Total			Progress				Notes
	JIRA	Requester	#cycles	Anchor	Events(M)	CPU(d @ 10k)	Disk (TB)	Status	Events done(M)	Percent done	CPU left (d @ 10k)	
-	ALIROOT-7271	PWGGA	15	16rs / 13def	800	132	844.8	OPEN	0	0	132	DPMJET
-	ALIROOT-7270	PWGGA	4	16qrst	320	52.8	337.9	OPEN	0	0	52.8	DPMJET
-	ALIROOT-7269	PWGGA	24	all pp 2016	320	63.4	415.4	OPEN	0	0	63.4	PYTHIA8
LHC17#8	ALIROOT-7268	PWGGA	11	all pp 2016	160	31.7	207.7	10% QA	54.2	31	21.9	PYTHIA8
-	ALIROOT-7267	PWGGA	6	12cdthi	160	31.7	207.7	OPEN	0	0	31.7	PYTHIA8, G3/G4
-	ALIROOT-7337	PWGGA	-	-	-	-	-	OPEN	0	0	0	PHOS Embedding into RAW ESDs
LHC17#9	ALIROOT-7313	Common	1	16e	17	1.9	7.1	Setup	0	0	1.9	AD
LHC17#2,3,4]b_fix	ALIROOT-7100	Common	9	16qrst	124.5	19.2	126	RUNNING FULL	40.4	29	13.6	new EPOS-LHC
LHC17d20b	ALIROOT-6873	Common	2	16kl	78.1	7.7	36.1	RUNNING FULL	8.2	10	6.9	new EPOS-LHC
LHC16j4	ALIROOT-6867	PWGDQ	2	16kl	56.8	5.6	28.7	RUNNING FULL	27	43	3.2	diamond vbx update
Common			12			28.8	169.2			22	22.4	
PWGDQ			2			5.6	28.7			43	3.2	
PWGGA			60			311.6	2013.5			3	301.8	
Total			74			346	2211.4			5	327.4	
RUNNING FULL											23.7	

MC anchored to 2016

LHC16d	DONE
LHC16e	MC starting
LHC16f	DONE
LHC16g	DONE
LHC16h	DONE
LHC16i	DONE
LHC16j	DONE
LHC16k	TPC clarification
LHC16l	DONE
LHC16m	At the very end, maybe
LHC16n	Won't be done (rare triggers only)
LHC16o	DONE
LHC16p	DONE

Productions in FinalQA

		Last week	Now	QA since... (only for productions in QA since long)/comment
Gen Purpose	2016 pp	1	0	since 18.06
	2016 pPb	0	0	
LF		5	3	since 04.05, 29.05, 04.06
HF		0		
DQ		3	1	since 06.06
G4		1	1	Detector QA DONE
Data	muon_calorimeter 2017	3	2	
	2016 pp	4	2	since 01.05, 11.04
	2017 pass1	3	3	
	2015 pass2	1	1	
Tot		21	13	

Huge work by the QA and detector teams!
42 productions in finalQA [1 month ago](#)

How to configure a MC generator with AliDPG

- **Updated TWiki:** <https://twiki.cern.ch/twiki/bin/view/ALICE/AliDPGMonteCarloTutorial>
- There are different possibilities for the generator configuration for running *alroot_dpgsim.sh*:
 - **general purpose configurations:**
 - defined in *MC/GeneratorConfig.C: GeneratorName[kNGenerators]*
 - example: `--generator Pythia8_Monash2013`
 - **PWG specific configurations:**
 - called via *PWGXX:MacroName[:optional argument to be passed]*
 - macro has to be located in *MC/CustomGenerators/PWGXX/MacroName.C*
 - example: `--generator PWGHF:Pythia6_Perugia2011_HF001:fv0`
 - **Custom generator:**
 - called via `--generator Custom`
 - macro *GeneratorCustom.C* has to be located in working directory
- For further information, comments, suggestions, ... **please contact us.**

M. Weber

Anchored pass in MC

- **Data:** information about pass is taken during reconstruction from the JDL and stored in the ESDs (AliProdInfo)
 - It is then used at analysis time to decide which TPC splines and in general which OADB object to load for the PID
- **Monte Carlo:** equivalent information not available in the JDL (and ESDs)
 - TuneOnData and splines not properly treated/loaded
- **Ongoing:**
 - **Force JIRA tickets** to contain the information about the pass name (data) and anchored pass (MC) to be then used by MonALISA to fill a dedicated JDL variable to be then used by AliProdInfo
 - The JIRA ticket has to be specified for the MC production
 - Automatically this can be read by the MC database
 - Drawback: Will need care when several productions/cycles are in the same JIRA ticket

AOD merging issue

- An analyzer spotted some issues with the AODs from the pPb MC productions
- Investigations lead to the conclusion that in some cases the merged file is corrupted
- Possible to spot this at merging time → validation script will now catch this
- Maybe an issue with input file (SE temporarily misbehaving, corrupted replica accessed...)
- No issue or memory problem found when running locally
- ROOT didn't trigger any real error, allowed the process to go on
- Analysis of old logs:
 - a. ~0.3‰ AOD merged files with issues found (among the “inspectable” ones)
 - 0.03‰ from alidaq (data)
 - 0.5‰ from aliproduct (MC)
 - b. No correlation with AliRoot/ROOT versions
 - c. No production affected more than the others
- Follow-up:
 - a. Delete the problematic files

Many thanks to Costin!

Upcoming changes in the RCT

- Goal:
 - Simplify the usage (both for the detector experts, the QA experts, and the analyzers) of the RCT table
 - **Currently:** several flags (up to ~50!) defined by each detector; no defined prescriptions when needing a new flag (--> detectors were simply adding it); information about specific responses (PID, trigger) of detectors embedded in the flags
 - New: reserve specific flags for specific cases (e.g.: flags > 100 → not full acceptance); separate specific responses (PID, trigger) from general detector status; allow free (searchable) text

Quality Pass 1								Detectors configuration																							
Field	A	A	E	F	H	H	M	M	P	P	S	S	S	T	T	Field	A	A	E	F	H	H	M	M	P	P	S	S	S	T	T
Field	C	D	M	M	L	M	C	T	H	M	P	D	D	D	D	O	O	C	A	D	T	P	R	S	D	D	D	D	F	C	
2,097	x	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	x	1	1	1	1	1	1	1	1	1	1	1	1	1	
2,334	x	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	x	1	1	1	1	1	1	1	1	1	1	1	1	1	
5,766	x	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	x	1	1	1	1	1	1	1	1	1	x	1	1	19	
8,030	x	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	x	1	1	1	1	1	1	1	1	1	1	1	1	1	

Change configuration of EMC for run# 265308, pass 1	
Run range:	265308
Configuration:	1 (All OK)
Editor:	29 (2 RCU missing) 30 (half RCU SM1C) 31 (1 RCU missing SM2) 32 (MB ok EMC ->care) 33 (2 RCU missing) 34 (1 RCU SM4 missing) 35 (EMC trig accept + RCU miss) 36 (noise in SM6) 37 (Pb in mult) 38 (not in pass2) 39 (BAD pedestals not properly sets) 40 (pending (LED/time)) 41 (pending) 42 (physics 2) 43 (no SM7 data) 44 (OK but part of detector should be masked)

RCT flagging used to create Runlist

M. Germain

Detector Column General convention:

- Flag "1": GOOD
- Flag < 0 : data not usable: bad data for this pass
- Flag ≥ 100 : Limited acceptance: data are OK BUT incomplete coverage
- **1 free text comment to store run/run information**

Same convention should be used for:

- **Detector Trigger Column:** 1, negative (<0) , acceptance(≥ 100)
- **Detector PID column:** 1, -1
- Was asked to detectors to propose a remapping with respect to those conventions;
- 1st round made for all detectors to insure backward compatibility;
- Still few points/detectors to clarify (TRD/TPC)
- Implementation discussion with developers (Costin) to be finalized in summer
- **Implementation expected early september**
- Future for RUN3:
 - Some proposals/discussions started (need of automatization in QA)

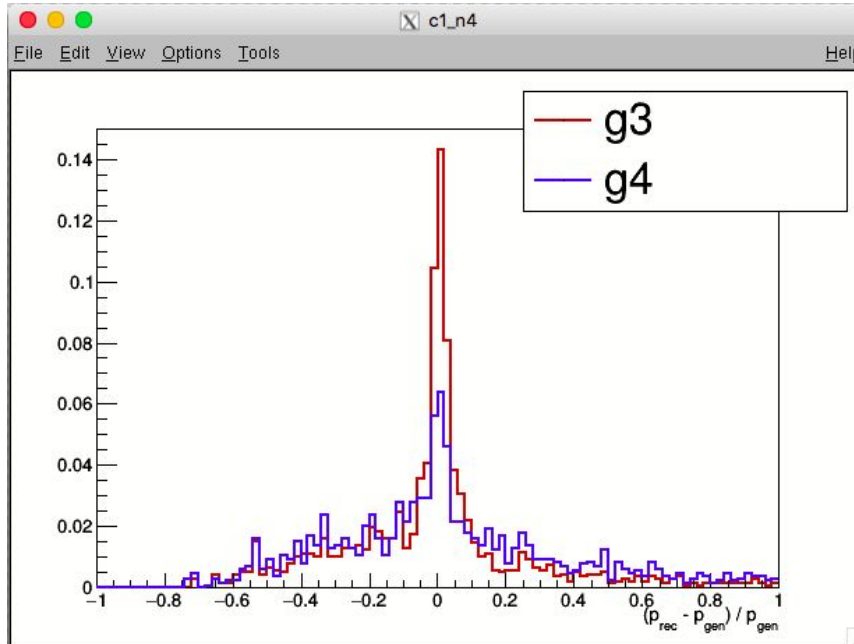
QA tools

- Several topics discussed and addressed
 - Performance generator, Database, Release validation including Monte Carlo, dashboards...
- more in the **next talks**

- Finalization of the studies for the **systematic uncertainty on the TPC-ITS matching efficiency** for the pp, p-Pb and Pb-Pb data at 5 TeV
 - Used for EPS and SQM results
 - Available at [link](#)
- Studies of the tracking **efficiency for primary particles vs Monte Carlo generator** (consequence of different particle composition)
 - Ongoing, see [PWGMM-21](#)
- Issue with **EPOS** spotted → resulted in some fixes in AliRoot, and re-running of some Monte Carlo productions

- Studies on **pileup**
 - Changes in the default for cuts for pileup rejection using SPD
 - To reduce the impact of false positives in p-Pb and high multiplicity pp collisions
 - Changes in default value for the maximum event multiplicity (number of SPD clusters) for which the search for multiple vertices with SPD is activated at reconstruction time
 - Extend pileup tagging with SPD to higher multiplicities
- Studies on **centrality**
 - ZNA estimator in pPb had issues due to wrong ZDC information in the AODs
 - Fixed with refiltering (see previous slides)

- **G4** → see next talk + issue with kinks still under investigation



e.g.: kaon-kink daughter momentum resolution much worse in G4 than G3

- **MC-to-MC embedding** → new tests ongoing
- **Analysis QA** on AODs and ESDs → two service tasks
- **QA-tools** → see next talks



- QA-tools: **Development of a trending and alarm framework for online HLT QA**
 - Several QA components on the HLT provide QA data for different detectors at discrete times as simple 1 or multi-dimensional histogram. These data are visualized using the overwatch web application. The scope of the task is to implement a component processing the histograms and extracting trending information. These trending information will be sent to an elasticsearch database for visualization. In addition, by comparing to limits defined by users, automatic alarms should be raised (E-Mail to detector responsible with alarm message + Log in database). The alarm handling should be in testing phase during the pp data taking period end of 2017.
- AOT-tracks: **Systematic monitoring of matching, overall efficiency and impact parameter in data and Monte Carlo**
 - The student should take care of the following tasks: 1) study and monitoring of TPC-ITS matching efficiency and its uncertainty 2) monitoring of overall (i.e. reconstructed/generated) efficiency in Monte Carlo 3) monitoring of impact parameter distribution in data and MC Activity 1) may include some development/new study while for tasks 2) and 3) one needs just to check plots produced automatically. The specific data samples to be covered will be defined in itinere and, up to a certain extent, may also be discussed and agreed with the student on the basis of her/his interest/thesis work. Our preference would be that the student takes care of the service task for 1 year at 50%, but we can consider different proposal.

Please, contact us if you are interested (or know someone who might be!)

Summary and conclusions

- A broad variety of activities and topics covered by the DPG
 - Trying to simplify the work of the analyzers
 - Providing as much information as possible to the relevant people
 - Maintaining documentation up-to-date as much as possible
- DPG cooperation with Offline, detectors, PWGs and analyzers very useful
 - Helps to spot issues, fix them, improve the processing

A huge thank you to all that contribute to our activities, and especially to the DPG coordination for their constant commitment!

BACKUP