Preparing for Run-2

Andrej Filipčič
Jožef Stefan Institute, Ljubljana

- CPU and Storage organization
- Specializing sites for different workflows
- Flat hierarchy

Run-1 model, briefly

- Strict hierarchical model (Monarc):
  - Clouds: T1 + T2s (+ T3s)
  - No direct transfers between foreign T2s
  - Relaxed towards the end of Run-1 (Multi-cloud production – T2s can process jobs of many clouds)
- Production organization:
  - Tasks assigned to T1s
  - T1 is the aggregation point for the output datasets of the tasks
  - T2 PRODDISK used for input/output transfers from/to T1
- T2 disk space:
  - distribute the final data to be used by analysis
  - store secondary replicas of precious datasets

Planning for Run-2 model - facts

- Network globally improved
  - Much higher bandwidth (an order of magnitude increase)
  - Most of the links between ATLAS sites provide sufficient throughput : full mesh for transfers can be used
- Many Tier-2 sites provide the Tier-1 level stability of computing, storage and WAN
  - Many in LHCONE or other high-throughput networks
  - Tape resource is the only difference between Tier-1s and large Tier-2s, as far as the usability for ATLAS is concerned
- CPU only (opportunistic) centers are fully integrated in ATLAS
  - Some run all kind of tasks, including data reprocessing
  - Have good connectivity to geographically close Storage Elements

CPU and Storage organization

- Breaking the barrier between the Storage Element and Computing Element:
  - Remote I/O, job overflow, remote fail-over of input or output file staging → storage not strictly bound to the site computing resource
  - Tier-1, Tier-2, Tier-3 storage classification does not make much sense anymore
- ATLAS Storage pool:
  - TAPE
  - STABLE disk storage – T1 + reliable T2 (former T2Ds)
  - UNSTABLE disk storage – less reliable T2s
  - VOLATILE disk storage – unreliable T2s, T3s, opportunistic storage

Using new storage classes

- TAPE:
  - will be addressed by Richard
- STABLE:
  - Common pool of Tier-1 and Tier-2 storage, NO differentiation of sites
  - Tier-1 and Tier-2 sites will be used at the same level for storing custodial, primary data
  - Production tasks will be assigned/brokered to Tier-2 sites as well
  - Tier-2 sites will be used to store the final outputs of the production chain

Using new storage classes 2

- UNSTABLE:
  - Will be used for secondary data (for analysis)
  - Will not store primary data
  - We will not rely on the sites as the source of dataset transfers, although they can/will still serve this functionality
  - Simply said, they will play a role of the old Monarc Tier-2s
- VOLATILE:
  - Will not be used for the planned replication of datasets
  - Will not serve as the source for the centrally-operated data transfers
  - Can still be used to broker the jobs to close CEs (Tier-3 analysis...)
  - LOCALGROUPDISK SEs, Rucio cache storage ...

Job optimizations 2

- Production / Analysis
  - Run-1: 75% / 25% (slots occupancy ~ cputime usage)
  - Run-2: 90% / 10% (not even a rough estimate)
    - Bulk of analysis (Derivation) moving to (group) production
    - Remaining analysis will be shorter and I/O intensive
- Reduce the merging
  - Avoid it if possible (simulation, reconstruction)
  - Local merging – merge on the site, where the files to be merged are
- Jobs will produce bigger outputs
  - Good for tape storage
  - Bigger files transferred – good for efficient transfers (but less files to transfer)

Tier-2 site classification

- Based on ASAP metric
  - ATLAS Site availability for analysis
    - Analysis tests do all relevant checks of CE and SE availability
  - See Martina's talk later today
- 3 types of Tier-2s: AN EXAMPLE, to be refined, rediscussed
  - T2S : STABLE, ASAP > 90% in the last 3 months
  - T2U : UNSTABLE, 90% > ASAP > 80% in the last 3 months
  - T2V : VOLATILE, ASAP <80% in the last 3 months
- ICB policy:
  - T2V will be exposed to ICB which will inform the corresponding funding agency
  - IF T2V has ASAP < 80% for more than 6 months, it will be put in degraded mod
    - Storage will be removed from ATLAS
    - Can continue to contribute as Tier-3 (CPU)
- Metric might be too simple (network throughput), further experience needed

Consequences for production

- STABLE storage effectively doubles the space available for production:
  - ~50 out of ~80 Tier-2 SEs will be part of it (today's T2D, in 2015 T2S)
  - Not limited to Tier-1 disk space for brokering
- Much larger space to consolidate the production data
  - Less complex rules for data placement policies, less need for data migration
- Solving the always problematic full Tier-1 space and less used Tier-2 space which did occasionally block the production of some tasks in the past

Consequences for analysis

- Decrease the analysis pressure on Tier-1s, where the important data was stored
- Adapt the new mechanisms for data replication (like PD2P) to reduce the “unnecessary” migration of data

Job optimizations

Job type	Run-1 evts/out [MB]	Run-1 walltime [h]	Run-2 evts/out [MB]	Run-1 walltime [h]
MC evgen	5000	0.1	5000	0.1
MC simul	50-100 / 100	6-24	500-1000 /1000	8-30h - mcore
MC HITSmerge	1000 / 1000	0.3	-	-
MC digi+reco	1000 / 500	10	5000 / 2500	6 - mcore
MC AODmerge	5000 / 2500	2	-	-
Data reco	1000 / 500	6	1000 / 500	6
			10000 / 5000	8 – mcore (?)
Group Prod	10000 / 100	1	10000 / N * 100	2

Numbers are very rough, Run-2 speed up not included

Job optimizations 2

- Production / Analysis
  - Run-1: 75% / 25% (slots occupancy ~ cputime usage)
  - Run-2: 90% / 10% (not even a rough estimate)
    - Bulk of analysis (Derivation) moving to (group) production
    - Remaining analysis will be shorter and I/O intensive
- Reduce the merging
  - Avoid it if possible (simulation, reconstruction)
  - Local merging – merge on the site, where the files to be merged are
- Jobs will produce bigger outputs
  - Good for tape storage
  - Bigger files transferred – good for efficient transfers (but less files to transfer)

Conclusions

- New production and data management system provides many possibilities for further improvements and dynamic optimizations
  - Unfortunately, the commissioning was delayed, to give us more time for big changes well in advance of the Run-2 startup
- Fortunately, many of the changes can be implemented before the Run-2 starts
  - Many hooks are present already, we just need to use and tune them
- And even during the Run-2 we can afford to bring drastic improvements to our distributed system
- BUT, the production STABILITY will be the FIRST PRIORITY during data taking
  - In the last 2 years, we got used to a bit relaxed modus of operandi
  - In the next few months, we need to gradually tighten the overall stability to be ready for Run-2

Job optimizations 2

- Massive multicore for ~80% of production
  - All G4 simulation and all digi+reco
  - Effective drop in running jobs from 200k to 60k (20k 8-core + 40k single-core)
- JEDI dynamic resizing – tune the jobs to 6-12h
  - Avoid failures and cpu losses for very long jobs
- Automatic healing:
  - Split jobs too long
  - Increase memory requirements for out-of-memory failing jobs

Balancing the site usage

- Provide a steady flow of multi-core/single-core jobs
- Shorter jobs (much less than 2 days) better for fast turnaround
  - High priority jobs can get resources faster
  - Borrowed cpus can be drained sooner
- … and better for sites
  - Less cpu lost due to a downtime
  - Faster node draining for reservations or maintenance

FLAT Hierarchy

- STABLE storage with “stable” computing resources and fast network connections – A set of reliable resources
- 2nd layer of the less reliable, sometimes unavailable, pool of computing resources

- ATLAS plans to use the STABLE layer in a completely FLAT way
  - optimizing all the workflows (cpus, transfers, storage) for fast turnaround while minimizing the resource usage (minimize the transfers, balance disk usage...)

FLAT hierarchy

- Rucio supports distributed datasets:
  - A dataset replica can be distributed over many sites
- Strict ATLAS cloud model does not make much sense any more
  - Tasks are brokered to all stable sites, the point of consolidation of the production chain output
- A task still needs to be processed by many sites – job brokering will rely on
  - Input data proximity
  - Transfer cost matrix
  - Dynamic evaluation of transfer time (number of assigned jobs, recent history of past activity)
- New Prodsys and DDM:
  - Intermediate datasets (middle of the chain) will stay unconsolidated – distributed among the sites, skipping the output transfers
  - This might have to be limited to T2S sites only
- Final datasets consolidation:
  - Primary replicas will be consolidated
  - Secondary replicas can stay distributed at the sites that produced the files

Global cloud

- Tasks do not need to be assigned to any site – global task
  - The final consolidation can be delayed
- Final (primary) datasets do not need to be consolidated at all
  - Will be evaluated
  - Might be too difficult to manage (migration to tape)
- Big global task can be managed in a better way
  - Less tasks to manage, better activity overview, clearer prioritization
  - Large production tasks have been artificially split in Run-2 to run everywhere
- Experience with the new system is needed to choose the best option

Specializing the sites for workloads

- The pre Run-1 constraints for job placement are gone
  - Frontier instead of direct DB access → data reprocessing runs anywhere
  - High priority jobs (HLT reprocessing, Tier-0 spillover) with a short deadline could run everywhere
- But not all the sites are equal
  - Tier-1 vs Tier-2 is definitely not the correct answer
- ALL the jobs are important,
  - But not all the job types run equally well on all the sites
  - Some sites are slow for analysis but they are good for data reprocessing
  - Some sites are very big but cannot run 100% of heavy I/O jobs
- Differentiation was already used during Run-1 by limiting the job types through the fairshare (AGIS settings)
  - e.g. evgensimul=60%,all=40%
- But not all the jobs are EQUALLY important:
  - Some tasks have short deadline
  - Some large activities have close deadline (physics conferences)

Future specialization

- Sites will still be able to limit the heavy jobs to protect their infrastructure from the overload
- Dynamic specialization:
  - I/O expensive jobs will be automatically throttled by the central system based on recent history – keeping track of data transferred to site and reduce the heavy job assignment
- Migration from fixed bamboo queues to per task/job heaviness estimates
- Forced specialization:
  - ADC will specialize sites for certain activities, if the site provides custom resources (more memory per cpu, GPU availability …)
- New specialization classes will be defined after gaining experience with the new production system for custom requests with short deadline
  - Further site categorization needed to address processing power, network throughput and fast completion of urgent tasks