The ATLAS Data Carousel Project

Alexei Klimentov, Mario Lassnig and Xin Zhao

WLCG Grid Deployment Board

Jun 10, 2020

BROOKHAVEN

CERN

Thanks

- Team Effort ---
  - workflow and workload management SW developers (ProdSys2/PanDA)
  - data management SW developers (Rucio)
  - distributed production and analysis team (DPAs)
  - distributed computing coordination, experts and sites operations
  - Rucio and BigPanDA monitoring teams
  - FTS, CTA and dCache SW developers and experts
  - Tier-0 and Tier-1s operations, storage and tape experts

Outline

- ATLAS Distributed Computing Software Stack
- Data Carousel objectives and motivation
  - Data carousel and HL-LHC R&D projects
- Data Carousel Phases :
  - Phase III (Y2020) highlights and results
- More challenges ahead

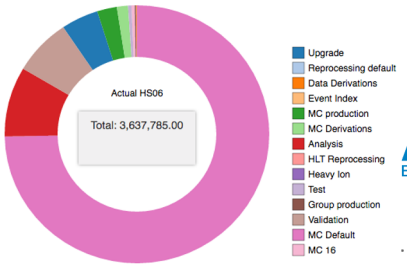
ATLAS Workflow and Workload Management. ProdSys2/PanDA

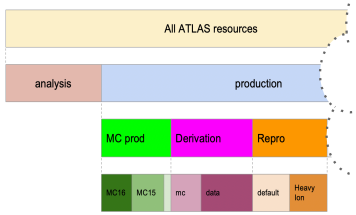
Global ATLAS operations
Up to ~1.2M concurrent jobs
25-30M jobs/month at >250 sites
~1400 ATLAS users

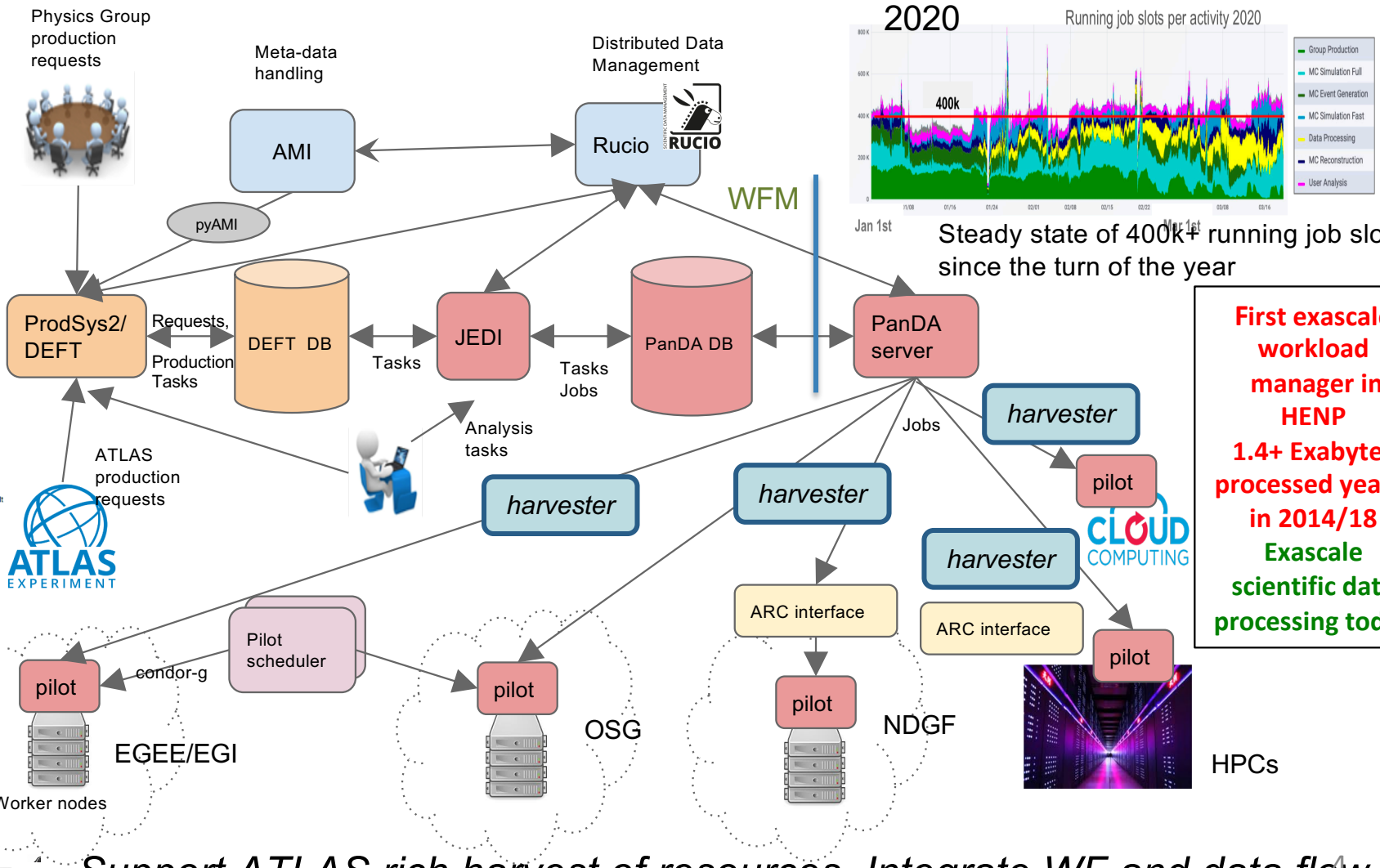
Orchestrate all ATLAS Workflows :

- *MC Production*
- *Physics Groups WF*
- *Data reprocessing*
- *T0 spill-over*
- *HLT processing*
- *SW validation*
- *User's analysis*
- *ART*

Shares/priorities







Steady state of 400k+ running job slots. since the turn of the year

First exascale workload manager in HENP
1.4+ Exabytes processed yearly in 2014/18
Exascale scientific data processing today

Support ATLAS rich harvest of resources. Integrate WF and data flow

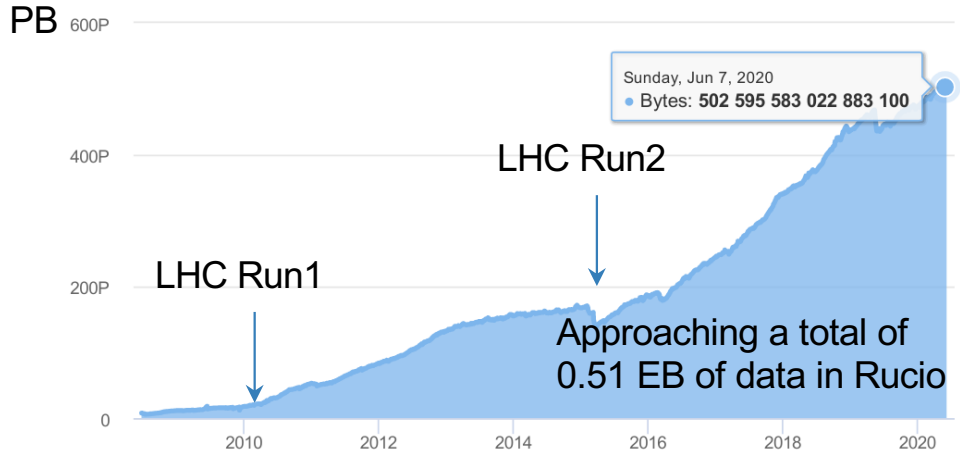
BROOKHAVEN

ATLAS Data Management. Rucio

A few numbers to set the scale

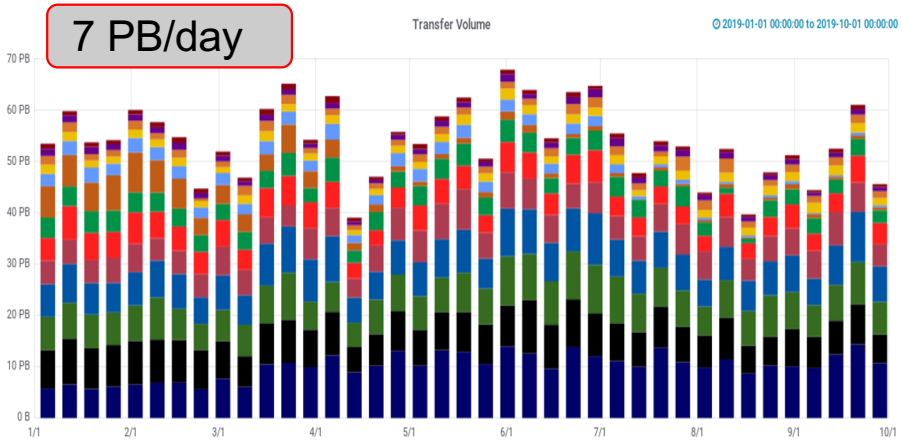
- Rucio
  - 1B+ files, 505+ PB of data, 400+ Hz interaction
  - 120 data centres, 5 HPCs, 2 clouds, 1000 users
  - 500 Petabytes/year transferred & deleted
  - 2.5 Exabytes/year downloaded & uploaded

PanDA and Rucio are highly scalable
The first exascale scientific data and workload management systems today



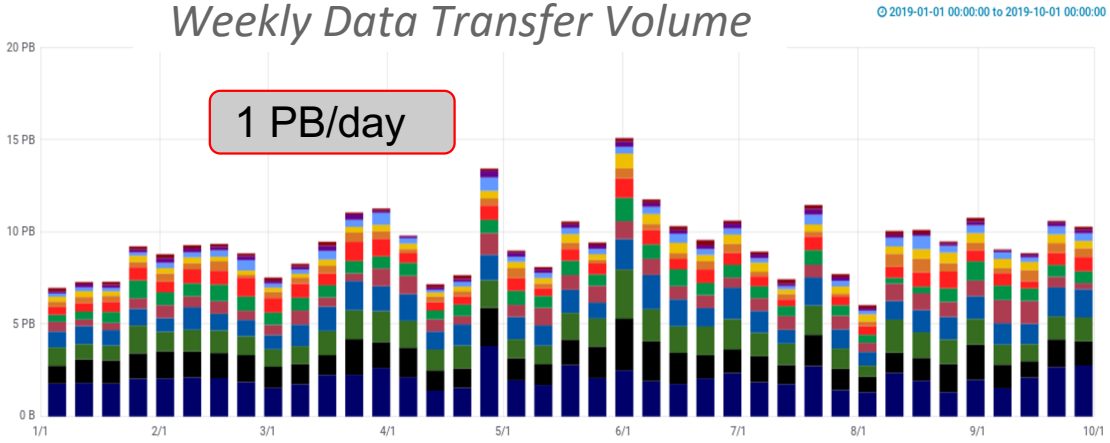
Weekly Data Access Volume

7 PB/day



Weekly Data Transfer Volume

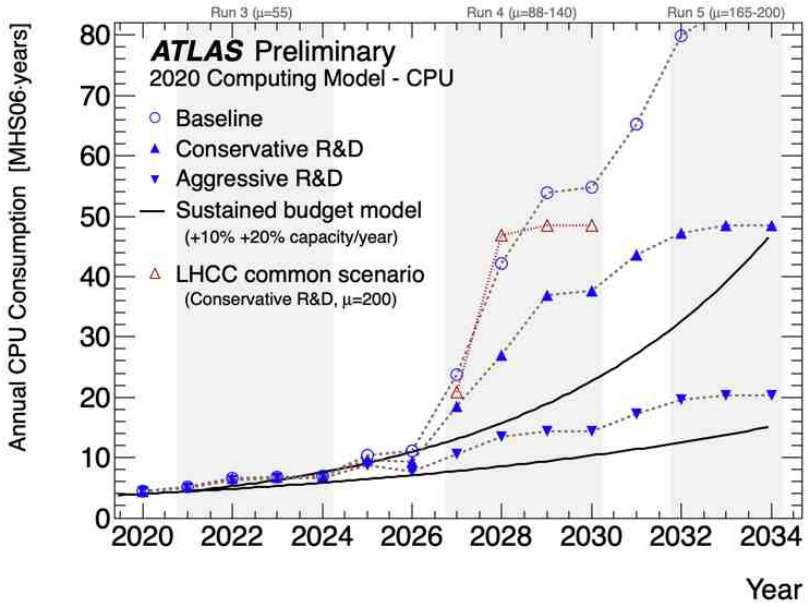
1 PB/day



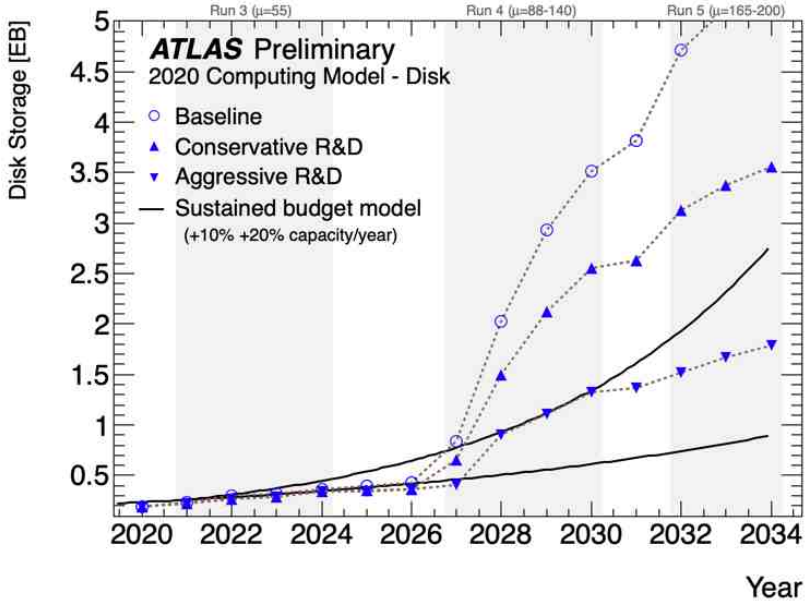
BROOKHAVEN

The High Luminosity LHC Challenge

Growth in CPU Needed



Growth in Disk Storage Needed



- High Luminosity LHC will be a multi-exabyte challenge where the envisaged Storage and Compute needs are a factor 10 to 100 above the expected technology evolution.
- LHC experiments have successfully integrated HPC facilities into its distributed computing system. “Opportunistic storage” basically does not exist for LHC experiments.

The HEP community needs to evolve current computing and data organization models in order to introduce changes in the way it uses and manages the infrastructure, focused on optimizations to bring performance and efficiency not forgetting simplification of operations.

BROOKHAVEN

Data Carousel – HL-LHC R&D Computing Projects

WLCG and experiments have launched R&D projects to address HL-LHC challenges

- ***Data Lake***. The aim is to consolidate geographically distributed data storage systems connected by fast network with low latency. The Data Lake model as an evolution of the current infrastructure bringing reduction of the storage and operational costs
- ***Intelligent Data Delivery Service*** (iDDS). The intelligent data delivery system will deliver events as opposed to delivering bytes. This allows an edge service to prepare data for production consumption, the on-disk data format to evolve independently of applications, and decrease the latency between the application and the storage. The first implementation in April-May 2020 for Data carousel and active ML workflows
- ***Hot/Cold storage.*** Data placement and data migration between "Hot-Cold" storage using data popularity information
- ***Data format and I/O.*** Evaluating new formats (f.e. parquet) and I/O performance for HENP data
- ***Third Party Copy***. Improve bulk data transfer between sites and find a viable replacement to the GridFTP protocol
- ***Operations Intelligence.*** Reduce HEP experiments computing operations effort by exploiting anomaly detection, time series and classification techniques to help the operators in their daily routines, and to improve the overall system efficiency and resource utilization
- ***Data Carousel.*** Use tape more effective and active in distributed computing context.

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 scans disks 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

Rucio dataset in ATLAS is a unit of data replication and processing (O(10-10K) files; file size O(1-10GB))

Data Carousel R&D Project

By 'data carousel', we mean an orchestration between workflow/workload management (WFMS), data management (DDM), 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 ~ X% of inputs are pinned on disk at any one time.

The project to use tape in effective way was initiated for RHIC experiments (in production for STAR and PHENIX for more than 15 years. They managed to fetch files at BNL pretty much at tape speed for weeks in a row).

Data Carousel is one of R&D projects to address High Luminosity LHC distributed data processing challenge (scope, context and scale are different from RHIC), we are working very closely with CERN, 9 Tier-1s, FTS and dCache teams

Ultimate goal : use tape more efficient and active

Cycle through tape data, processing all queued jobs requiring currently staged data

We focus on ***efficiently*** using the ***available*** tape capacities

- Introduce no or little performance penalty to tape throughput, after integrating tapes into our workflow
- Improve efficiency and throughput of tape systems, by orchestrating the various components in the whole system stack, starting from better organization of writing to tapes
- Solutions should scale proportionally with future growth of capacities of tape resources

'Data Carousel' LHC R&D was started in the second half of 2018 → to study the feasibility to use tape as the input to various I/O intensive workflows, such as derivation production and RAW data re-processing

…and "tape" could be any "cold" storage (it is led by A.Klimentov, M.Lassnig and X.Zhao)

Data Carousel R&D Project. Cont'd

- DDM system : Rucio → more intelligent tape I/O
  - Bulk data staging requests handling
  - Use FTS features on more intelligent way
- File Transfer Service → optimize scheduling of transfers between tape and other storage endpoints
- DDM / WFM / Facilities integration. Optimize data placement to tape
  - Do data grouping for files known to be re-read from tape
  - Optimize file size (Larger file size, 10GB+ *preferred*, see recent CTA studies)
  - Use novel (or request new) features of storage systems (dCache, EOS, CTA,…)

Data Carousel Project Phases

- Phase I : Tape Sites Evaluation (Y2018)
  - completed
    - Conduct tape staging tests, understand tape system performance at sites and define primary metrics
- Phase II : ProdSys2/Rucio/Facilities integration (Y2019-2020)
  - Address issues found in Phase I
  - completed
    - Deeper integration between workflow, workload and data management systems (ProdSys2/PanDA/Rucio), plus facilities
  - Identify missing software components
- Phase III : Run production, at scale, for selected workflows (Y2020)
  - in progress
    - Address it in cold/hot storage context

Now we are in the middle of Phase III ('run production at scale' was demonstrated and now we increase number of workflows running in Data Carousel mode)

Goal : to have data carousel in full production for LHC Run3

Data Carousel Phase I. Jun-Nov 2018*

- Established baseline measurement of current tape capacities
- 9 ATLAS T1s and CERN participated
- Overall throughput from all T1s (as of Nov, 2018) reached ~600TB/day, and Y2020 throughput is over 1.1 PB/day
- CERN conducted its own Tape Archive (CTA) test, reached ~2GB/s throughput

- Average Tape Throughput Y2018 (site w/o ATLAS) : throughput directly from local site tape monitoring
- Stable Rucio Throughput Y2018 (ATLAS) : from Rucio dashboard, over a "stable" run time
- Stable Rucio Throughput Y2020 (ATLAS) : Take a period of time, during which a site delivered good throughput over a sustained period of time (>5 hours)

Site	Tape Drives used	Average Tape (re)mounts #	Average Tape throughput GB/s	Stable Rucio throughput Y2018 GB/s	Stable Rucio throughput Y2020 GB/s
BNL	31 LTO6/7	2.6	1~2.5	0.87	3.4
FZK	8 T10KC/D	>20	~0.40	0.30	1.6
INFN	2 T10KD	Majority tapes mounted once	0.28	0.30	1.1
PIC	5~6 T10KD	Some outliers (>40 times)	0.50	0.38	0.54
TRIUMF	11 LTO7	Very low (near 0) remounts	1.1	1.0	1.6
CCIN2P3	36 T10KD	~5.33	2.2	3.0	3.0
SARA-NIKHEF	10 T10KD	2.6~4.8	0.50~0.70	0.64	1.1
RAL	10 T10KD	n/a	1.6	2.0	2.0
NDGF	10 IBM Jaguar/LTO-5/6 (@4 sites)	~3	0.20~0.80	0.50	0.60

* Y2020 numbers

CERN CTA Y2020 : 20 enterprise IBM, 7 LTO tape drivers 4.6 GB/s (4.3 GB/s)

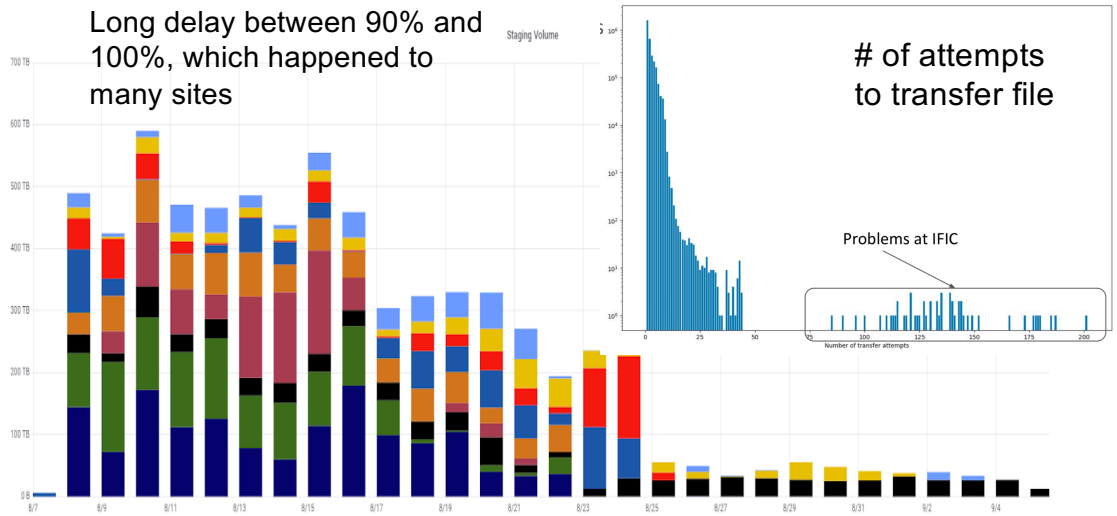
Data Carousel Phase I. Lessons Learned

- Tape frontend --- a potential bottleneck for an effective tape usage
  - Limiting number of
    - incoming staging requests
    - staging requests to pass to backend tape
    - files to retrieve from tape disk buffer
    - files to transfer to the final destination
- Data organization (file placement on tape) is vital
  - Good throughput seen from sites who organize writing to tape (especially in case grouping data by datasets)
    - Usually the reason for performance difference between two sites that have similar hardware and software setup

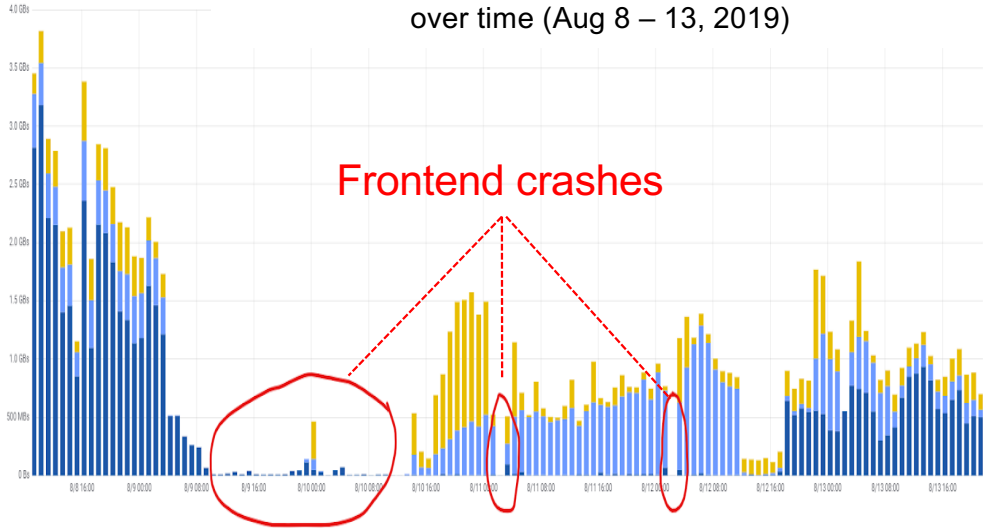
Data Carousel Phase II. Aug-Oct 2019

- Deeper integration of workflow/workload management (ProdSys2/JEDI/PanDA), data management (Rucio) systems and facilities
  - ProdSys2/Rucio communication protocol
  - New algorithms for data staging to respect global shares and priorities, resources, and sites tape performance (staging profile)
- Two rounds of data carousel exercises have been conducted :
  - the second round was combined with data reprocessing campaign
  - It took 5 days to have 70-90% data staged
- FTS and dCache limitations

Long delay between 90% and 100%, which happened to many sites



Staging throughput (GB/s) from three Tier-1s (colored) over time (Aug 8 – 13, 2019)



BROOKHAVEN

Software Development to address Phase III Data Carousel challenge

Rucio

- Notification extension
  - Fine grained progress notifications for replication rules
  - Selective AMQ notifications for Prodsys for each 10% of transfer progress
- Throttler improvements
  - Improved throttler for throttling of STAGING requests
  - Introduced source-based throttling of links
    - Unfortunately this feature did not scale well and had to be disabled
    - Currently running only with destination-based throttling
- Metadata prototype implementation to group data on tape

Monitoring : Rucio, ProdSys2

ProdSys2/iDDS

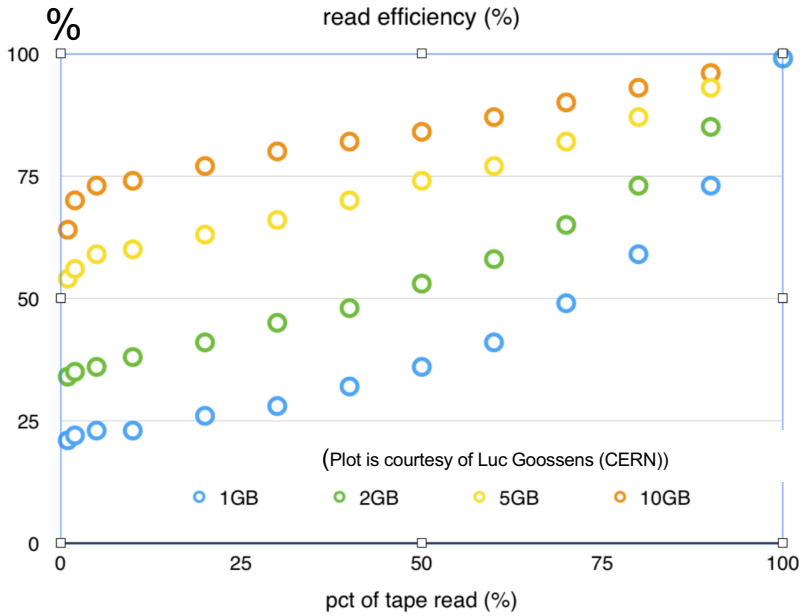
FTS

CTA

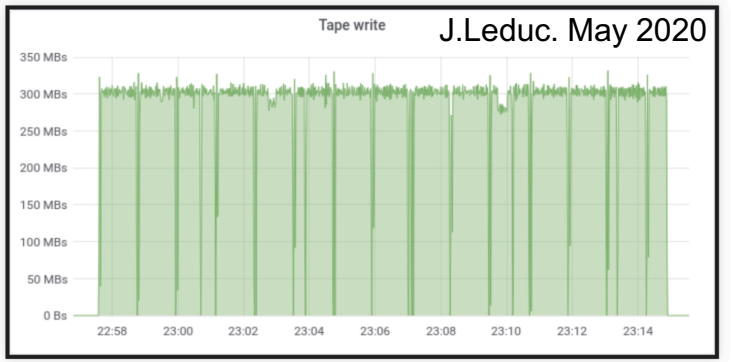
BROOKHAVEN

Data Carousel. Smart writing

- Efficient data carousel is not possible without smart writing
- It is a team effort between storage SW developers, sites and experiments (TRIUMF and CTA have a very interesting experience)
- It is still under discussion how ‘ATLAS’ can pass meta information to ‘sites’
- Possible options
  - Tape families --- too high of a layer than datasets, won’t help much
  - Bigger files
    - Zip small output files before writing to tape.
    - Target 10GB
    - CTA team studies are very interesting and we need deeper studying
  - Co-locating files from the same dataset on tape
    - Since they will be recalled together, equivalent to “bigger fat file”
    - We have a site that put all files of a dataset on one tape (or 1+ for bigger dataset). Reach almost stream reading speed of a tape drive per tape mount



Writing small files to CTA

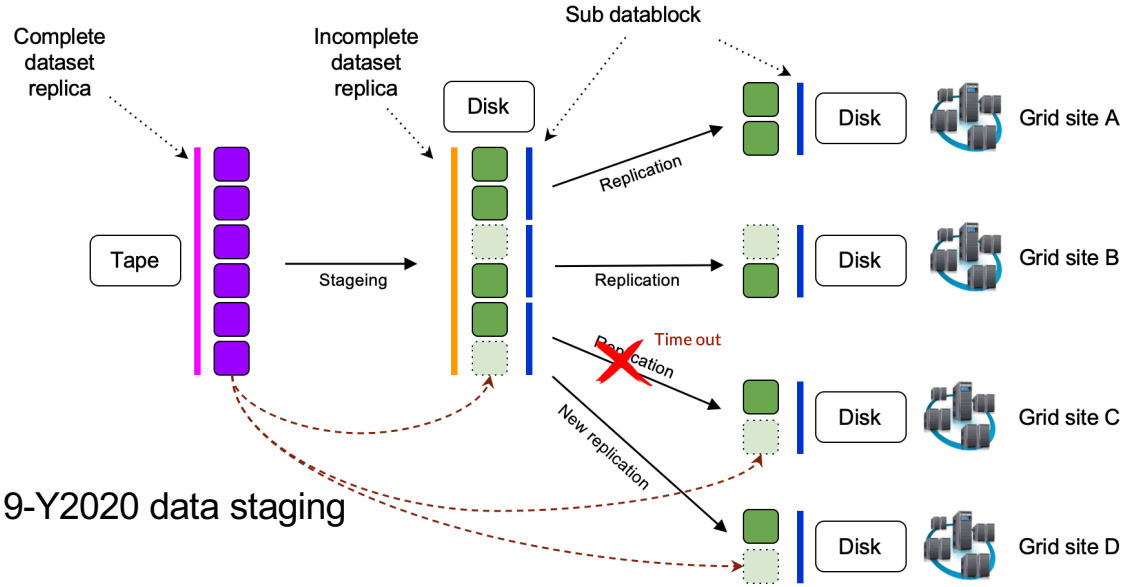


Archiving 3000 x 100MB files to 1 LTO tape drive at nominal drive speed
Achieving 95% efficiency

BROOKHAVEN

Software Development to address Phase III Data Carousel challenge. iDDS

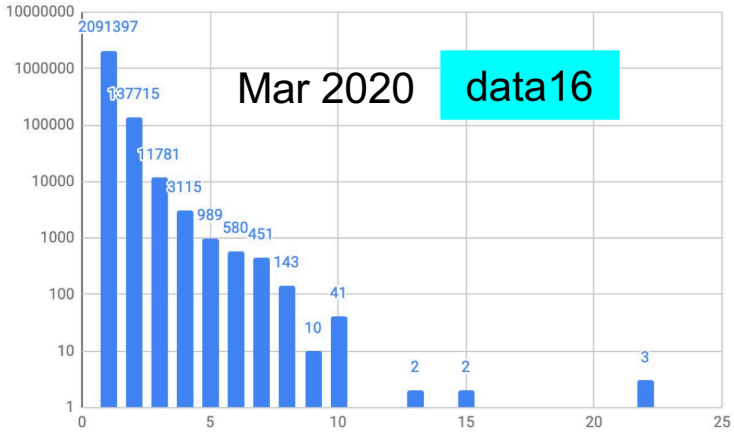
- A dataset is a unit of ATLAS data processing and replication
- Data carousel works with datasets and ProdSys2 sends staging request per dataset although files are used in downstream systems (ATLAS Dataset O(10-10k files) [file O(1-10GB)]
  - ✓ Files in each dataset are prestaged by the tape system rather randomly
- Potential issues : with jobs submission (delay with files pre-staging by FTS), PanDA queues saturation (many jobs are assigned for execution, but not started), longer occupancy for temporary data on disk



Y2019-Y2020 data staging

Courtesy of T.Maeno

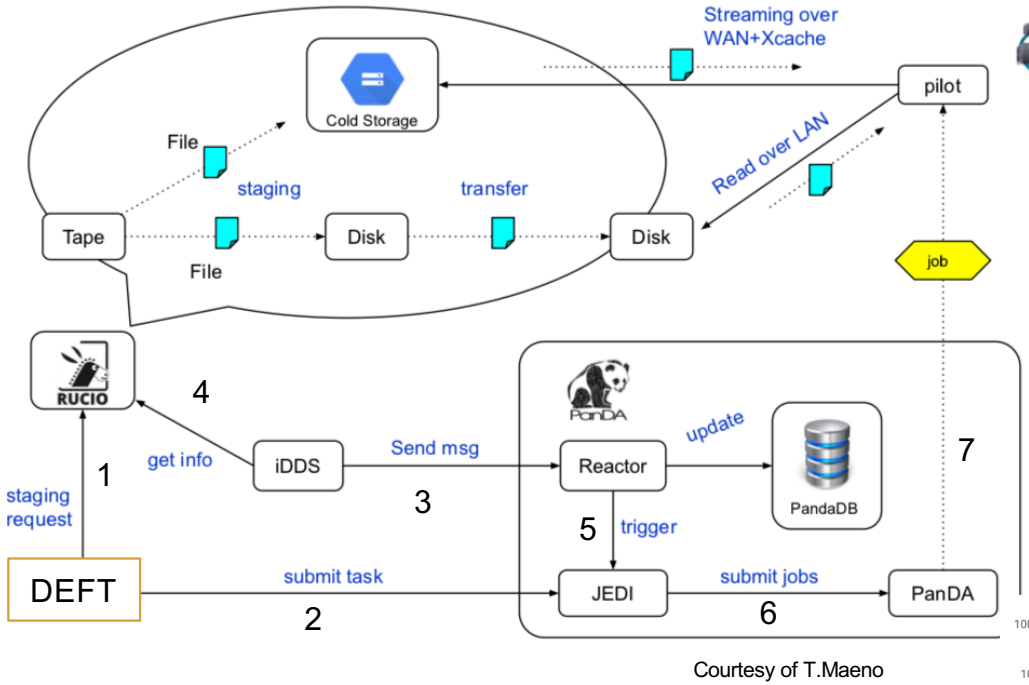
Number of attempts for each job (Data Carousel w/o iDDS)



Intelligent Data Delivery Service R&D *(iDDS). The intelligent data delivery system will deliver events as opposed to delivering bytes. This allows an edge service to prepare data for production consumption, the on-disk data format to evolve independently of applications, and decrease the latency between the application and the storage.*

BROOKHAVEN

Software Development to address Phase III Data Carousel challenge. iDDS. Cont'd



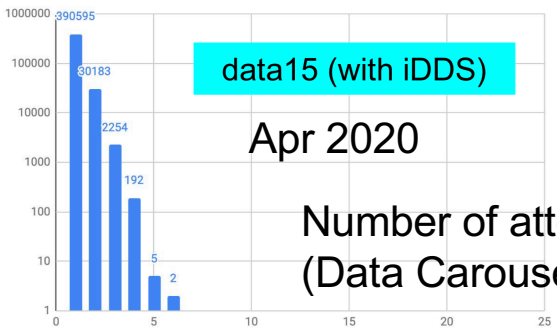
Courtesy of T.Maeno

Y2020 data staging with iDDS

DEFT – Production System part responsible for workflow management
JEDI -- Production System part responsible for Workload Management (tasks and jobs brokering)

Data carousel + iDDS algorithm

DEFT defines task : Task is in staging state (waiting)
(1) DEFT sends request to DDM (Rucio) to start staging
(2) DEFT notifies JEDI
JEDI releases task
(3) JEDI find Rucio pre-staging rule for the task and sends request to iDDS
(4) iDDS communicates with Rucio, finds staged files and reports (5) it to JEDI
(6) JEDI generates jobs to process staged files
(7) Jobs brokerage and files transfer are done as usual

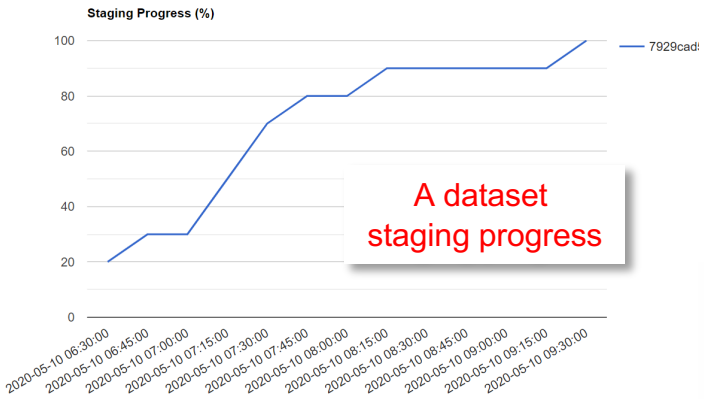


Number of attempts for each job (Data Carousel + iDDS)

BROOKHAVEN

Software Development to address Phase III Data Carousel challenge. Monitoring

A new development, provides number of different perspectives and



A dataset staging progress

Step Name	DRAW_RPVLL r11782 99.99%	DRAW_RPVLL p4071 100.00%	DAOD_RPVLL p4072 100.00%	DAOD_RPVLL r11784 100.00%
Tasks total (running)	596 (596)	596 (596)	596 (596)	596 (596)
Input events (running tasks)	17,155,408,824	1,402,798,424	1,402,793,218	1,402,784,870
Processed events	17,155,371,807	1,402,806,401	1,402,808,942	1,402,801,705
Output events	1,402,836,476	1,402,956,006	1,402,836,253	1,402,826,206
Running/Pending/Not started	0%/0%/0%	0%/0%/0%	0%/0%/0%	0%/0%/0%
Input bytes (running tasks)	18.27 PB (596 tasks)	1.61 PB (596 tasks)	1.38 PB (596 tasks)	1.61 PB (596 tasks)
Input bytes (done tasks)	18.27 PB (596 tasks)	1.61 PB (596 tasks)	1.38 PB (596 tasks)	1.61 PB (596 tasks)
Output bytes	1.61 PB (596 tasks)	1.61 PB (596 tasks)	1.38 PB (595 tasks)	1.38 PB (596 tasks)
Average HS06 per event	396	5	18	1318
Duration (finished tasks)	3.80 days	3.20 days	1.28 days	4.19 days

SARA-MATRIX_DATATAPE		-->	0	0	57 (+0)	0	0	31302
FZK-LCG2_DATATAPE		-->	0	0	76 (+0)	0	0	46958
BNL-OSG2_DATATAPE		-->	0	0	132 (+0)	0	0	66675
TRIUMF-LCG2_DATATAPE		-->	0	0	53 (+0)	0	0	28916
INFN-T1_DATATAPE		-->	0	0	48 (+0)	0	0	25158
IN2P3-CC_DATATAPE		-->	0	0	79 (+0)	0	0	49220
RAL-LCG2_DATATAPE		-->	0	0	84 (+0)	0	0	48474
PIC_DATATAPE		-->	0	0	30 (+0)	0	0	16128

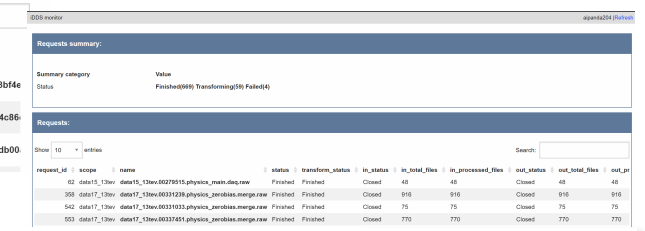
Data Carousel activity overview
ATLAS global and Tier-1s

Dataset staging information	
Staged / Total files	406 / 406
dataset	data15_13TeV:data15_13TeV.00276329.physics_ZeroBias.merge.RAW
status	done
start_time	2020-05-09T10:49:03.398755
end_time	2020-05-10T12:25:13.455162
rse	34f4f94ccc0f4b579cd2698ca3bf4eff
update_time	2020-05-10T12:25:13.455162
source_rse	NDGF-T1_DATATAPE
campaign	data15_13TeV
pr_id	30747

Staging info in tasks monitoring

iDDS information	
request_id	192
scope	data15_13TeV
name	data15_13TeV.00276329.physics_ZeroBias.merge.RAW
request_type	StageIn
transform_tag	2
workload_id	21243561
status	Finished
request_created_at	2020-05-09T12:07:05
request_updated_at	2020-05-10T12:25:06

iDDS info for task



iDDS dashboard

… and way, way more

S.Padolski
M.Borodin

BROOKHAVEN

Software Development to address Phase III Data Carousel challenge. FTS and Networking

- Networking is and has been one of the rock-solid highly reliable building block of ATLAS computing successes
- FTS is one of vital services for Data Carousel success
  - FTS support including monitoring is super important for Data Carousel success (also FTS supports majority of HENP experiments and integrated with Rucio)
  - New feature is being implemented to report a transfer as completed only when file has been migrated to tape successfully
- Minor to almost no FTS issues during Data Carousel Phase III
  - DB response slowed down from 1-4 mins to 20 mins
    - CERN database tuning on Jan 28th and FTS scheduler performance was improved
    - Number of ATLAS VMs was increased to 30 (xlarge flavoured VMs) in April (to be compared with 10 large flavoured VMs in Jan)
- FTS team fully supported Data Carousel

Data Carousel Phase III.

Run Production at scale. Feb-Apr 2020. Phase III - 1

- Reprocess a complete LHC Run2 ATLAS RAW data sample (~18.5 PB in total)
  - Perform in data carousel mode to avoid data staging in advance
  - Respect reprocessing share and priority vs other workflows
    - Group production
    - Monte-Carlo simulation
    - Users analysis

completed

- Demonstrate 18 PB RAW data reprocessing with 1 PB disk buffer

Run Derivation Production in Data Carousel mode. Jun-Oct 2020. Phase III - 2

- PoC phase. June 10th, 2020. Produce DAOD from AOD (0.3 PB data sample)
- Run derivation production for Y2016-2018 AOD and produce DAOD_PHYS and DAOD_PHYSLITE

DAOD – Derived Analysis Object Data
Primary data format for physics analysis
DAOD_PHYS(LITE) – DAOD for Run3

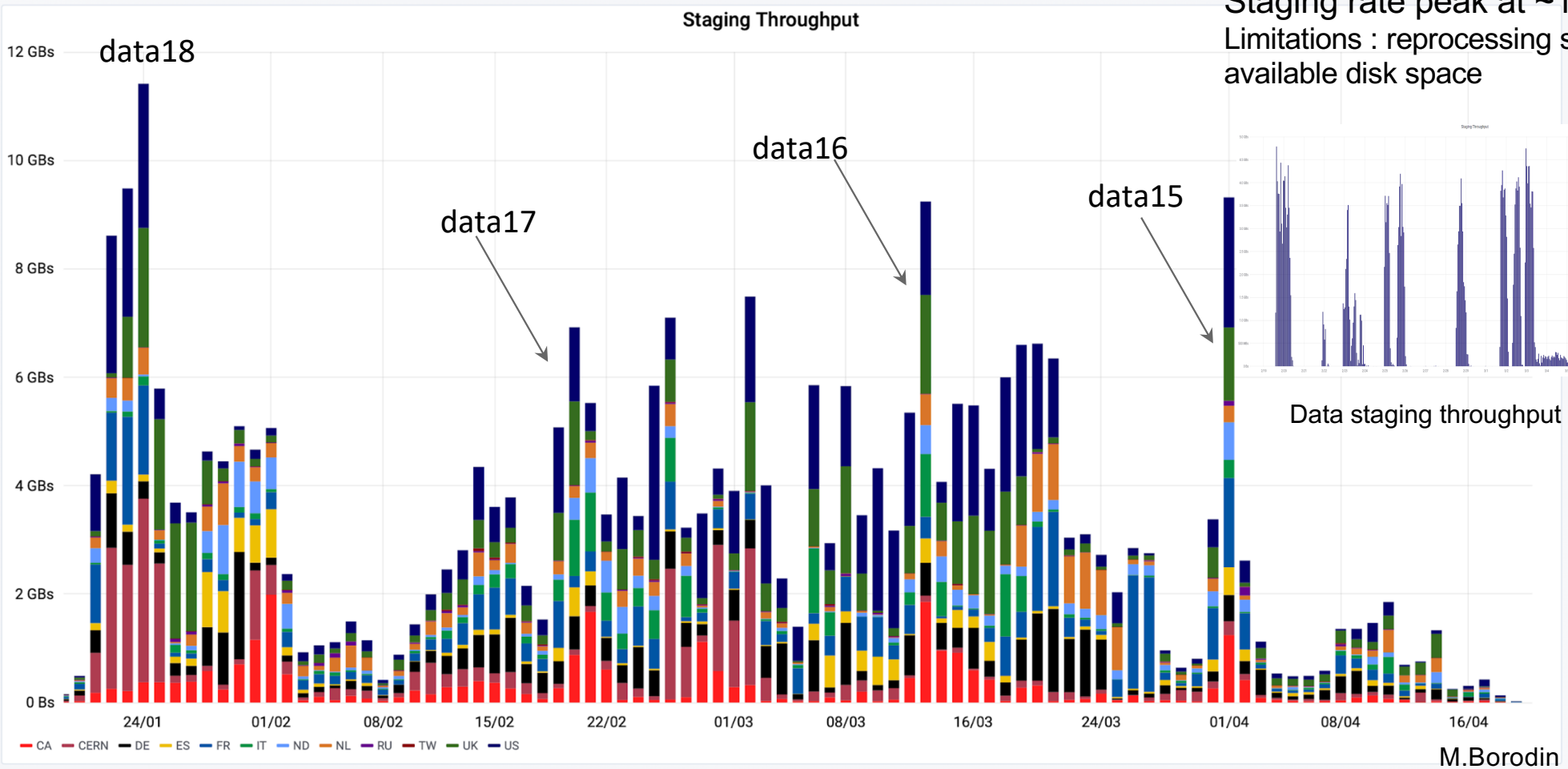
BROOKHAVEN

Data Carousel Phase III-1 Highlights

- Workflow and Data sample. Reprocess a complete Run 2 (2015-2018) data sample : 596 runs (datasets). It was started on 19 of January and finished on 4 of April. RAW data : ~18.5 PB
  - Reprocessing was done in steps : Y2018 data, Y2017,… and for different scenarios
    - 9 Tier1s and CERN
    - 9 Tier1s only
    - Data Carousel with iDDS component
- Staging scenario. Data were staged to Tier-1s and Tier-2s (aka *nucleus* of the Tape site). Tier-1s were asked in advance for a preferable staging profile (lessons learned from Phase II, when we did staging in bulk mode)
  - Staging profile :
    - Upper/Lower limits of number of concurrent requests + “time delay between bunches of staging requests”
    - New bunch won’t start until the previous bunch falls below a threshold (e.g. 50% done)
    - Limits and time delay were defined by sites:
      - Make requests more bulky
      - Control bulk size, reduce load on FTS and site frontend
  - Input data have been deleted as soon as reconstruction step is done

Data Carousel Phase III-1. Staging throughput

Staging rate peak at ~11+ GB/s
Limitations : reprocessing share, available disk space



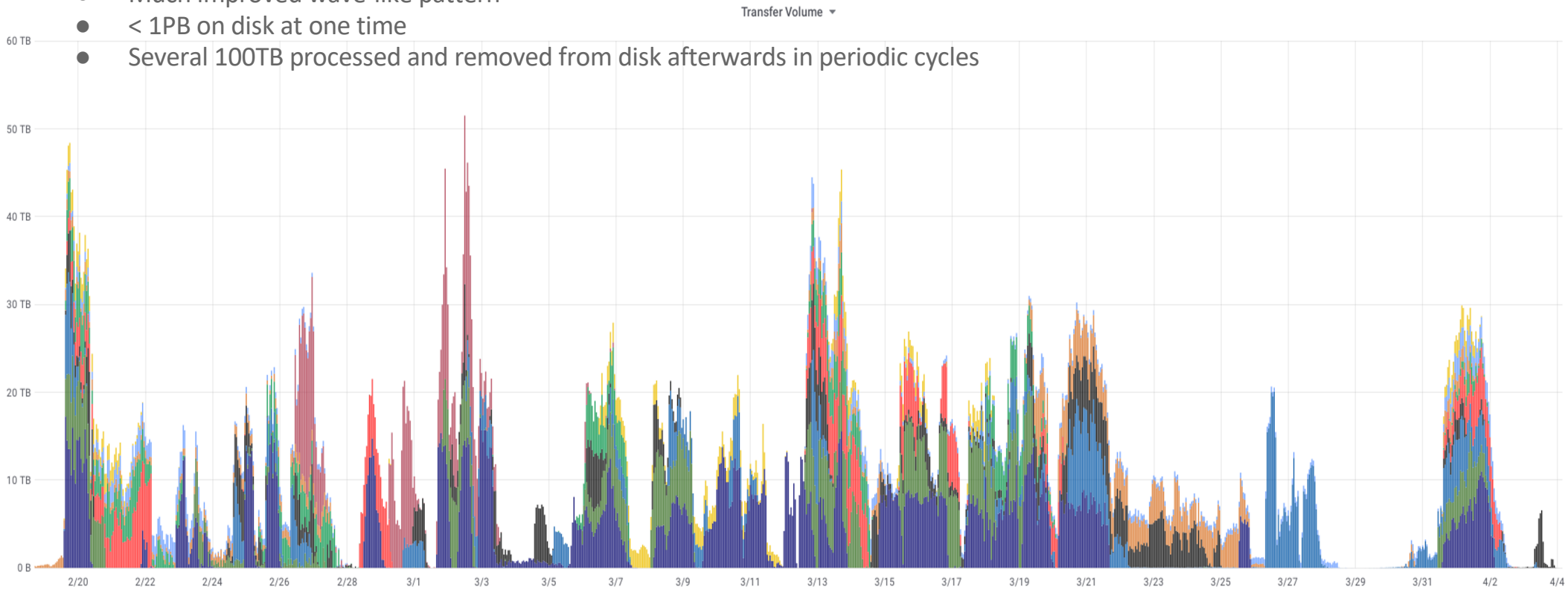
Data staging throughput of one site

M.Borodin

BROOKHAVEN

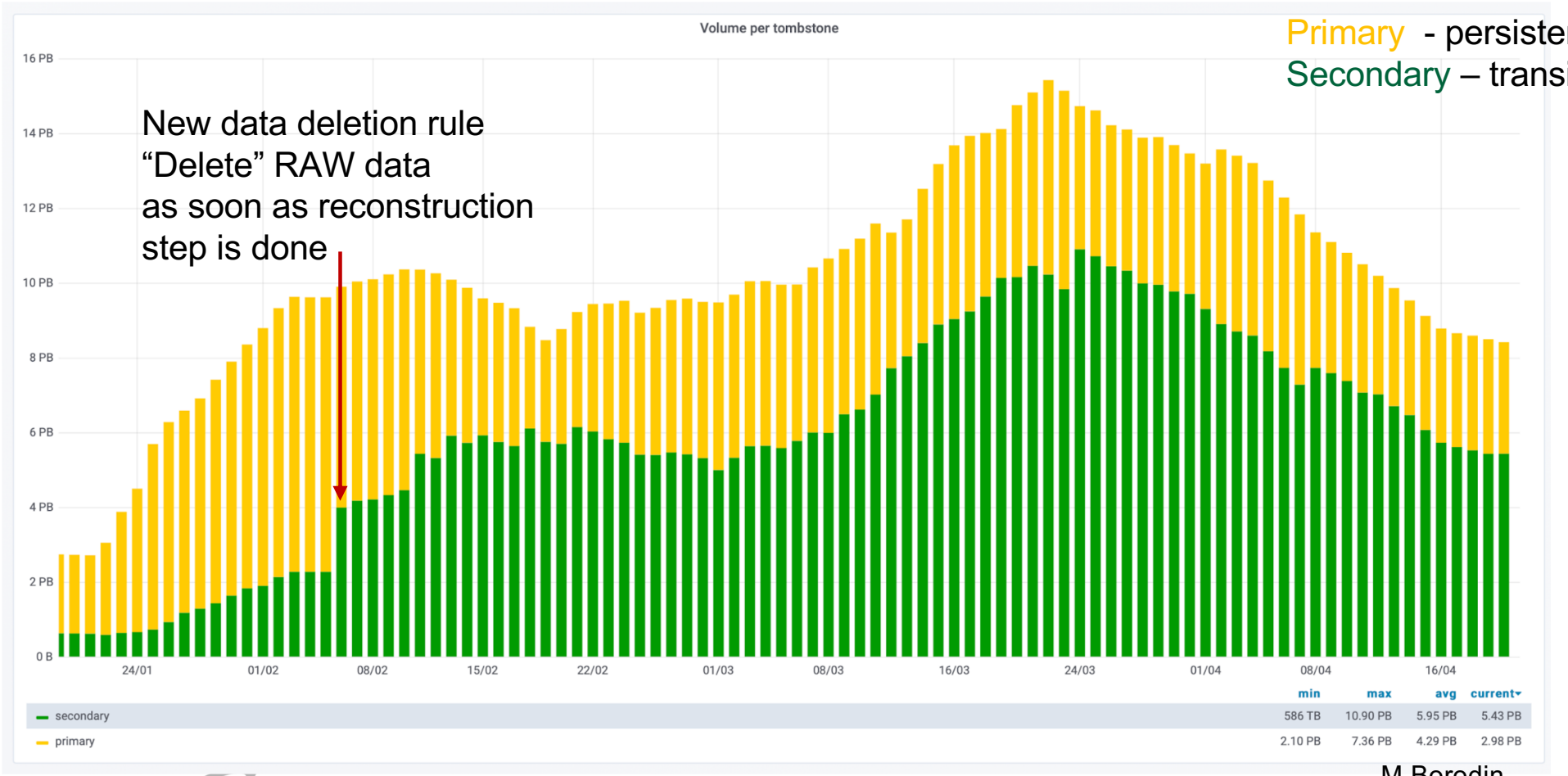
Data Carousel Phase III-1. Staging throughput

- Run 2 Y2017 RAW reprocessing - 5.5 million files to be staged , total volume : 5.7PB
- Much improved wave-like pattern
- < 1PB on disk at one time
- Several 100TB processed and removed from disk afterwards in periodic cycles



BROOKHAVEN

RAW Data on ATLAS disks (Jan – Apr 2020)



New data deletion rule
"Delete" RAW data
as soon as reconstruction
step is done

Primary - persistent data
Secondary – transient data

BROOKHAVEN

	data18			data17			data16			data15		
Site	**days**	N	PB	**days**	N	PB	**days**	N	PB	**days**	N	PB
BNL	**2.1**	29	0.7	**1**	58	1.8	**0.7**	39	1.2	**1**	17	0.3
CERN	**4.5**	26	1.0	**2.2**	22	0.7						
FZK	**3.1**	14	0.5	**3.6**	19	0.5	**2**	22	0.7	**3**	8	0.07
IN2P3	**5.7**	21	0.6	**1.5**	19	0.5	**5.5***	18	0.6	**2**	9	0.1
INFN				**5**	15	0.5	**6.6**	7	0.2	**4.5**	2	0.03
NDGF	**14**	12	0.4	**8.7**	7	0.2	**22***	11	0.3	**4**	5	0.09
PIC	**10**	10	0.4	**2.2**	10	0.3	**2.5**	4	0.1	**1**	4	0.04
RAL	**4.3**	25	0.8	**1.4**	21	0.8	**1.8**	18	0.8	**1**	9	0.1
SARA	**16***	21	0.6				**2.8**	14	0.6	**1**	5	0.04
TRIUMF	**14**	10	0.4	**3.2**	15	0.4	**2.7**	12	0.5	**1**	5	0.1

M.Borodin

Data Carousel Phase III-1. Data Staged per Site. **) – minor issues with sites*

N – total datasets, PB – total size in PB, days – average number of days to stage a dataset

BROOKHAVEN

Data Carousel plans for 2020

- Demonstrate Data Carousel for Derivation Production : PoC (started today) and at scale
- Software development
  - Fine tuning in Production System and iDDS (for instance, staging requests distribution between Tier-1s in case of multiple tape replicas)
  - Rucio : Source-based throttling
    - Requires a substantial re-write of the code, otherwise we would hit the scalability issue again
    - Still on the roadmap, but currently lack of expert person to do this development
  - Tape metadata to group data on tape (together with dCache, Tier-1s and FTS teams)
    - Communicate colocation metadata to FTS & Tape system
    - dCache team proposal for tape recall efficiency
- Continue tape throughput and “Big Files” studies together with Tier-1s and CERN
- Operations intelligence and more automation (automatic tasks rebrokering in case of tails)
- DOMA ACCESS discussion about a joint LHC experiments test(s)
  - Our vision, that it isn’t only about tape, but also about networking (FTS and more)

More Challenges Ahead

- We successfully and quickly passed “a pilot project phase” between ATLAS, FTS, dCache, CTA and T0, T1 centers
  - …and obtain metrics vital for the project
  - Many unknown unknowns problem retired/solved (FTS database limitation is only an example). Known unknowns (smart writing, meta-information passing…) still remain
- ATLAS demonstrated a “real Data Carousel” mode in action, in a production environment with many other concurrent activities (data writing, data rebalancing, data consolidation, etc)
- New software module (iDDS) is being evaluated and integrated with the ATLAS Production System, which can potentially mitigate the latency issue of staging inputs from tape directly. A good example how HL-LHC R&Ds work together
- New algorithms to be developed for an intelligent decision making

More Challenges Ahead. Cont'd

- We respected reprocessing share and we are limited by the available disk space, that's why we didn't stress tape services hard enough. In the future, we need to put more pressure on tape sites, to find new bottlenecks, with the goal to see the same pulse-shape performance, at a much larger scale.
- We (as WLCG community) need to address Data Carousel topic in a global (multi-VO) context
  - Tape throughput studies
  - FTS data movement (Rucio+FTS for ATLAS and CMS)
  - Data grouping and smart writing (together with CTA, dCache and Tier-1s)
    - dCache team presented ideas how to improve tape recall efficiency
    - CTA team works on improving LTO read efficiency

Thanks

- It is a collaborative effort in ATLAS (Operations, Distributed Computing, Software developers), sites (T0 and T1s), dCache, FTS and CTA teams. Thanks to all
- Thanks to M.Borodin, D.Cameron, A.DiGirolamo, J.Elmsheuser, E.Karavakis, J.Leduc, T.Maeno and S.Padolski for slides and materials

Back up slides