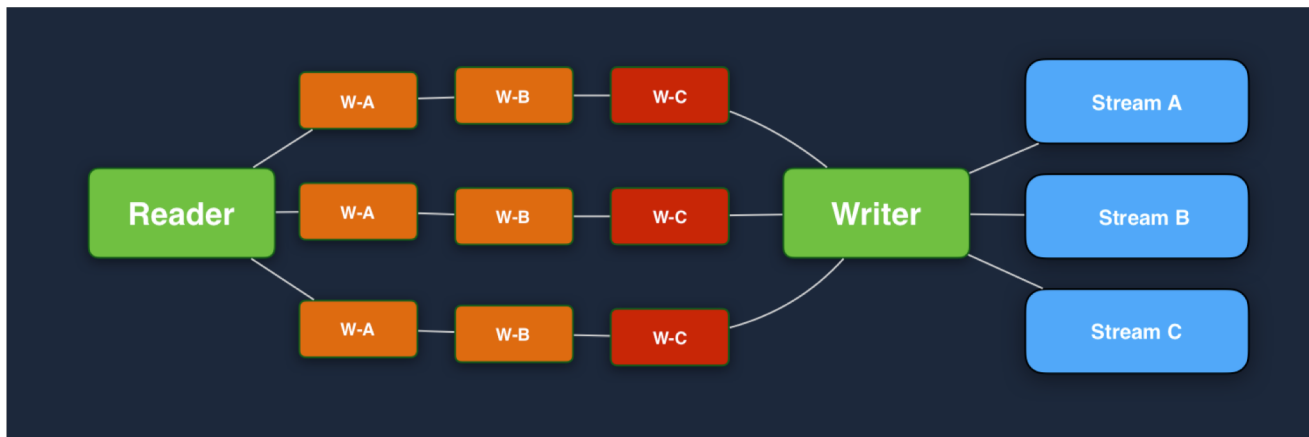


STATUS OF ANALYSIS TRAINS

William Phelps (GWU), Nathan Baltzell, Raffaella De Vita, Gagik Gavalian

What is an Analysis Train?

- Analysis trains are used to efficiently skim large datasets
- Why do we need them?
 - Large (Many TB) datasets can be read one time instead of many times
- Consists of “wagons” as services in CLARA: Multithreaded



Preliminary Benchmarks

- Processing time on one 2016 node
 - ~3 hours per run
 - 2.2 hours per 100M triggers
 - 12K events per second
- The Train took one evening to run on the farm for all DNP DSTs
 - **Quite fast!**
- Total size of skimmed DSTs so far ~2.5 TB

Status of Analysis Trains

- **Inbending data:** Run 4013-4078 (**5b.6.1**, 10% of Spring 2018)
 - 100% complete **including file merging**
 - Added wagon for CND calibration
 - Merged files are located in
/work/clas12/clas12/data/trains/v1/5bp6p1/
- **Outbending data:** Run 3933-3977 (**5b.6.2**)
 - 100% complete **including file merging**
 - Added wagon for CND calibration
 - Merged files are located in
/work/clas12/clas12/data/trains/v1/5bp6p2/

Merged Files/Skims

- One file per run
- One directory per skim
- A readme.txt for basic documentation

File Sizes For Run 4013

245M	skim1_jpsi_tcs/4013_1.hipo
346M	skim2_ft_pi0/4013_2.hipo
25G	skim3_mesonx_vs/4013_3.hipo
14G	skim4_inclusive/4013_4.hipo
3.8G	skim8_ep/4013_8.hipo
1.5G	skim9_ppbar/4013_9.hipo

Directory Structure

```
v0
├── README.txt
├── skim1_jpsi_tcs
├── skim2_ft_pi0
├── skim3_mesonx_vs
├── skim4_inclusive
├── skim8_ep
└── skim9_ppbar
```

File Listing

```
skim2_4013.hipo
skim2_4014.hipo
skim2_4015.hipo
skim2_4016.hipo
skim2_4017.hipo
```

Skim #

Run

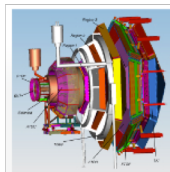
Skim Information

Skim #	Title	Description	Fraction*
Skim 1	J/ψ /TCS	Custom Wagon	0.0052
Skim 2	FT/π^0	$e^- \gamma\gamma$ in FT	0.0078
Skim 3	MesonX/VS	e^- in FT + 2 charged tracks	0.5489
Skim 4	Inclusive	e^- in the Forward Detector	0.3187
Skim 8	$e^- P$	e^- in FD Proton FD/Central	0.0874
Skim 9	$p\bar{p}$	$p\bar{p}X$	0.0321

*Fraction of the total output, not of the input. Numbers will be provided at a later point.

Summary

- Trains have been run over majority of cooked Spring 2018 data
- New web page showing current status that will be kept up to date
- Ready to run trains on reconstructed files!



Navigation

[Main Page](#)
[Community portal](#)
[Current events](#)
[Recent changes](#)
[Random page](#)
[Help](#)

Toolbox

[What links here](#)
[Related changes](#)
[Special pages](#)
[Printable version](#)
[Permanent link](#)

Page **Discussion**

Run Group A

[Status Summary](#)

[Data Processing](#)

Trains

[Useful Links](#)

[Useful Code](#)

Train Introduction

- The Analysis Trains run in CLARA and produce DST skims based on adding "Wagons" to the skim.
 - Analysis train documentation here: [Train documentation](#) 📄
 - The DSTs produced keep only the REC* banks. Documentation is here: [DST documentation](#) 📄

Train Information

- v0 - 9/5/18
 - /work/clas12/clas12/data/trains/v0/
 - First run. **This skim has been replaced by v1.**
- v1 - 9/20/18
 - /work/clas12/clas12/data/trains/v1/
 - Includes a new train for CND calibration . Email sent to clas12_first_exp@jlab.org mailing list.
 - New self documenting notation for the skimmed files: skim1_4013.hipo

```

v1
├── 5bp6p1
│   ├── README.txt
│   ├── skim1_jpsi_tcs
│   ├── skim2_ft_pi0
│   ├── skim3_mesox_vs
│   ├── skim4_inclusive
│   ├── skim7_epiCND
│   ├── skim8_ep
│   └── skim9_ppbar
└── 5bp6p2
    ├── README.txt
    ├── skim1_jpsi_tcs
    ├── skim2_ft_pi0
    ├── skim3_mesox_vs
    ├── skim4_inclusive
    ├── skim7_epiCND
    ├── skim8_ep
    └── skim9_ppbar
  
```

Future work

- We are currently running tests using the new slurm job handling system (replaces PBS)
 - Lower memory footprint
- Please let us know if you need skims that were not listed!
- Regularly scheduled trains in the future



Acela Express ~2011