



Condor scripts update

R. Zarrella



Condor scripts, where were we?

DATA

shoe/Reconstruction/scripts/runShoeBatchT1.sh

- Experimental data processing
- Documentation inside the file
- Possible to process multiple runs (even all!!) for a single campaign

Arguments

-i	Input directory (in /storage/gpfs_data/foot/shared
-o	Output directory (in /storage/gpfs_data/foot/\${USER})
-c	Campaign name
-r	First run number
-l	Last run number (optional)
-m	Merge output files (optional, default "0")

MC

shoe/Reconstruction/scripts/runShoeBatchT1_MC.sh

- MC files processing
- **1 job = 20k events**
- **Campaign and run number are retrieved automatically!**
- Possible to process multiple files for a single run (-f option)

Arguments

-i	Input file (in /storage/gpfs_data/foot/shared/SimulatedData)
-o	Output directory (in /storage/gpfs_data/foot/\${USER})
-m	Merge output files (optional, default "0")
-f	Use full statistics (optional, default "0")

All jobs submitted together → no guarantees that merge starts after processing!

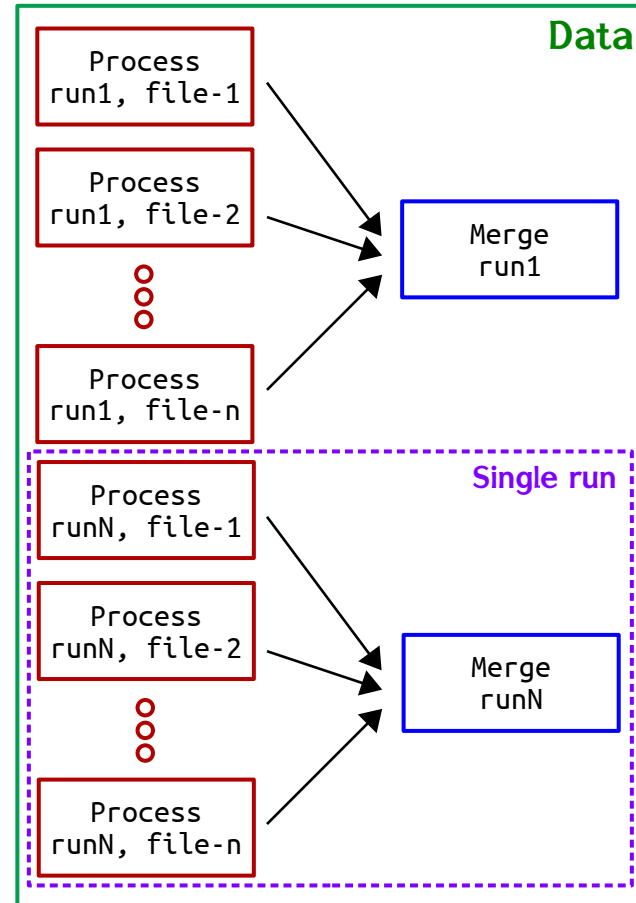
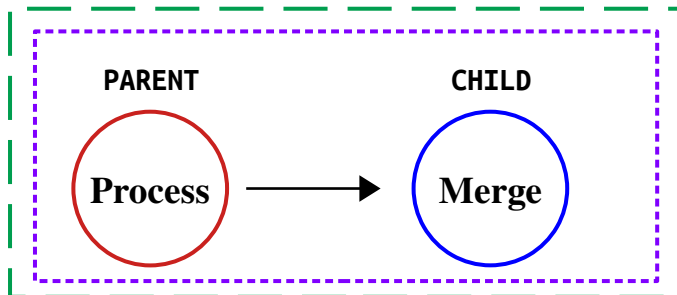
Timeout needed...



DAG (Directed Acyclic Graph)

Job flow control directly inside condor!

- Possible to decide the order of job submission
- No time-out anymore!
 - ➔ Merge killed if processing fails
- Comes with a couple of features:
 - ➔ Auxiliary files (.out/.err/.log) automatically downloaded
 - ➔ Completed jobs automatically removed
- Different scheme for data and MC

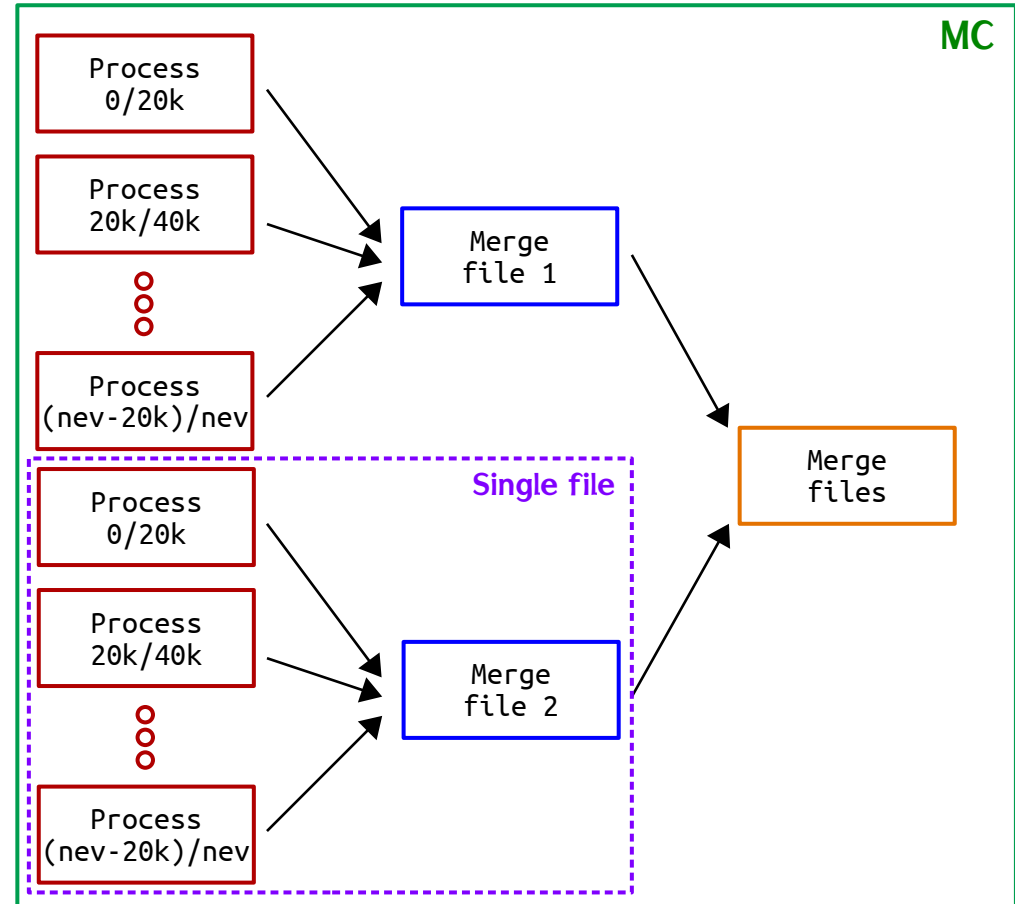
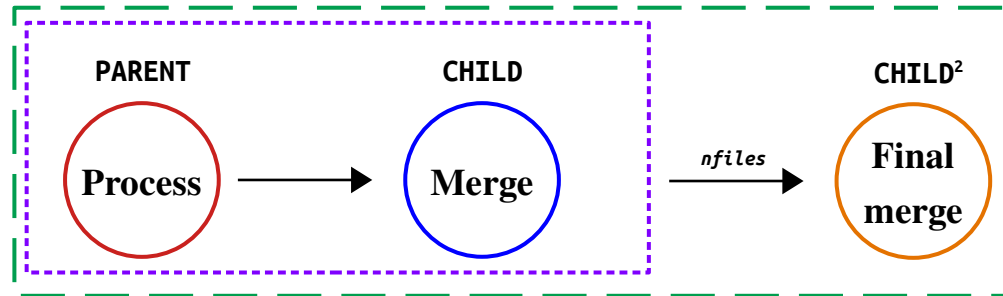




DAG (Directed Acyclic Graph)

Job flow control directly inside condor!

- Possible to decide the order of job submission
- No time-out anymore!
 - ➔ Merge killed if processing fails
- Comes with a couple of features:
 - ➔ Auxiliary files (.out/.err/.log) automatically downloaded
 - ➔ Completed jobs automatically removed
- Different scheme for data and MC



Condor scripts update!



DATA

`shoe/Reconstruction/scripts/runShoeBatchT1.sh`

- Experimental data processing
- Documentation inside the file
- Possible to process multiple runs (even all!!) for a single campaign

Arguments

<code>-i</code>	Input directory (in <code>/storage/gpfs_data/foot/shared</code>)
<code>-o</code>	Output directory (in <code>/storage/gpfs_data/foot</code>)
<code>-c</code>	Campaign name
<code>-r</code>	First run number
<code>-l</code>	Last run number (optional)
<code>-m</code>	Merge output files (optional, default “1”)

MC

`shoe/Reconstruction/scripts/runShoeBatchT1_MC.sh`

- MC files processing
- **1 job = 20k events**
- **Campaign and run number are retrieved automatically!**
- Possible to process multiple files for a single run (`-f` option)

Arguments

<code>-i</code>	Input file (in <code>/storage/gpfs_data/foot/shared/SimulatedData</code>)
<code>-o</code>	Output directory (in <code>/storage/gpfs_data/foot</code>)
<code>-m</code>	Merge output files (optional, default “1”)
<code>-f</code>	Use full statistics (optional, default “0”)

Internal documentation updated in both scripts!



Production?

Need to agree on a set of reconstruction parameters → **production!**

(default in SHOE repo?)

I/O

EnableTree: y
EnableFlatTree: n
EnableHisto: y
EnableTracking: y

EnableSaveHits: y
EnableRootObject: y
EnableRegionMc: y
EnableElecNoiseMc: ?

Global Reco

EnableKalman: ?
Kalman preselection strategy: ?
N measure in global tracking: ?

IncludeTOE: ?
TOE cuts: ?

IncludeStraight: ?
Parameters: ?

Detectors

IncludeDI: y
IncludeST: y
IncludeBM: y
IncludeTG: y
IncludeVT: y
IncludeIT: y
IncludeMSD: y
IncludeTW: y
IncludeCA: y

From last GM

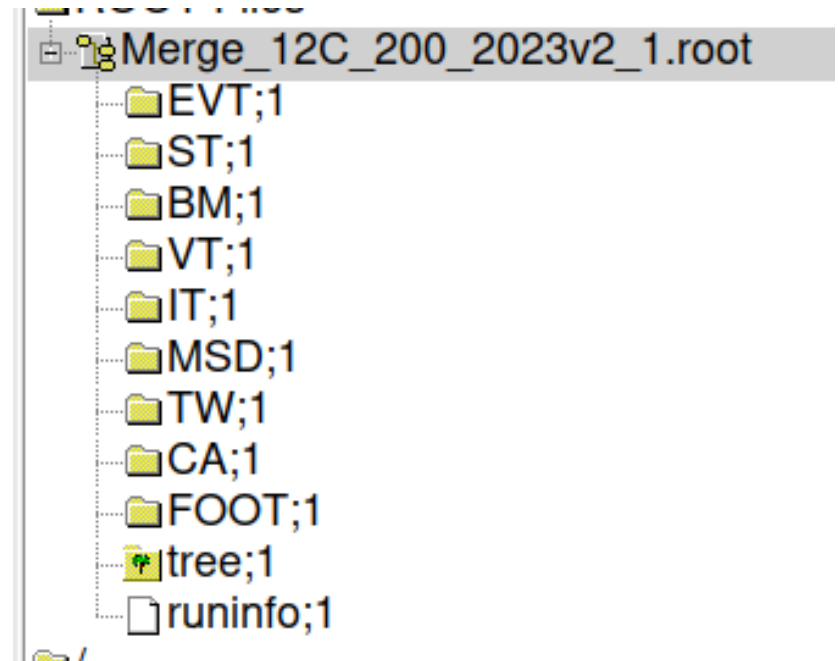


Analysis script for Tier1 in the making...

shoe/Reconstruction/scripts/runAnalysisBatchT1.sh

-i Input file (in /storage/gpfs_data/foot/
-o Output file (in /storage/gpfs_data/foot/, optional)
-m Is MC (optional, default "0")

- No need to indicate campaign and run → from input file
- Need to indicate if campaign is MC (add it to runinfo?)
- Output file name optional
→ default = "input_folder/MergeAna_campaign_run.root"
- Possible to analyze in batch but...
- Output of analysis can contain a lot of objects (histos/folders/trees), **merge can be very slow!!**

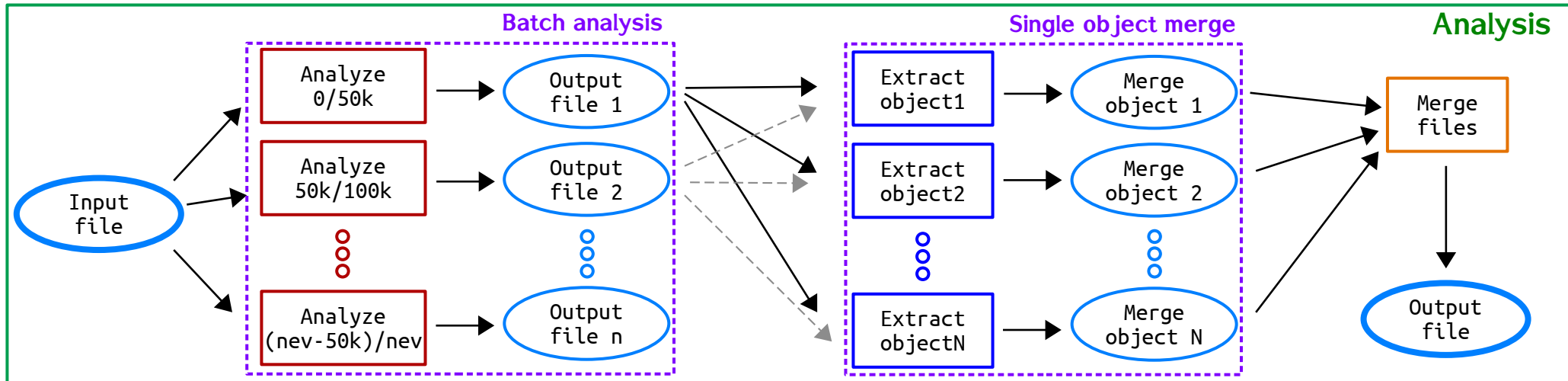
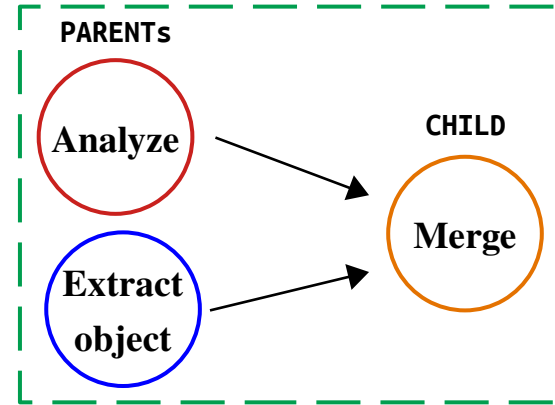


How do we manage that?

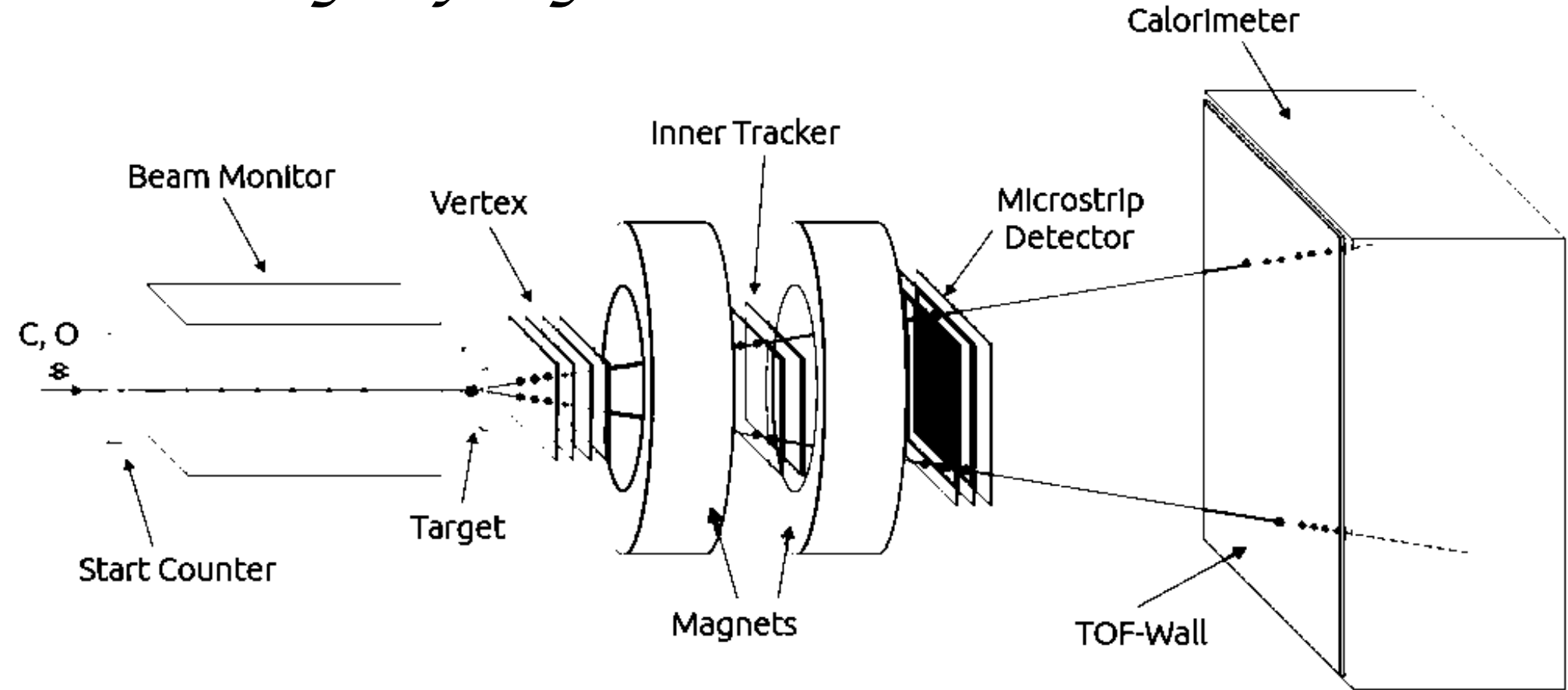


DAG of analysis

- **Split merge of output file for single objects!**
- Merge time is almost independent of # of objects
- Analysis and object-merge start together (faster object extraction from single files)



Thank you for your attention





Backup slides