

# Table of Contents

|                              |          |
|------------------------------|----------|
| <b>Question.....</b>         | <b>1</b> |
| <b>Full answers.....</b>     | <b>2</b> |
| CERN.....                    | 2        |
| hephy-Vienna.....            | 2        |
| KI-LT2-QMUL.....             | 2        |
| UKI-LT2-RHUL.....            | 2        |
| RO-13-ISS.....               | 2        |
| Nebraska.....                | 2        |
| INFN-ROMA1.....              | 2        |
| NDGF-T1.....                 | 2        |
| BEgrid-ULB-VUB.....          | 3        |
| NCG-INGRID-PT.....           | 3        |
| IN2P3-IRES.....              | 3        |
| LRZ-LMU.....                 | 3        |
| CA-WATERLOO-T2.....          | 3        |
| CA-VICTORIA-WESTGRID-T2..... | 3        |
| Taiwan_LCG2.....             | 3        |
| IN2P3-SUBATECH.....          | 3        |
| asd.....                     | 3        |
| MPPMU.....                   | 3        |
| INFN-LNL-2.....              | 4        |
| Australia-ATLAS.....         | 4        |
| KR-KISTI-GSDC-02.....        | 4        |
| UKI-LT2-IC-HEP.....          | 4        |
| -UCL.....                    | 4        |
| UKI-SOUTHGRID-BRIS-HEP.....  | 4        |
| GR-07-UOI-HEPLAB.....        | 4        |
| UKI-SOUTHGRID-CAM-HEP.....   | 4        |
| USC-LCG2.....                | 4        |
| EELA-UTFSM.....              | 4        |
| DESY-ZN.....                 | 4        |
| PSNC.....                    | 5        |
| UAM-LCG2.....                | 5        |
| T2_HU_BUDAPEST.....          | 5        |
| INFN-Bari.....               | 5        |
| IEPSAS-Kosice.....           | 5        |
| IN2P3-CC.....                | 5        |
| NONE_DUMMY.....              | 5        |
| WEIZMANN-LCG2.....           | 5        |
| RU-SPbSU.....                | 5        |
| USCMS_FNAL_WC1.....          | 5        |
| RRC-KI-T1.....               | 5        |
| vanderbilt.....              | 5        |
| UNIBE-LHEP.....              | 6        |
| CA-SFU-T2.....               | 6        |
| _CSCS-LCG2.....              | 6        |
| T2_BR_SPRACE.....            | 6        |
| T2_BR_UERJ.....              | 6        |
| GSJ-LCG2.....                | 6        |
| UKI-NORTHGRID-LIV-HEP.....   | 6        |
| CIEMAT-LCG2.....             | 6        |
| a.....                       | 6        |
| T2_US_Purdue.....            | 6        |

# Table of Contents

## Full answers

|                                       |    |
|---------------------------------------|----|
| IN2P3-LAPP.....                       | 7  |
| TRIUMF-LCG2.....                      | 7  |
| KR-KISTI-GSDC-01.....                 | 7  |
| GRIF.....                             | 7  |
| IN2P3-CPPM.....                       | 7  |
| IN2P3-LPC.....                        | 8  |
| IN2P3-LPSC.....                       | 8  |
| ZA-CHPC.....                          | 8  |
| JINR-T1.....                          | 8  |
| praguelcg2.....                       | 8  |
| UKI-NORTHGRID-LIV-HEP.....            | 8  |
| INDIACMS-TIFR.....                    | 8  |
| TR-10-ULAKBIM.....                    | 8  |
| prague_cesnet_lcg2.....               | 8  |
| TR-03-METU.....                       | 8  |
| aurora-grid.lunarc.lu.se.....         | 8  |
| SARA-MATRIX_NKHEF-ELPROD__NL-T1_..... | 8  |
| -UNIBA.....                           | 9  |
| DESY-HH.....                          | 9  |
| T3_PSI_CH.....                        | 9  |
| SAMPA.....                            | 9  |
| INFN-T1.....                          | 9  |
| GLOW.....                             | 9  |
| UNI-FREIBURG.....                     | 9  |
| Ru-Troitsk-INR-LCG2.....              | 9  |
| T2_Estonia.....                       | 9  |
| pic.....                              | 9  |
| ifae.....                             | 10 |
| NCBJ-CIS.....                         | 10 |
| RAL-LCG2.....                         | 10 |
| T2_IT_Rome.....                       | 11 |
| BNL-ATLAS.....                        | 11 |
| FZK-LCG2.....                         | 11 |
| INFN-NAPOLI-ATLAS.....                | 11 |

# Question

Question 5. A storage option that is currently being explored is so called "storage less" sites. These are sites that provide no reliable, permanent storage for data. However, they may provide some storage in the form of an autonomous cache, as an optimization. Are you aware of features that are currently missing in the storage stack that you would find useful, either as a site offering traditional storage, or as a site adopting the storage-less model? Do you believe that deploying an autonomous cache would be feasible and would save you money or support effort?

# Full answers

## CERN

n/a

## hephy-Vienna

We would not gain from a storageless effort. While we run a grid site, the support for the local physics community is paramount. They need a permanent storage to store their data. If the storage allocated for the experiments is maintained by some caching algorithm, it would be fine with us.

## KI-LT2-QMUL

Our site is expected to help support the storage requirements of storage less sites. The WAN network connection is important and understanding the network requirements would be very useful.

## UKI-LT2-RHUL

Would definitely be interested in setting up an autonomous cache. It should interface to an underlying POSIX-like system and make use of the direct access to that available on worker nodes so as not to introduce a bottleneck through the cache. The current storage (DPM) and other backend storage options we would use are distributed across many servers to achieve the required throughput. A service that is simpler to install and configure than DPM would require less effort. DPM seems to have a lot of major upgrades which take a lot of time and effort. Apart from that it is fairly low effort to run.

## RO-13-ISS

For ALICE usage a cache would not make sense

## Nebraska

Deploying an autonomous cache may benefit us in allowing us to reduce replication factors and thus grow our effective storage size without cost. Deploying a subset of our storage as a cache is a likely move soon but we cannot do that without interactions and approvals from our VO (CMS).

## INFN-ROMA1

At the current level of integration of services, implementing a cache in the storage stack in our site would not save significant effort, as the bulk of the work is coming from non-grid usage.

## NDGF-T1

Lack of support from the experiments other than ATLAS for running with just cache and no local storage system. We already run ATLAS computing on some sites in this fashion.

## **BEgrid-ULB-VUB**

We have a big mass storage dedicated to CMS, but also to some smaller experiments, for some of which we act as T0, so we will never go storage-less. Our users require a big amount of that storage too, that have helped them a lot. For storage-less sites, as AAA model advocated by CMS, bandwidth of other sites are taxed, which is a way to deport budgets on other sites that then have to accommodate the increase in WAN. We have seen this for the past few months with sometimes large xrootd requests coming from outside.

## **NCG-INGRID-PT**

Might be feasible but it will not save money.

## **IN2P3-IRES**

Deploying and maintaining the storage requires only few manpower. However, switching to a disk-less site would made for us much more complicated to get funding for the CPU (having a storage gives our center a real added value when dealing with the local funding agencies).

## **LRZ-LMU**

Yes, autonomous cache would be feasible. It might save some support effort. It will increase storage cost since capacity for primary data must be provided in any case and additional cache storage comes on top.

## **CA-WATERLOO-T2**

Probably no benefit in our setup as a "cache" would need allocation and dedication to be useful. Over a decade of dcache support so no short/medium term benefit going to another technology at this point either.

## **CA-VICTORIA-WESTGRID-T2**

Hard to say without knowing more about what such an autonomous cache is. This sounds like a different approach to DDM, but if the infrastructure and storage system run at sites is still the same, it will not simplify or reduce effort. A dCache SE can be marked as a volatile Rucio SE and used like a cache, but that doesn't change anything for the site, you're still running dCache.

## **Taiwan\_LCG2**

I am not aware of any missing for autonomous cache. And, I do believe deploying autonomous cache would save money and efforts for small sites

## **IN2P3-SUBATECH**

Would be feasible but would not spare much effort

**asd**

## **MPPMU**

no

## **INFN-LNL-2**

### **Australia-ATLAS**

Definitely save money, but our networking to the world could cause higher job latencies

We already use ARC-Cache for >10 years and two other T3 clusters (ARNES, SiGNET-NSC) are cache only clusters. This mode certainly makes opportunistic resources clusters trivial to use for ATLAS. Depending on the WAN bandwidth, T3 clusters can run full ATLAS workload or they can be throttled to less I/O intensive payloads.

### **KR-KISTI-GSDC-02**

Of course, I think the labor effort of storage-less sites can be reduced. On our site, quite a few SAS disks became a problem and we were able to replace them. If we replace these disks with SSDs and do not use disks, I think it will be a great help to reduce replacement work. But I do not know for sure about money.

### **UKI-LT2-IC-HEP**

No.

### **-UCL**

### **UKI-SOUTHGRID-BRIS-HEP**

A light storage element for HDFS - directly using the name nodes for file system information instead of own DB - just checking the permissions for a given path

### **GR-07-UOI-HEPLAB**

### **UKI-SOUTHGRID-CAM-HEP**

Planning to move to "storage-less" configuration soon. Autonomous cache feasible and would save some effort (but not enough). Ideal solution to minimise support effort would be true storage-less solution.

### **USC-LCG2**

No

### **EELA-UTFSM**

### **DESY-ZN**

Not really. To keep the operational costs low at the site, we prefer reliable hardware

## **PSNC**

YES, believe that deploying an autonomous cache would be feasible and would save FTEs

## **UAM-LCG2**

## **T2\_HU\_BUDAPEST**

feasibility: not sure, savings: yes

## **INFN-Bari**

Xrootd and dCache

## **IEPSAS-Kosice**

I am not aware of features

## **IN2P3-CC**

As Tier 1, we are definitely not on the way to move to "storage less site". But we agree to said that a feature to do it easily will be useful for some other sites

## **NONE\_DUMMY**

blah

## **WEIZMANN-LCG2**

We need to have significant local storage for local users.

## **RU-SPbSU**

maybe

## **USCMS\_FNAL\_WC1**

Unclear -- possibly for analysis use, most production processing accesses events once & moves on.

## **RRC-KI-T1**

Mechanism of delivering tasks to data is good enough. Cache for "storage less" sites is data storage too, that should be supported by site.

## **vanderbilt**

no

## **UNIBE-LHEP**

using ARC cache for 10 years now. Can work equally with or without local SE

## **CA-SFU-T2**

not feasible to us

## **\_CSCS-LCG2**

I'm not aware and have not enough info about the "autonomous cache" to answer

## **T2\_BR\_SPRACE**

I have no opinion on that.

## **T2\_BR\_UERJ**

We are not aware of those features.

## **GSJ-LCG2**

An autonomous cache would not save us money or effort (the computing model foresees sending jobs to data anyway).

## **UKI-NORTHGRID-LIV-HEP**

No idea

## **CIEMAT-LCG2**

I am not sure if more features would be required to better support the storage-less model, we have not explored caching technologies yet.

If using an autonomous cache meant offering significantly less disk space for CMS, then that would reduce money/effort, since there is less hardware (and fewer hardware types) to buy/maintain. However, much greater reduction in support effort would be achieved if this meant to get rid of a service like dCache, but that might prove difficult, given that it holds data for other non-WLCG communities (even if their total stored data is much lower than that of CMS).

**a**

## **T2\_US\_Purdue**

Erasure-coding for HDFS would be useful for providing redundancy using fewer disks.

So far we have considered deploying autonomous cache only as a performance enhancer (network bandwidth saver) to our existing traditional storage. We see it as feasible for our site, and even have plans to deploy such cache, for the purposes of improving performance. We do not know if it will save us money or support effort.



We don't think our current mode of operation as a CMS Tier-2 site will be possible without traditional local storage (i.e. caches alone). The observed network bandwidth on site between storage and compute systematically exceeds the WAN speed of the whole site.

## **IN2P3-LAPP**

We are aware of possible options and concerned by the medium/ long term future of DPM solution in the context of DOMA. We intend to study the usability of DPM volatile pools and see if we can benefit from integrated cache. Acting as a Tier 2 Nucleus for ATLAS, the site intend to take part in the DOMA R&D activities and expect to be able to maintain a reliable and permanent storage for ATLAS. Nevertheless, the site may also experiment autonomous cache as a common solution for WLCG and non-WLCG ESFRI projects, to access data elsewhere.

## **TRIUMF-LCG2**

Besides financial considerations, storage less sites is an option going forward into the future to perhaps increase site's reliabilities with easier cache setup and maintenance, and further cache utilization effectively optimized by the VO in principle. However, we have to be careful and ensure the overall systems is still scalable as data would still have to come from traditional sites and most likely primarily from Tier-1 centres. There has to be a proper balance with a proper optimization.

For the Canadian Tier-1 centre, we are operating a federation with production in two different locations. One location is configured to run only simulation tasks, which require the least I/O thereby minimizing transfers between the two location since the bulk of the disk storage is located at the other location. This mode of operation is quite optimal. Storage less sites could be assigned tasks that minimizes I/O (i.e. event generator and simulation tasks) and will be effective. We know from ATLAS, the great majority of cpu resources used on the Grid are for simulation tasks.

As for autonomous cache, it's good idea to have it at site or for a regional if full supported and can be easily integrated within workload and data management systems of the VO. At TRIUMF, we also have a Ceph based xcache setup available, using retired storage systems, however, ATLAS has no effective model to utilize it for a region or a site.

## **KR-KISTI-GSDC-01**

The idea of "Storage less" site would be helpful for those sites whom cannot afford to buy, install, operate any storage systems with the lack of manpower, physical space, and so on. Also a cache can be deployed on the compute nodes along with its local disks (hdd or ssd) or in-memory space. However without have a good network connection (low latency and large bandwidth) towards the nearest storage site, the "caching" idea would not quite effective. Having a good connection should be prior to the idea of "storage less" site.

## **GRIF**

Not planning on becoming a "storage less" site

## **IN2P3-CPPM**

no

## **IN2P3-LPC**

missing : high throughput network. Yes, deploying an autonomous cache would save money and support effort.

## **IN2P3-LPSC**

storage less sites could allow to spare money and support

## **ZA-CHPC**

Once setup EOS is pretty stable and just ticks over. The cost is the deployed storage admin costs are minimal.

## **JINR-T1**

N/A

## **prague\_lcg2**

WLCG storage less site would not help in our case, we host other VOs which want a permanent storage

## **UKI-NORTHGRID-LIV-HEP**

NO IDEA

## **INDIACMS-TIFR**

We are a storage site and would like to keep contributing our storage resources to WLCG sites.

## **TR-10-ULAKBIM**

Yes, I believe.

## **prague\_cesnet\_lcg2**

I do not think so.

## **TR-03-METU**

Yes, I believe

## **aurora-grid.lunarc.lu.se**

## **SARA-MATRIX\_NKHEF-ELPROD\_\_NL-T1\_\_**

NA, we will not be a storageless site

## **-UNIBA**

## **DESY-HH**

Desy will stay a site with reliable storage for data. Desy hosts data for many communities and experiments outside WLCG and acts as T0 or T1 for them.

Desy is worried about storage-less sites, because this would put additional load on the network and on the traditional sites.

## **T3\_PSI\_CH**

-

## **SAMPA**

we are offering a traditional storage site. I believe that deploy an autonomous cache can be feasible and I could support that effort.

## **INFN-T1**

In general, any cache infrastructure to be efficient should be more performant than the "permanent storage"

## **GLOW**

Deploying an autonomous cache would be feasible but with minimal impact on the money or support effort.

## **UNI-FREIBURG**

no - tested an Xcache instance here - not enough cache hits (yet) improvement: job distributing site is "aware" of the contents of the cache jobs requesting files from different sites can benefit from a dynafed instance

## **Ru-Troitsk-INR-LCG2**

No

## **T2\_Estonia**

International network speed between sites rises fast, so way not share cache/storage and provide anykind of autonomous cache.

## **pic**

We operate a Tier-1 site for ATLAS, CMS and LHCb. Also, a fraction of the ATLAS Tier-2 in Spain is co-hosted in PIC, as well as a Tier-3 infrastructure for IFAE ATLAS users. We are indeed studying in detail the data access patterns in PIC storage for CMS, and how CMS jobs that run in PIC access the input data, to better understand how to improve the a) use of the storage we provide to CMS, and b) if the deployment of a cache might be useful for some kind of CMS workflows (aka analysis jobs). It might be the case that the deployment of a cache might reduce the net capacity deployed in the site, still to know how big this impact

could be, but the support effort is not likely to decrease in our site. Some other sites in the region might opt to deploy caches, and PIC might become a data exporter for these caches. This might have an impact on how the storage is deployed, in its performance, and in the underlying network. All these effects need to be addressed.

A general feature that we miss in the software we use to manage storage (dCache in our case) are more knobs to control or protect us from high load from users. We would like for instance to try and serve several VOs from the same disk server, but currently we don't do that because we are not able to set max movers limits per VO, and any VO filling up the queue could alter the transfers of all other VOs in the same server

## ifae

We operate a Tier-1 site for ATLAS, CMS and LHCb. Also, a fraction of the ATLAS Tier-2 in Spain is co-hosted in PIC, as well as a Tier-3 infrastructure for IFAE ATLAS users. We are indeed studying in detail the data access patterns in PIC storage for CMS, and how CMS jobs that run in PIC access the input data, to better understand how to improve the a) use of the storage we provide to CMS, and b) if the deployment of a cache might be useful for some kind of CMS workflows (aka analysis jobs). It might be the case that the deployment of a cache might reduce the net capacity deployed in the site, still to know how big this impact could be, but the support effort is not likely to decrease in our site. Some other sites in the region might opt to deploy caches, and PIC might become a data exporter for these caches. This might have an impact on how the storage is deployed, in its performance, and in the underlying network. All these effects need to be addressed.

A general feature that we miss in the software we use to manage storage (dCache in our case) are more knobs to control or protect us from high load from users. We would like for instance to try and serve several VOs from the same disk server, but currently we don't do that because we are not able to set max movers limits per VO, and any VO filling up the queue could alter the transfers of all other VOs in the same server

## NCBJ-CIS

Our costs are not very high at the moment so we do not think about storage-less models.

## RAL-LCG2

Most features that a storage stack might need are "available" but in no fit state for production. One feature of Caches that I would like to see implemented is the ability to cache files only once they have been requested more than once. At RAL, observation of our XRootD caches has shown that approximately 95% of files are used just once before they are removed from the cache. However the 5% that are used more than once made up about 30% of the total usage. Caching files when they are requested for the second time would allow a much smaller cache to fulfil the same job.

Experiments need to build much more flexibility into their workflows management system. A large fraction of jobs (MC simulation) require very little data access and so could be run at a "storage less" site with little drop in efficiency. If Experiments were better at matching jobs to the available resources, there would be far less need to deploy caches to enable different types of workflows to run at certain sites.

Experiments could also put effort into fetching data more effectively within their jobs. One of the stated reasons to have a cache is to reduce latency and hence improve efficiency of jobs. If a job could start fetching the data it will need before the job actually requires it, this would achieve the same goal. This might not be possible for all job types (e.g. analysis) but if it could happen with some. These jobs could be run efficiently at storage-less (and cache-less) sites.

It is feasible to deploy caches, but it would cost money and effort to support them.

## T2\_IT\_Rome

No. No.

### BNL-ATLAS

As a T1, our main focus is to provide reliable, scalable and archival storage services. But we do have caching service deployed as well, based on XCache. We expect majority data transfers between dCache and grid clients will still go through the dCache doors, but a caching layer in front of users/clients can help reduce cost and latency for applications.

### FZK-LCG2

### INFN-NAPOLI-ATLAS

I think that deploying an autonomous cache would be feasible. Still some effort is needed to optimize the caching solutions, to integrate with the experiment environments and to train users to get advantages from the use of local caching. I don't think this solution would save effort in larger sites, like large T2s.

-- OliverKeeble - 2019-08-22

-- MarioLassnig - 2019-08-23

---

This topic: LCG > QosSurveyAnswersQ5  
Topic revision: r3 - 2019-09-20 - OliverKeeble



Copyright &© 2008-2021 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.  
or Ideas, requests, problems regarding TWiki? use Discourse or Send feedback