Custom merging per component - shell scripts and JDL

ALICE Offline week

Alice analysis on ”train”

ALICE analysis/calibration/derived data preparation done usually on trains

- CPass0
- CPass1
- QA
- Lego trains
- AOD filtering

Several “wagon” tasks connected in “train” to optimally use computing resources

Data merging.

Data merged at the end of the process, per run, per period

- Standard merging tools works well for standard data for which data volume is independent of the number of events (THn, list of THn)
- Standard tools merge per files - problems in case of different containers THnSparse, TTree which scales with Nr of events
  - THnSparse - data volume as function of the Nevent - saturation curve
  - TTree - linear scaling with the number of events

Extension of the standard tools or new tool for merging needed

- Forbidding THnSparse and TTree not solution
- ThnSparse and TTree are really needed

Merging has to be parameterized

Numbers: Consideration for the TTree containers

N runs - O(10^4)

Nchunks - O(10^2-10^4) per run

Nevents (pp)

- O(10^4-10^7) events per run
- O(4x10^7) events per event (1kHz x 10hours)

TPC space point calibration tree - 10-40 GBy per run/fill considered

- per cluster residuals
  - per fill (10^5-4x10^7) (events) x 10 (tracks) x 10^2 (clusters) x (~20 bytes)- (10^9-4x10^11 clusters)
    - we are considering 10-100 Hz of the calibration trigger
- 10 floats, 1 int
  - position: r,z,phi, time
  - angle: q/pt
  - residuals: deltay, deltaz TPC-ITS, TPC-(ITS+TRD), TPC-(ITS+TOF)
    - TPC-(ITS+TRD)/TPC-(ITS+TOF) - precision ~2 times better

ATO-126: THnSparse/THn:Distortion and efficiency maps

For the TPC local distortions very local

- Smooth distortion close to the boundaries of E field holding elements - tau ~2 (1D structures) -10 cm (2D structures)

Distortion maps to be sampled with precision comparable with the tau

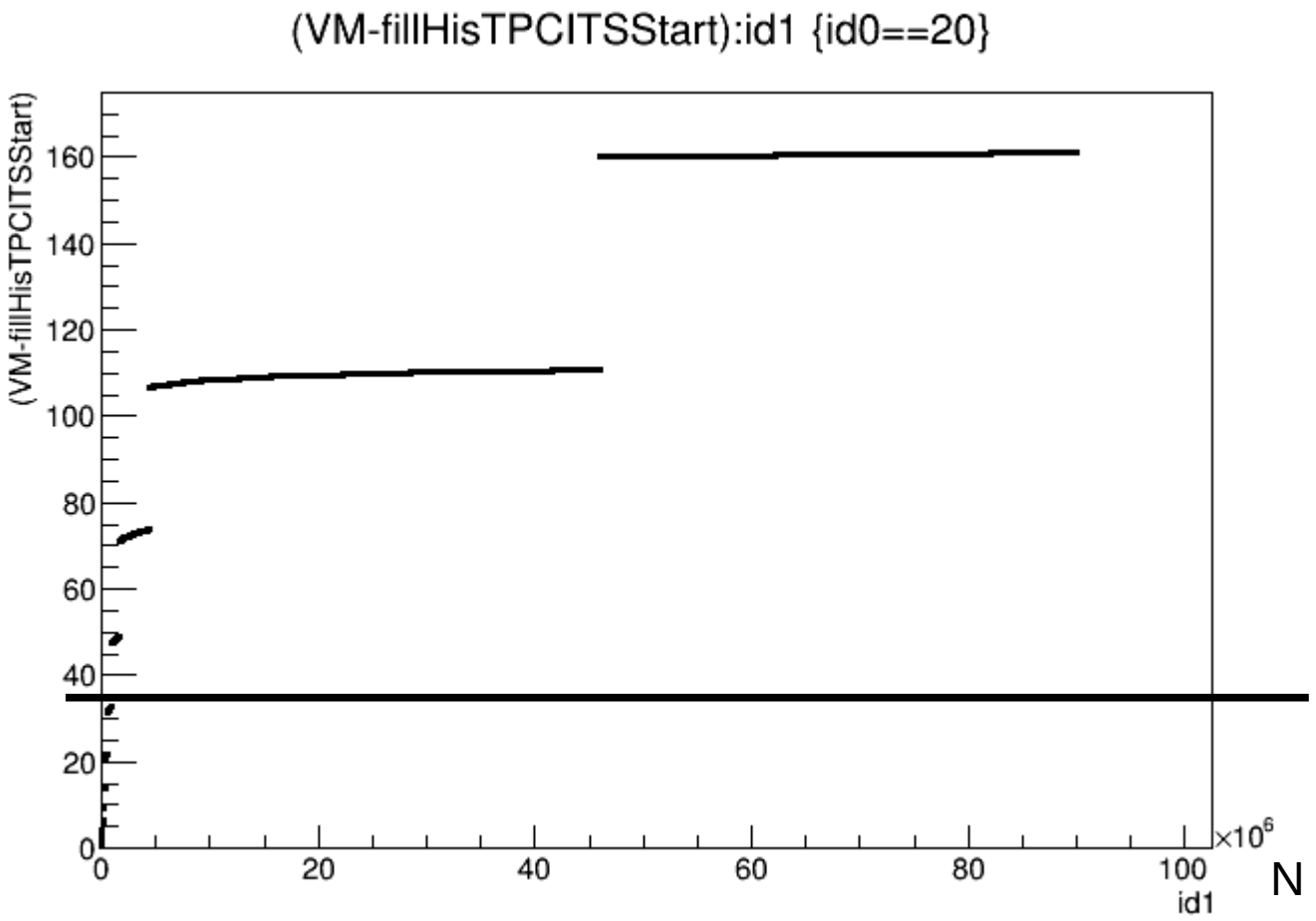
CPass0/Cpass1 calibration distortion map example

AliTPCCalibTime Booking Results:

- 1.5 GBy needed in case of THn representation (5 (hisCE)*219.34375 (MBy)+18.26953125(MBy)*20. (detector))
- almost -0 for the ThNSparse representation
- We can not use such representation in the calibration train

ATO-126: merging THnSparse/THn comparison. Memory

- CPU time for merging and VM for the data representation is depending on the histogram occupancy nbins>0
  - for THn constant by quite high ~ nBins
    - coefficient proportional to number of axis to represent



Current tools (MI understanding)

- hadd
  - per file
  - sequential
- TFileMerger
  - per file
  - sequential
- AliFileMerger
  - per component
  - sequential
- Lego train merges
  - recursive
  - per file
- User defined tools on top of standard tools
  - parallel per component merging done e.g for the CPass0/CPass1

ALIROOT-5727: Proposal: command-line and jdl interface for custom merging

Current hadd

```
Usage: hadd [-f[0-9]] [-k] [-T] [-O] [-n maxopenedfiles] [-v verbosity] targetfile source1 [source2 source3 ...]
```

Marian Ivanov added a comment - 10/Dec/14 1:03 PM - edited

New (ali)hadd

```
Usage: alihadd [-f[0-9]] [-k] [-T] [-O] [-n maxopenedfiles] [-v verbosity] targetfile source1 [source2 source3 ...] -p
pattern
where pattern is a regular expression describing list of objects
Example pattern:
TPCcalib
TPCcalib/calibTime/
TPCcalib/calibTime/fResHistoTPCITS[5];
```

alihadd.jdl - interface for merging on grid

- can be based extension of the existing AliFileMerger.C

Extension - additional parameters:

- time outs to access data files - accepted efficiency
- local asynchronous caching or asynchOpen (is it still working)
- max disk space - splitting parameter (for trees)

aliadd usage example

- Lego train
  - Users should be able to configure merging process
  - PWGPP AddHocCalibration train example:
    - StandardQA (THn) - merged per run
      - full statistic needed
    - ExpertQA - FilteredTrees.root (TTree*)
      - merged in chunks <nThr, e.g 2 GBy
      - eff threshold can be configured
- AOD filtering for the expert QA:
  - currently too many small chunks of the FilteredTrees.root
  - ~ 2 MBy => to be merged in groups ~2GBy
    - eff. threshold >95 %
- CPass0/CPass1 automatic calibration
  - TPC space point calibration - merged in groups <1-2GBy
  - TOF and T0 calibration trees …
    - presently done by hand
  - following calibration procedure to be also automatic

End