ATLAS WORLD-cloud and networking in PanDA

Fernando H. Barreiro Megino¹, Kaushik De¹, Alessandro Di Girolamo², Tadashi Maeno³, Rodney Walker⁴ on behalf of the ATLAS Collaboration

¹University of Texas at Arlington, ²European Organization for Nuclear Research, ³Brookhaven National Laboratory, ⁴Ludwig-Maximilians-Universität München

CHEP 2016 San Francisco
Rationale

- Original ATLAS Computing Model was designed as static clouds (=mostly national or geographical groupings of sites), setting data transfer perimeters
- Hierarchic model with clear distinctions in Tier 1-2-3 level
- Particular policies enforced in the workload management system:
  - Output of tasks (=set of jobs) had to be aggregated in the Tier 1s \(O(10)\)
  - Tasks have to be inflexibly executed within a static cloud
- This model works, but is getting outdated and has a series of disadvantages
  - WLCG networks have evolved significantly in the last two decades and bandwidth has increased \(O(1000)\): limiting transfers within a cloud is no longer needed
  - Usage of sites is uneven. In particular Tier 2 storage was not optimally exploited
  - High priority tasks occasionally stuck at small clouds
- First try was to allow sites to belong to multiple clouds. WORLD cloud is completely breaking the boundaries
WORLD cloud

- Dynamic model, defining dynamically the grouping of processing sites for each task

- **Task nucleus:**
  - Task brokerage will choose a nucleus for each task based on various criteria
  - The task output will be aggregated in the nucleus
  - The capability of a site to be a nucleus is defined manually in AGIS (ATLAS Grid Information System): Tier 1s and the bigger Tier 2s are defined as nuclei

- **Task satellites:**
  - Run jobs and ship the output to the nucleus
  - Job brokerage selects satellites for each task, based on usual criteria (e.g. number of jobs and data availability)
  - Satellites are selected across the globe: a network weight will bias towards well connected nuclei and satellites
Configurator

Rucio

Storage occupancy

NWS

Network data

AGIS

Site role (nucleus/satellite), semi-static network link classification

Configurator

PanDA

Task brokerage

Job brokerage

Chooses the nucleus for a task

Chooses satellite sites for the task, where the jobs will run
Network data

- Rucio, FAX, PerfSonar events are collected in the ATLAS analytics platforms [1].
- The Network Weather Service [2] (DDM team) aggregates information from the platform.
  - Per source-destination pair:
    - #files transferred in last hour
    - #files queued
    - Throughput according to FTS (aggregations for last 1h, 1d, 1w)
    - Throughput according to FAX
    - PerfSonar metrics (latency, packet loss, throughput)
- AGIS also provides semi-static link classification to be used as a backup.
- Configurator agent downloads and processes this information every 30 min. Data is cached in a key-value table in PanDA DB.
  - Table is extensible for any new metric without modifying the structure

Presentations in this conference:
[1] Ilija Vukotic et al., “Big Data Analytics Tools as Applied to ATLAS Event Data”
[2] Mario Lassnig et al., “Using machine learning algorithms to forecast network and system load metrics for ATLAS Distributed Computing”
Task brokerage: nucleus selection

- One nucleus is chosen for each task
- Nuclei must fulfill all hard limits:
  - Be in active state and be able to execute the workload
  - Have over 5TB free space - the free space includes an estimation of the space to be filled by its pending tasks
  - The number of output files transferring to the nucleus must be below 2k
- Eligible nuclei compete through a combined weight, based on data locality, total RW (remaining work) and available storage size in the nucleus.

\[
RW = (n_{Events} - n_{EventsUsed}) \times cpuTime
\]

\[
weight = \frac{1}{totalRW} \times \frac{availableInputSize}{totalInputSize} \times tapeWeight \times \frac{freeSpace}{totalSpace}
\]
Job brokerage: satellite selection

- 10 satellites are selected to execute the task
- The satellites must be able to run the jobs (RAM, walltime, core count, queues, SW releases and other settings)
- Sites must have <150 files in the transfer queue to the nucleus
- Eligible sites compete on a weight basis (#jobs, available & missing data)
- We have added the network weight, looking for balance between good throughput and queue length of output files
  - Throughput of FTS transfers satellite→nucleus
  - Queued output files in Rucio/FTS satellite→nucleus
Example: Top connected sites to Nucleus **AGLT2** (Michigan)

Average network weight last 24h (8 Sept 2016)

Plot shows best network weights belong to some of the “obvious” links (e.g. intracloud in the US), but also exploits other links with good throughput and low queues.
Status: Impact on T2 disk space usage

- WORLD cloud was fully activated end March 2016
- Nuclei being added progressively
  - Currently T1s and ~20% of T2s (tentatively more T2s will be added)
- Extending task output recipients to T2s is starting to have a positive impact on the overall disk usage (more primary data on T2s)

Storage plots from David Cameron’s space monitoring:
http://adc-ddm-mon.cern.ch/ddmusr01/plots/
Status: Impact on output file transfer duration

Implementing dynamic, controlled world-wide clouds does not have penalties on the transfer durations.

End March: WORLD cloud turned 100% active

NW weight and hard queue limit implementations
Observations and future work

- Some sites suffered initially under transfer load during heavy campaigns
  - Not because of inter-cloud transfers, but because their bandwidth was insufficient for the Nucleus role
- Hard limit queue controls worked fine to alleviate issues and deviate the traffic from blocked sites
  - Limits trigger also during unrelated, “accidental” massive transfers
- Further downstream controls could be implemented to e.g. avoid already assigned jobs to run while their nucleus is stuck
  - Pause overloaded nucleus in job brokerage (using a higher queued file threshold to avoid waves)
- We have tried to optimize **output** file transfer, but still need to include some optimization for **input** file transfers
  - This case needs to be solved together with the DDM team, since it involves further uncertainties (multiple copies, tape staging, etc)
- Reduced operational effort/manual interventions to re-broker tasks
- Need to improve analytics data for gridwise analysis