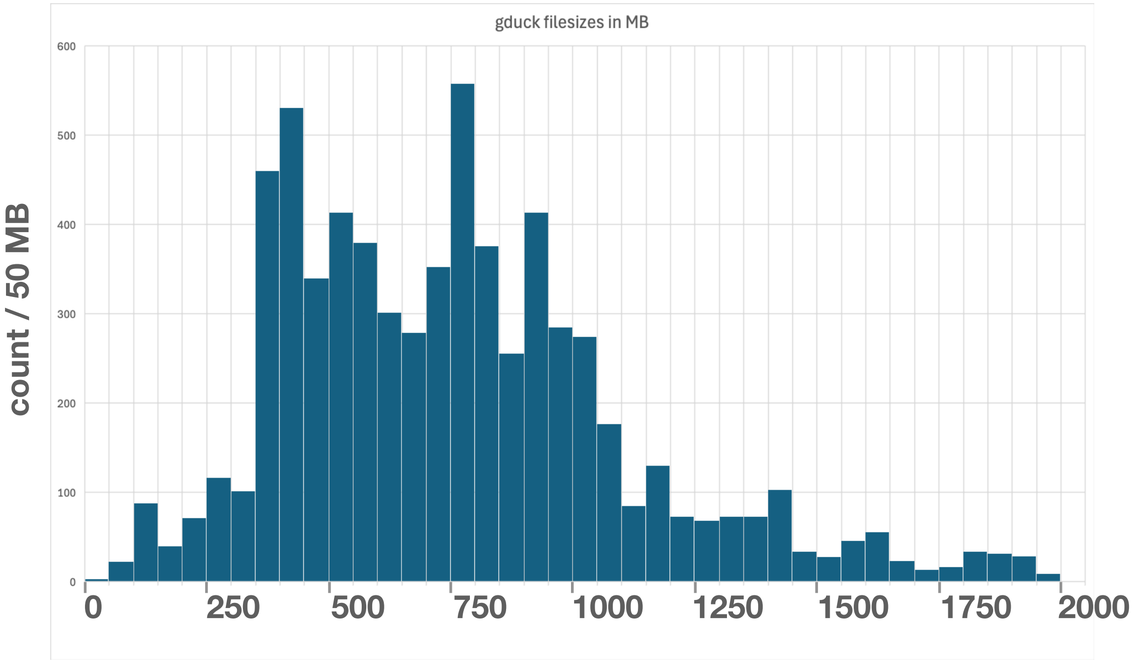
dCache - ARC stress test plans

AHM Spring 2026

Datasets

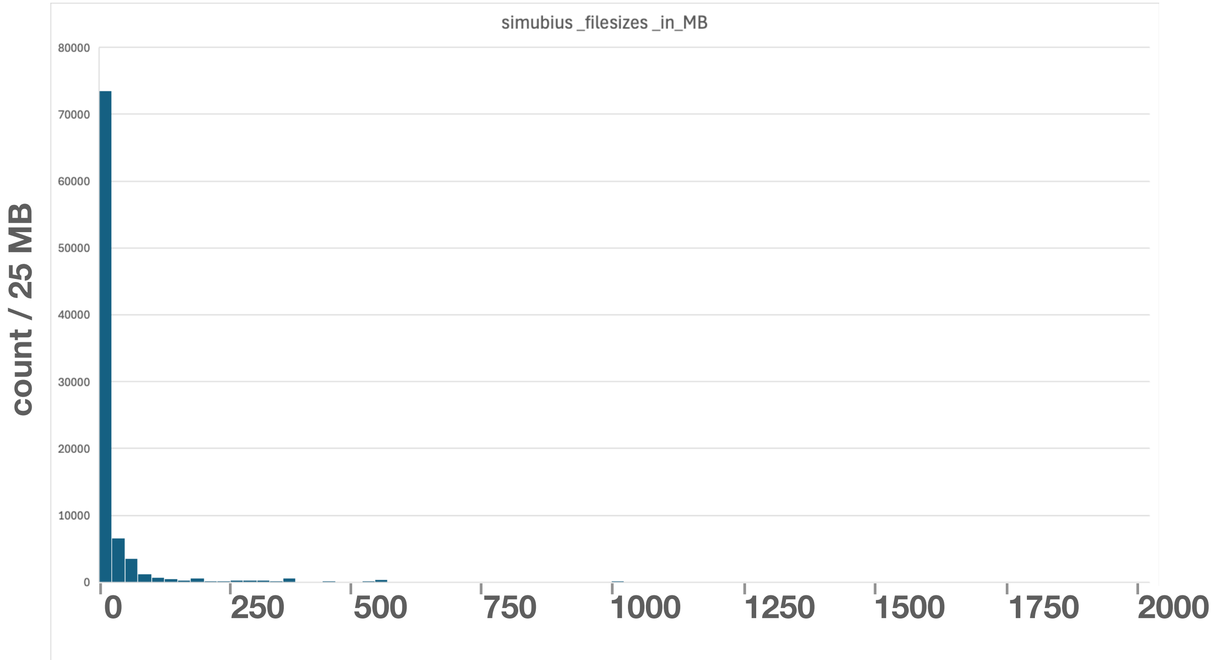


data24 DAOD_PHYSLITE

`user.gduck:data24_13p6TeV.physics_Main.deriv.DAOD_PHYSLITE.NHR_test_1`

```
Total files : 6748
Total size : 5.118 TB (avg.750 MB / file)
Total events : 612,991,779
Size Range: 20 MB to 2030 MB
```

- Reads InDetTrackParticles
- Only a subset of branches used to mimic realistic jobs 20-30 %
- Fills some histograms, output `.root` files, few kB size



Run3 - backgrounds - flat ntuples

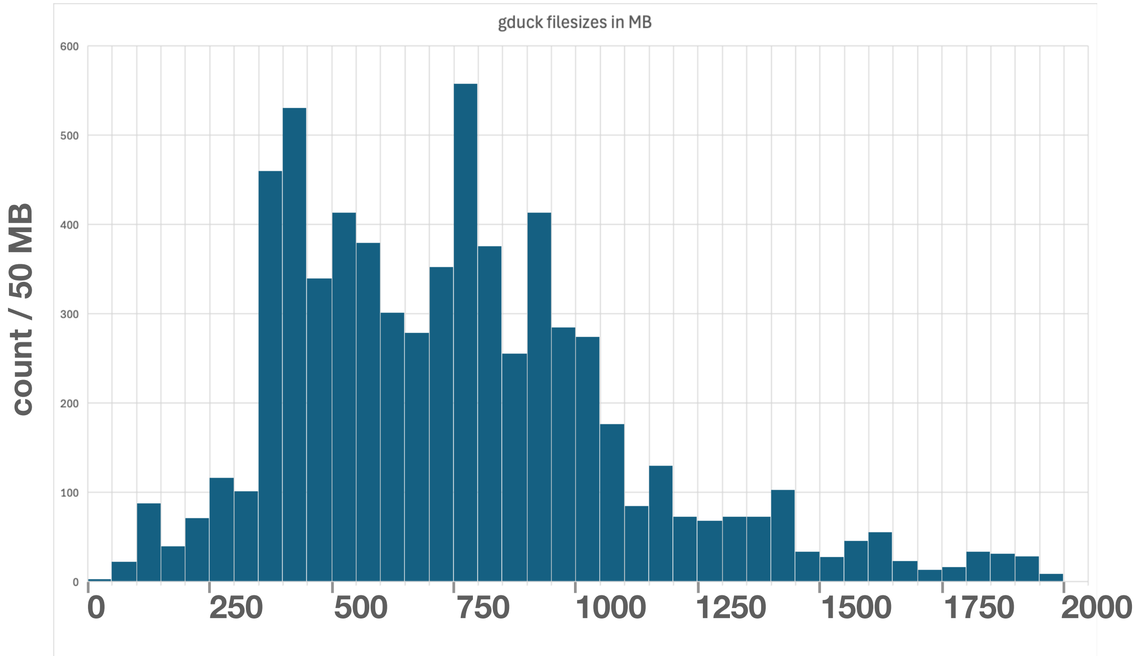
`user.ddaniele:flat_MEL_benchmark_v1`

```
Total files : 90972
Total size : 3.445 TB (avg.38 MB / file)

Size Range: 1 kB to 4000 MB
```

- Reads flat ntuples `.root` files
- applies (close to) final cuts for signal searches

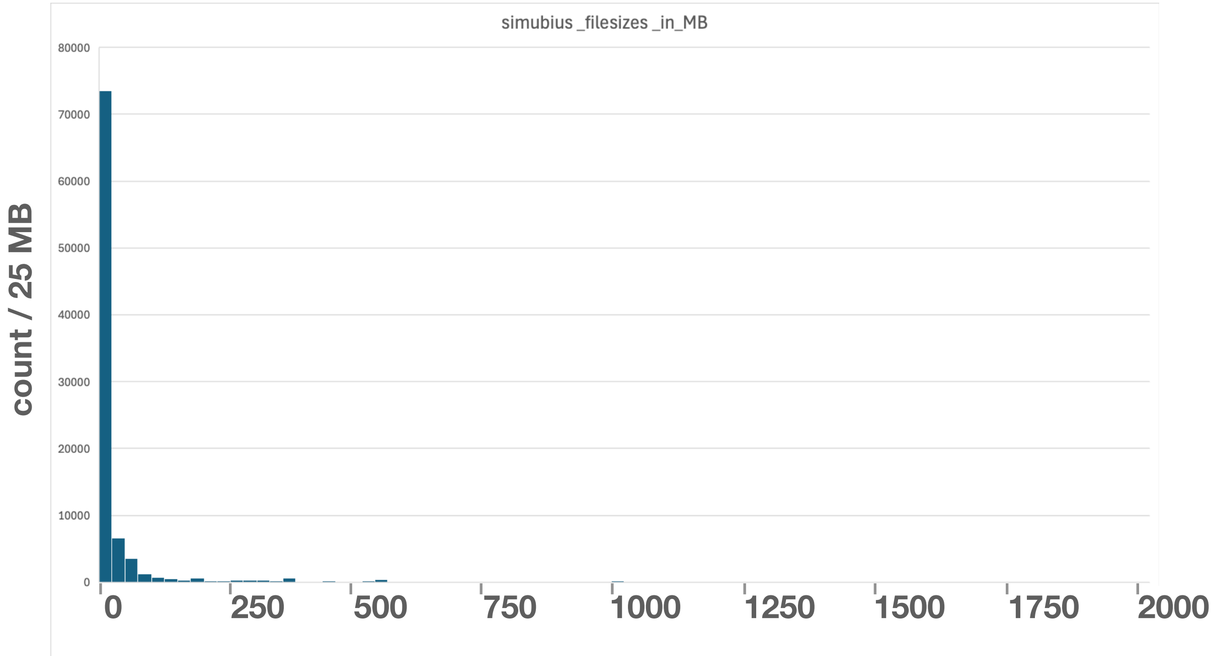
Datasets



DATASET: user.gduck:data24_13p6TeV.physics_Main.deriv.DAOD_PHYSLITE.NHR_test_1

RSE	FOUND	TOTAL
GRIF_DATADISK	162	6748
TRIUMF-LCG2_DATADISK	27	6748
IN2P3-LPC_DATADISK	646	6748
RAL-LCG2-ECHO_DATADISK	513	6748
NDGF-T1_DATADISK	**6748**	**6748**
INFN-T1_DATADISK	2163	6748
FZK-LCG2_DATADISK	6748	6748
LRZ-LMU_DATADISK	5719	6748
LRZ-LMU_LOCALGROUPDISK	6748	6748
NET2_DATADISK	297	6748
UAM-LCG2_SCRATCHDISK	1	6748
CERN-PROD_DATADISK	914	6748

● Active



DATASET: user.ddaniele:flat_MEL_benchmark_v1

RSE	FOUND	TOTAL
UNIBE-LHEP_LOCALGROUPDISK	**90961**	**90972**
AU-MELBOURNE_DATADISK	5	90972
INFN-NAPOLI-ATLAS_SCRATCHDISK	411	90972

● In progress

Tuning the HC tests

- HC test template: 50 input files per job (randomly selected from dataset) → dataset exhausted with 135 jobs
  - can create one test per PanDA queue, with tuneable targets for running and queued jobs
  - for the test duration time (tuneable), HC aims at keeping a ~stable number of running and queued on each queue
- HC creates the jobs → sends them to PanDA which then sends them to the queues:

Test date/ duration	Job template type	Target job running/queued	Test job sent total	Test notes
10 Feb/24h	directIO	6k/4k	1-4k	50 to 80% failed. Duplicated files in the generated jobs (HC bug → fixed)
24 Feb/24h	copy2scratch	6k/4k	None	DNS requests on the HC celery cluster exceeded some threshold → fixed
3 Mar/24h	copy2scratch	6k/4k	7-8k	Many jobs completed, no reliable info on job counts, status, efficiency, etc (*)
9 Mar/24h	copy2scratch	3k/2k	2.5-3k	as above
25 Mar/24h	copy2scratch	6k/4k	8-9.7k	as above

ARNES, DCSC, HPC2N, LUMI, NSC, PUHTI, SiGNET, SiGNET_NSC, UIO_CLOUD, UNIBE-LHEP, Vega

(*) “From the logs it seems that the request to PanDA to get all the job statuses was too big and failed. This probably created a cascade of bad events that should explain the behaviours observed.”

Tuning the HC tests

- 300k file requests per site → 3.3M per set of 11 simultaneous tests
- When no file is in the ARC cache, the first 6748 requests per site go to dCache → 74k within ~1h
- In some cases, some files could still come from other sites, but **`preferredpattern=`** in arc.conf should minimise that
- From there on, ARC bears the load for copying/linking from cache
- With the second dataset we can have a factor 13+ more