ATLAS Data Carousel

Xin Zhao(BNL), Alexei Klimentov(BNL), Mario Lassnig (CERN)

OSG All Hands Meeting
September 4th, 2020

Team effort ---

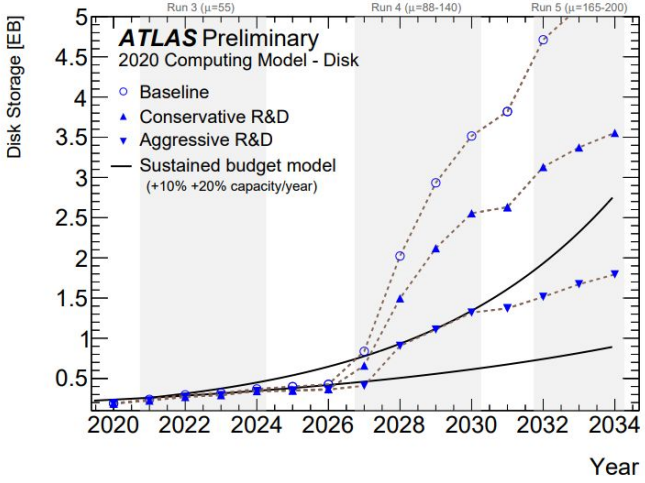
- Workflow management team (WFM)
- Distributed Data Management team (DDM/Rucio)
- Distributed production and analysis team (DPAs)
- Operations team (Grid/T0)
- Monitoring team
- ADC coordinators and experts
- CERN Tier0 and all Tier1 storage and tape experts
- dCache and FTS experts

Outline

- Data Carousel and our objectives
  - What is Data Carousel ?
  - Objectives
  - Three phases
- Progress and improvements
  - Integration with key ATLAS distributed computing components
  - FTS
  - Tape sites
- Exercises and results
- Next steps
- Summary

What is Data Carousel ?

- Facing the data storage challenge in HL-LHC, for the purpose of cost saving, ATLAS started the Data Carousel R&D in the second half of 2018
  - to study the feasibility to use tape as input to various ATLAS workflows, such as RAW data reprocessing and derivation production.



By 'data carousel', we mean an orchestration between workflow management (WFMS), data management (DDM/Rucio), and data transfer (FTS) and data archiving services whereby a bulk production campaign with its inputs resident on tape, is executed by staging and promptly processing a sliding window of X% (5%?, 10%?) of inputs onto buffer disk, such that only a small fraction of inputs are pinned on disk at any given time.

Objectives

- Integration of tape resources with ATLAS distributed computing workflow
  - No more pre-stage campaign
  - Tape systems available at T0 and T1 sites (see the table)
    - Upcoming NESE tape tier at one USATLAS T2 (NET2)
- Use ***available*** tape resources ***efficiently***
  - Metrics : Tape recall efficiency -- ratio of throughput delivered to end users over the vendor-specified nominal tape throughput
  - Examine the various services involved in tape staging, including production system, DDM, FTS, and site SE services, minimize performance penalties to the throughput out of tape system
  - Improving the throughput out of tape itself, by studying file placements on tapes, so called “smart writing”

Sites	tape	SE	Pledge (2019)
CERN	CTA	EOS	100%
BNL	HPSS	dCache	23%
FZK	TSM → HPSS	dCache	13%
RAL	CTA	Echo	13%
CCIN2P3	HPSS	dCache	12%
TRIUMF	Tapeguy	dCache	10%
INFN	TSM	StoRM	8%
NL-T1	DMF	dCache	8%
NDGF	TSM	dCache	6%
PIC	Enstore	dCache	4%

Three Phases

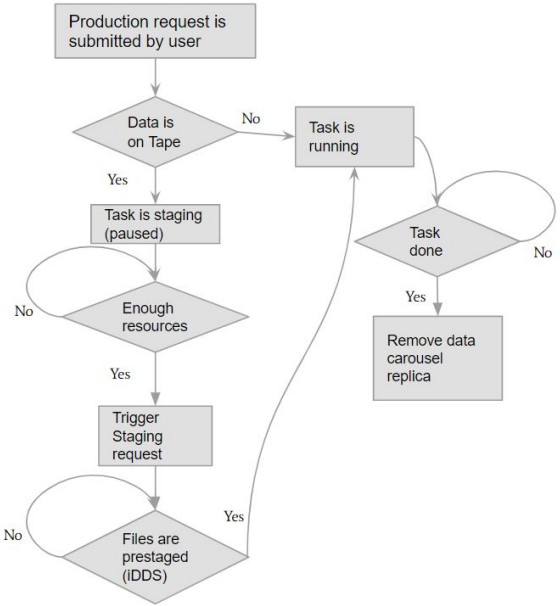
- Phase I : tape site evaluation (Y2018)
  - Conduct tape staging tests, understand tape system performance at sites and find bottlenecks
- Phase II : ProdSys2/Rucio/Facilities integration (Y2019~2020)
  - Address issues found in Phase I
  - Deeper integration between workflow, workload and data management systems, and facilities
  - Identify missing software components
- Phase III : Run production, at scale, for selected workflows (Y2020)

Now we are in the middle of Phase III. The goal is to have Data Carousel in production for LHC Run3.

…… Next we will go through improvements and progress made so far, at various services

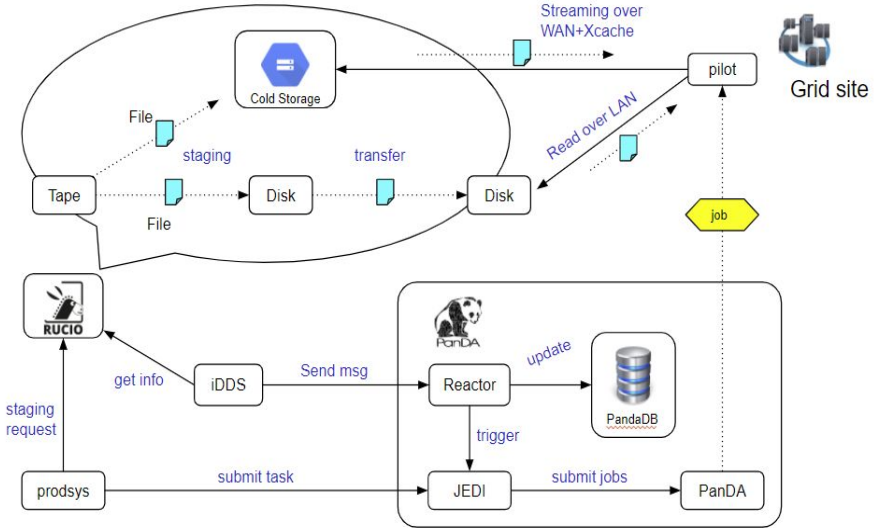
Integration of tape into ATLAS workflow (1/3)

- Changes to the workflow management system (ProdSys2)
  - No more pre-stage campaigns
  - ProdSys2 makes decision on staging requests dynamically, among all tape sites, based on user requests and availability of disk and CPU resources.
- Fine grained communication between ProdSys2 and Rucio
  - Rucio provides notifications to ProdSys2 for each 10% of transfer progress



Integration of tape into ATLAS workflow (2/3)

- Integration with iDDS* (intelligent Data Delivery Service)
- iDDS talks to Rucio to collect and digest file information, and lets JEDI/PanDA process jobs whose inputs are already staged from tape, not waiting for the whole dataset/task to be ready.



** https://indico.cern.ch/event/773049/contributions/3474484/*

Integration of tape into ATLAS workflow (3/3)

- Site staging profile
  - Production systems decides on how many files to stage from each tape sites. On the other hand, sites also tell ATLAS how the staging operation should be conducted, in the “staging profile”
    - Bulk staging requests is preferred, in order to utilize tape system efficiently
    - Avoid overloading tape system and its frontend storage services
  - Site decides :
    - the size of each submission bunch, ie. the number of concurrent active staging requests it can handle at any given time
      - upper limit : 10k~200k ; lower limit : default 5k
    - how the bunches should be submitted.
      - Some sites prefer batch mode, ie. new requests don’t come until the active ones are served to a certain point, e.g. 50% already staged.
      - Some sites prefer to keep the number of active requests at a constant level
  - Staging profile is configurable on CRIC

FTS

- One of vital services for Data Carousel success
- Scalability and stability improved greatly over the last several months. Minor or almost no issues with FTS in recent campaigns
  - Scale horizontally. FTS ATLAS instance at CERN went from 10 large flavoured VMs in January to 30 xlarge flavoured VMs (8 cores, 16 GB RAM, 80 GB HD) in April
  - Together with some tweaks on its DB, the scheduler speed is faster than ever, reaching >6K active links
  - Fixed the long standing pin leak issue (FTS 3.9.4 release or beyond)
    - To avoid cases where staged files are purged from disk buffer before transferred to final destination
- Other important features to Data Carousel
  - Throttling, monitoring & documentation
  - New feature to report a transfer as completed only when file has been successfully migrated to tape system

Sites (1/2)

- Good tape throughput is crucial to Data Carousel success
- Two lessons learned from Phase I tape site evaluation
  - Bottleneck on the tape frontend : limiting factor for fully utilizing tape capacities
  - Writing is important : good performance observed from sites who organize writing to tapes (especially grouping files on tape by datasets)
- An area with a lot of work to do:
  - Sites are all different from each other, in terms of design and implementation of tape infrastructure and experience in supporting different user communities
    - BNL has been running Data Carousel for RHIC experiments for 15+ years
    - TRIUMF originally designed their own tape system with dataset concept in mind
    - Some sites use commercial tape system, treat it as a black box
  - Tape frontend, e.g. dCache, originally designed for archival usage of tapes, not in Data Carousel mode.
  - Many sites support multiple VOs

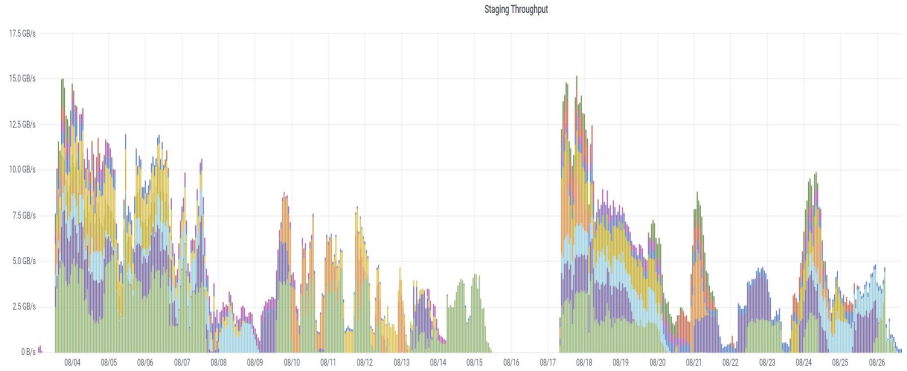
Sites (2/2)

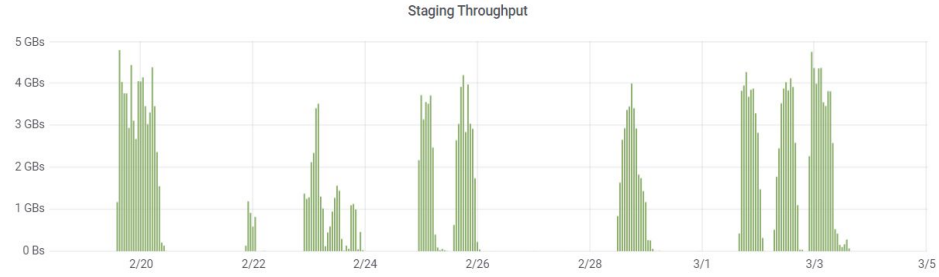
- Steady improvements have been made by almost all sites since Phase I
  - CERN has migrated their tape system from Castor to CTA.
  - Some T1s increase bulk size limit from O(1k) to O(100k)
    - Adding more staging pools in dCache
    - Increasing size of disk buffer and improve disk buffer I/O to match tape I/O
    - Switch from the default dCache HSM script interface to ENDIT interface
  - Simulation and real tape tests being conducted at various sites, w.r.t smart writing
  - Some new features from tape vendors look promising, and being evaluated.
- Experience sharing among sites and collaboration with other service providers are important

…… Next we will go through some of the recent Data Carousel exercises and their results

2020 Run2 RAW data reprocessing (1/2)

- Complete Run2 RAW (~18.5PB), staged & reprocessed
  - Phase III.1 exercise, run at scale
  - Finished on time, no complains from data preparation group
- “Real Data Carousel” mode
  - Staging was based on disk space usage, processing progress (gshare) and site staging profile
  - As a result
    - Wave-like throughput pattern observed (bulk-on-demand)
    - Much less disk space occupied than pre-staging mode



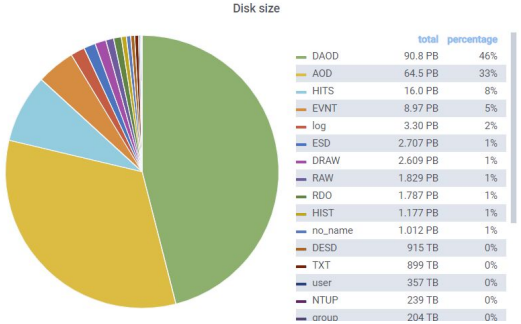


2020 Run2 RAW data reprocessing (2/2)

- Throughput
  - Sum of T1s stable staging throughput, ~15GB/s.
  - Recall efficiency : 30~40% (*roughly est.*)
- Great progress made since 2018

Sites	Stable Rucio throughput in Phase I test (2018)	Stable Rucio throughput in Run2 reprocessing campaign (2020)
CERN	2GB/s (CTA test instance)	4.3 GB/s (DDM link)
BNL	866MB/s	3.4 GB/s (DDM link)
FZK	300MB/s	1.6 GB/s (DDM link)
INFN	300MB/s	1.1 GB/s (DDM link)
PIC	380MB/s	540 MB/s (DDM link)
TRIUMF	1GB/s	1.6 GB/s (DDM link)
CCIN2P3	3GB/s	3 GB/s (DDM link)
SARA-NIKHEF	640MB/s	1.1 GB/s (DDM link)
RAL	2GB/s	2 GB/s (DDM link)
NDGF	500MB/s	600 MB/s (DDM link)

Next Steps (1/2) : more ATLAS workloads under Data Carousel mode



- Derivation production (AODs as inputs)
  - Derivation production uses AODs as inputs, which take 1/3 of our DISK space. Can we move them to tape ?
  - AOD replication policy
    - Currently AOD datasets are all pinned on disk, with 1 tape replica
    - Changes : increase tape replica to 2 for data AOD, and secondarize all AODs on disk with limited lifetime.
      - Approved by CREM2 (ATLAS Computing Resource Management).
      - To be implemented Oct~Dec 2020. 25+ PB of disk space gained, under current condition.
- Progressively add more workloads, like reconstruction and user analysis to Data Carousel mode, before or during Run3.

Next Steps (2/2) : increase tape throughput

- Looking at the landscape, several paths forward
  - Common solution from dCache, for both smart writing and efficient reading
    - 7 T1s run dCache. Will continue the discussion with dCache team
  - Common solutions from ATLAS :
    - ATLAS provides meta-info as grouping hints
    - bigger files : will make the landscape much greener, for mid-term. Not a long term solution. More evaluation/discussion with sites to come.
    - Other ideas ?
  - Sites go their own way ...
    - ATLAS provides meta-info as grouping hints

Sites	tape	SE	Bulk request	Smart writing	Pledge (2019)
CERN	CTA	EOS	100k	OK	100%
BNL	HPSS	dCache	100k	Plan in place	23%
FZK	TSM → HPSS	dCache	30k	Plan in place	13%
RAL	CTA	Echo	100k	Follow T0 ?	13%
CCIN2P3	HPSS	dCache	10k	Plan in place	12%
TRIUMF	Tapeguy	dCache	100k	OK	10%
INFN	TSM	StoRM	No limit	OK	8%
NL-T1	DMF	dCache	20k		8%
NDGF	TSM	dCache	200k		6%
PIC	Enstore	dCache	10k		4%

Summary

- Since the beginning of this HL-LHC R&D project, we have come a long way, with great progress made at all fronts.
  - Tape resources have been integrated into the ADC major components
  - Overall tape throughput improved from 10GB/s to 15GB/s (T1s)
- Still a lot of work to do
  - Overall tape recall efficiency is <50%
  - Continue to work on covering more ATLAS workloads under Data Carousel, smart writing, and monitoring …
- As we progressively roll out Data Carousel into production for Run3, we will continue to gain experience and collect statistics, to improve tape recall efficiency and grow tape capacities, to meet the needs of HL-LHC.

Backup Slides

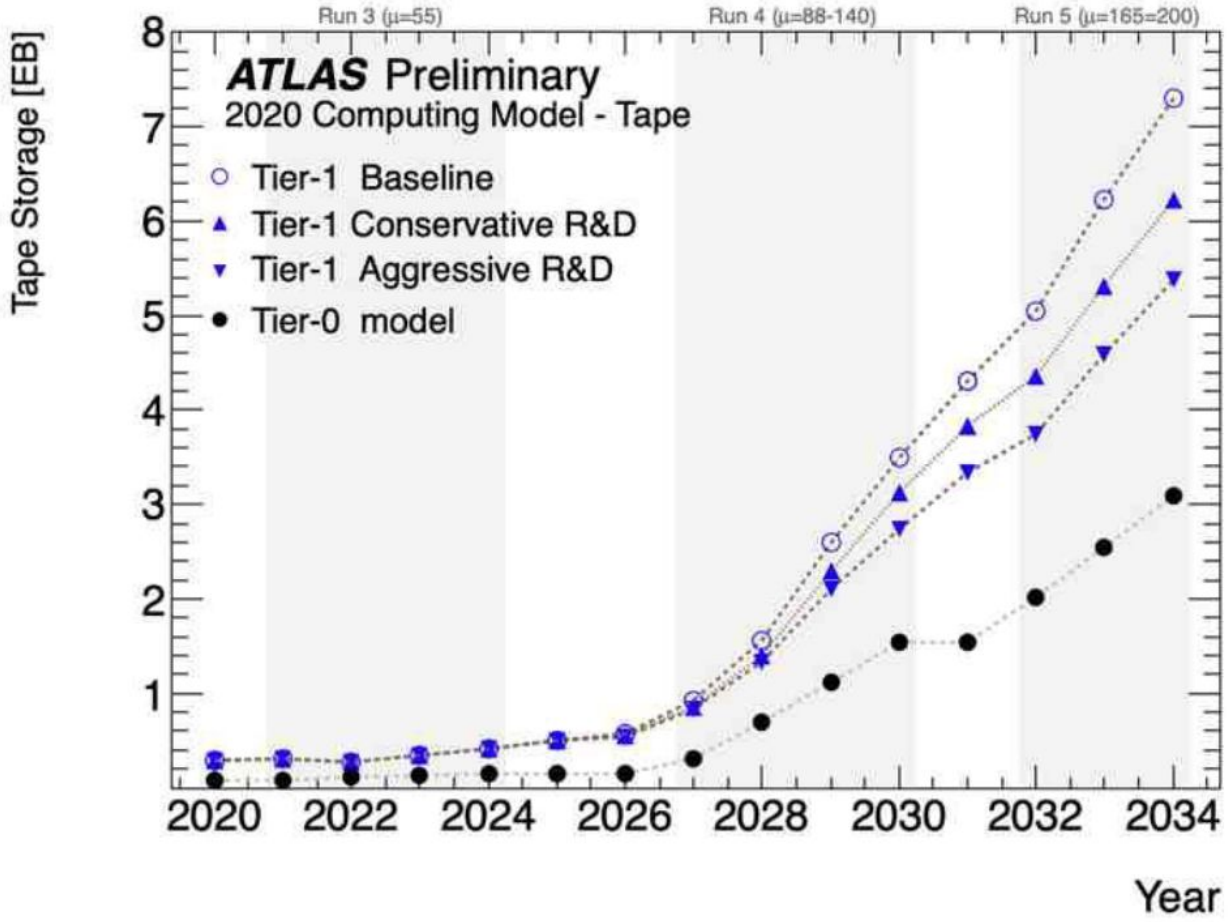

Run 3 (μ=55)
Run 4 (μ=88-140)
Run 5 (μ=165-200)
ATLAS Preliminary
2020 Computing Model - Tape
Tier-1 Baseline
Tier-1 Conservative R&D
Tier-1 Aggressive R&D
Tier-0 model
Tape Storage [EB]
Year
1
2
3
4
5
6
7
8
2020
2022
2024
2026
2028
2030
2032
2034

Table 1. Phase I tape test results (Tier1s)

Tier1 Sites	Tape Drives used	Average Tape (re)mounts	Average Tape throughput	Stable Rucio throughput	Test Average throughput
BNL	31 LTO6/7 drives	2.6 times	1~2.5GB/s	866MB/s	545MB/s (47TB/day)
FZK	8 T10KC/D drives	>20 times	~400MB/s	300MB/s	286MB/s (25TB/day)
INFN	2 T10KD drives	Majority tapes mounted once	277MB/s	300MB/s	255MB/s (22TB/day)
PIC	5-6 T10KD drives	Some outliers (>40 times)	500MB/s	380MB/s	400MB/s (35TB/day)
TRIUMF	11 LTO7 drives	Very low (near 0) remounts	1.1GB/s	1GB/s	700MB/s (60TB/day)
CCIN2P3	36 T10KD drives	~5.33 times	2.2GB/s	3GB/s	2.1GB/s (180TB/day)
SARA-NIKHEF	10 T10KD drives	2.6~4.8 times	500~700MB/s	640MB/s	630MB/s (54TB/day)
RAL	10 T10KD drives	N/A	1.6GB/s	2GB/s	1.6GB/s (138TB/day)
NDGF	10 IBM Jaguar/LTO-5/6 drives, from 4 sites	~3 times	200~800MB/s	500MB/s	300MB/s (26TB/day)

* https://cds.cern.ch/record/2709950/files/ATL-SOFT-PROC-2020-014.pdf

ATLAS Tape Writing Policy

- Write on tape everything that is not too small or too short lived soon after it is created
  - RAW datasets 2 copies (CERN + Tier-1s)
  - AOD datasets 1 copy (Tier-1s)
  - Zip and archive small size (long lived) data – 1 copy (Tier-1s)
- Always copy data to another site and write to tape (Y2018 policy change)
- No direct tape writing from production tasks
- A dedicated agent (RPG) scans disks once per day for eligible datasets not on tape
  - It makes Rucio rules for these datasets
    - Datasets distribution is based on pledge
      - Not more than N rules per Tier-1 to not overload tape buffers
        - 5-10 AOD datasets per Tier-1 at once

D.Cameron