GridPP Storage and Data Management

October 31, 2018

(This document normally lives at `http://storage.esc.rl.ac.uk/weekly/`)

Attendance, in roughly joinological order Jens (chair+mins), Matt, Rob, Duncan, Dan, Chris Hawkes (Bham), Raja, Teng, Mark, JohnH, Sam, Tim, Winnie, Alastair, Elena, Robert, Roger Jones, Alessandra, PeteC, Simon, Brian

Note the chatlog is pasted in directly and could do with a bit of tidying up, since Vidyo does not make this easy.

Problem Statement

- Birmingham has two SEs, but will be decommissioning their DPM SE
  - Birmingham was 40% ATLAS and 60% ALICE, with ATLAS using the DPM and ALICE using EOS
  - The capacity of the DPM is about 350 TB
  - EOS was set up for ALICE because it was "easier" than to migrate them to DPM
- ATLAS intend to consolidate their storage endpoints
- Heavy network traffic into Manchester
  - At least some of which is caused by ATLAS jobs at Birmingham

Discussion

Taking the last problem first, the network traffic to Manchester:

- May need improved monitoring at Manchester
  - Could take some of the load to RAL
  - Still need to fully understand the problem
  - Tim can monitor ATLAS usage of RAL, could help with the monitoring at Manchester
- For now, continue with the current setup

Mark suggests setting up a EOS-based SE for ATLAS at Birmingham, once the DPM has been decommissioned (as its hardware will be needed to set up a new SE.)

> We have agreed to propose setting up an EOS-based SE at Birmingham to ATLAS-UK and let them decide whether they want this or not. This document could be the proposal, or the basis for the proposal.

In favour of the proposal are:

- Setting up a second SE will be relatively easy, because Birmingham already run an EOS instance for ALICE.
- The consolidation of ATLAS endpoints will happen on a longer timescale.
- The proposed SE starts empty, so no migration of existing DPM data (migration is hard and risky)
- Running two EOS SEs would be less effort than an EOS SE and a DPM

Against the proposal are:

- ATLAS will be consolidating their endpoints in the UK, so the short term effort may not be worth it.
  - The overall contribution of Birmingham to ATLAS is relatively small
  - Any additional contribution to the ATLAS pledge could be done within SouthGrid
  - Running one EOS SE will necessarily be less effort than running two
- Some unknowns, e.g. will it be dual stacked, will it support WebDAV?

Proposed Timeline

- Early October - decommissioning of DPM begun.
- 7 December - DPM turned off
- December - set up (proposed) EOS SE for ATLAS
- December - initial testing
- Q1 2019 - configured AGIS to use endpoint

Risks

The following is a slightly simplistic summary of the risks identified during the call.

n	descr.	LI	IM	II	Comment
1	AGIS configuration difficult	4	3	12	None
2	Problems due to lack of EOS skills in GridPP	1	4	4	Ask CERN for help?
3	EOS deployment for ATLAS at Bham causing problems for ALICE	2	3	6	Turn ATLAS SE off again
4	EOS doesn't support ATLAS workloads well	2	4	8	Tests suggest it will be OK

A general mitigation is that the proposed SE starts empty: thus, if there are any problems with ATLAS using it, we can turn it off. We also noted that CERN use EOS for all experiments, so if problems are found, they would be specific to GridPP or Birmingham.

Chatlog

```
Mark
He's from Bham
MS
Robert
Peter Clarke is in the US so might not make it
RA
OK, ta.
https://twiki.cern.ch/twiki/bin/view/LCG/ContentDeliveryCaching
Dewhurst
Is there a link to where the slides are stored, or will people be sharing their screens?
D
Teng
I didn't upload the slides. I could send by email if needed.
T
Daniel
just to confirm that Bham compute is all vac based now
DP
Mark
Technically no, but the 2 WNs left on CREAM are just for tests basically.
And I'm going to turn these off by the end of the year :)
MS
Samuel
Also, Mark, how are you going to configure your EOS instance, if you make one for ATLAS?
(Would this be a resilient store with replicas, or?)
SC
Mark
It would be single replica but backed by hardware RAID
MS
Samuel
Okay, so lower-resilience than the Alice EOS instance?
SC
Mark
Need to do that because otherwise we can't provide enough storage
```

MS
Samuel
Right, that's what I was thinking.
SC
Daniel
is the limitation at the manchester side WAN / DPM / hardware storage perfomance?
DP
Samuel
Dan: this is something which I (and Roger) would both like to understand....
SC
Duncan
How many job slots at Bham?
DR
Alessandra
we are looking at increasing the network bandwidth. But htis is not only a manchester prob
the green is incoming traffic btw
so these are jobs writing to the storage
AF
Samuel
Writing *to* the storage? So they're completed jobs staging back?
SC
Duncan
I'm not sure that is the case
DR
Alessandra
yes
we are THE Bham storage
we = manchester
AF
Samuel
So, the problem isn't remote data access from files at the start, it's the stageout?
SC
Alessandra
both
a diskless site have to access data and write them back somewhere
AF
Samuel
Okay, but which is worst? Caching would only really fix the stage in
SC
Duncan
Imperial also reads from QMUL and I see about 10 Mb/s per job so 1000 jobs = 10 Gb/s
Which is why I asked how many slots at Bham
DR
Alessandra
which at the moment fills our pipe
AF
Teng
Caching could cut ~50% of the stage in traffic, but cant do anything to the staging out
T
Alessandra

jobs concurrent
AF
Samuel
Exactly, Teng.
So, we're definitely limited by the ability of ATLAS to manage stageouts and do something
SC
Today at 10:28 AM
[Which is really a "datalake" problem]
SC
Alessandra
this is why we need to wait for what they come out with from the WLCG access WG
before adopting a solution
IMO
Incidentally Stephane is co-coordinator of that WG with Ilija
AF
Samuel
Indeed.
SC
Alessandra
so ATLAS is not a passive actor there
AF
Duncan
It was not entirely unpredictable that reading data over the WAN is likely to increase WAN
DR
Alessandra
Mario also heavily involved
I attend the meetings
AF
Samuel
I should note that Teng is one of the people doing the test work for the Access TF, though
[Which is really a "datalake" problem]
SC
Alessandra
this is why we need to wait for what they come out with from the WLCG access WG
before adopting a solution
IMO
Incidentally Stephane is co-coordinator of that WG with Ilija
AF
Samuel
Indeed.
SC
Alessandra
so ATLAS is not a passive actor there
AF
Duncan
It was not entirely unpredictable that reading data over the WAN is likely to increase WAN
DR
Alessandra
Mario also heavily involved
I attend the meetings

AF
Samuel
I should note that Teng is one of the people doing the test work for the Access TF, though
Duncan
Manchester needs a network upgrade
DR
Simon
so, slight tangent, but the conclusionin this meeting the other week that RHUL should swit
Samuel
I think our conclusion was that you should switch to being a site with simpler storage pro
SC
Simon
I would also be interested to know what cache size the simulation would predict for RHUL,
SG
Duncan
4600 slots could draw around 46 Gbps!
DR
Simon
I misunderstood then, thanks Sam.
SG
Roger
I'm afraid I have to leave
RJ
Elena
IC is using QM storage for atlas. IC has settings type=evgensimul:100%,type=evgen:100%,typ
EK
Teng
My suggestion is to deploy a transparent cache because that's easy and only very small amo
T
Duncan
QMUL have 20 Gbps network...
DR
Samuel
Sure, but we still need a solution for the outputs, Teng.
SC
Elena
@duncan: and Manchester?
EK
Duncan
10 Gbps
See the plot Elena
Today at 10:35 AM
Answering my own question Bham has 1584 cores
DR
Raja
Apologies - I have to leave now.
RN
Alessandra
we should save this chat
AF

jens
I always save the chat for the minutes
JJ
Teng
@Simon sorry didn't see your question. That will be 100-200T if that 4600 slots are for pr
I could simulate that in minutes
T
Today at 10:42 AM
Simon
Thanks Teng
SG
Today at 10:51 AM
Brian
perhaps tuning amount of MC is something to be done over next 2-3 months
BD
Today at 10:53 AM
Paige
Apologies, got another meeting
PW
Simon
Teng, we mainly run analysis jobs, what does that mean?
SG
Teng
I'm not sure. I can run the simulation now. What's the name of your queue(s)?
T
Today at 11:00 AM
Duncan
ANALY_RHUL
DR
Today at 11:03 AM
Daniel
have to go
DP
Teng
OK. I'll email the plots later
T
Today at 11:06 AM
Elena
Configuring EOS in AGIS is tricky because there is no much examples
EK
Today at 11:07 AM