



7th TF-Storage meeting
Thursday-Friday, September 9-10, 2010
Poznan, Poland

Table of contents

Table of contents	1
Minutes	1
1. Welcome and apologies	1
2. Approval of agenda	1
3. Vendors' presentations	2
4. User communities	3
5. National updates	4
6. FileSender update	6
7. EC project calls and funding sources	6
8. Next meeting, AOB, and close	7
List of participants	8

Minutes

1. Welcome and apologies

The seventh TERENA Storage Task Force meeting was held on 9-10 September, 2010 at PSNC, Poznan, Poland. Jan Meijer (UNINETT), the chair of TF-Storage, welcomed the 25 participants (see the list of participants enclosed) and asked for a roll call. Paul Dekkers, Rogier Spoor (SURFnet), Cezary Dubnicki (9LivesData), Andreas Landhäußer (T-Systems Sfr GmbH) and David Corney (RAL) had apologised before the meeting. The TERENA secretary was Peter Szegedi.

2. Approval of agenda

The proposed meeting agenda was agreed with the minor change that the *"Measuring storage performance - theory and practice (TF-Storage Action Item)"* discussion was postponed to the next meeting. The presentations are available on the TF-Storage website:

<http://www.terena.org/activities/tf-storage/ws9/agenda.html>

3. Vendors' presentations

- Toine Beckers (DataDirect Networks) gave a presentation about DDN's high performance, high capacity network storage solutions. He introduced three products; the entry level storage S2A6620, the high performance streaming and archiving storage platform S2A9900, and the embedded HPC solution SFA10K.

DDN's largest customer base is the LHC community. DDN provides more bandwidth to the top500 list than all other vendors together. DDN is the leading provider of affordable, high-availability storage for the next generation of particle physics research.

Among the products, S2A9900 has low latency and high performance all implemented in silicon. The performance in terms of IO speed benchmarks is sustainable for long periods of time (longer than 20 hours). Direct connection and RAID striping provide maximum data availability. DDN enclosure RAIDing can withstand the loss of 20% of system enclosures and drives while delivering full data availability. The Parallel Data Recovery Engine allows data reconstruction and integrity checking. Drives that issue SMART warnings or grow defects at an increasing rate are copied to spare drives. There is no statistical information available on SATA disk failures but 1/1000 disk failure per week can be a good estimate. S2A6620 is an entry level mixed workload platform. SFA10K is using SFA as Block Storage device. Storage Fusion Architecture (SFA) shortens the IO path from the application to storage, reducing latency, and increasing IOPS performance.

Slides:

<http://www.terena.org/activities/tf-storage/ws9/slides/10092010-7tf-storage-beckers.pdf>

- David Ranson (BlueArc) spoke about BlueArc's unified network storage systems. First, he presented the BlueArc's view on storage workflow in higher education and academic research environment. That starts with historical data storage, computational clusters with workspace storage, permanent storage and finally ends at long term archive solutions. There is a strong need to optimise and consolidate solutions, manage complexity, and simplify administration in that sector.

Regarding BlueArc's storage products David mentioned that SiliconFS file system is used as a common point of integration for all elements of the storage solution. That is the central engine that manages all data movement and it has advanced metadata optimisation. Cost-efficient performance can be achieved by moving disk-intensive metadata ops to a high-speed storage tier. BlueArc servers use dedicated metadata caches with very aggressive caching algorithms.

David introduced the BlueArc Mercury and Titan platforms as potential storage solutions tailored to higher education needs. BlueArc software deals with data management, data protection, and virtualisation. More about these features can be found in the presentation: <http://www.terena.org/activities/tf-storage/ws9/slides/10092010-7tf-storage-ranson.pdf>

In summary it can be said that BlueArc's solution can match with the exponential growth in data and the increasing pressure to reach conclusions faster. In a collaborative research

environment better control on costs can be achieved.

There was a question about the Parallel Adaptive Filesystem (PAFS) and its usage in BlueArc platforms. David answered that a new product will be launched in February 2011 using PAFS whether the standard is finalised or not by that time.

- Lukasz Heldt (9LivesData) introduced HYDRAsstor as a scalable secondary storage solution. HYDRAsstor is a scalable, disk based storage for backup with global de-duplication. It is owned by NEC but the critical backend component was developed by 9LivesData in Poland.

A secondary storage size could be 5x-20x bigger than the primary storage, may include many copies of the same data, and each data chunk can be stored 5-10+ times. There is a high potential for the de-duplication technology there. De-duplication saves disk space by eliminating duplicates in general. HYDRAsstor product provides global de-duplication using DataRedux, as well as enhanced performance, storage scalability, and data resiliency using Distributed Resilient Data (DRD). In the HYDRAsstor architecture accelerator nodes realise performance while storage nodes realise capacity. The file system is built by trees of blocks. There is no distinction between data and file system metadata. HYDRAsstor backend design details can be found on the slides:

<http://www.terena.org/activities/tf-storage/ws9/slides/10092010-7tf-storage-heldt.pdf>

4. User communities

- Paul Boven (JIVE) gave a presentation about high-speed continuous recording and playback for VLBI community. He introduced JIVE and talked about radio astronomy in general and Very Long Baseline Interferometry (VLBI) in more details.

Currently, for VLBI, special purpose recording machines (Mark5) and special purpose packs with 8 hard disks are used. Each telescope records at up to 1024 Mb/s that means 5 TB in a typical 12h observing run and after correlation the resulting dataset is a few GB. The total EVN storage pool (i.e., physical hard drives) is 1.7 PB, which is constantly on the move.

Paul spoke about the NEXPreS project where a dedicated work package deals with "Provisioning High Bandwidth, High Capacity Storage on demand". NEXPreS is a combined real-time and disk-based VLBI. More telescopes will join in the future to the project.

Slides:

<http://www.terena.org/activities/tf-storage/ws9/slides/10092010-7tf-storage-boven.pdf>

- Marcin Mielnicki (PSNC) presented the storage issues in the largest Polish digital library (Digital Library of Wielkopolska). The goals of digitalisation in libraries are twofold; to ensure unified access to content, and to provide long-term preservation. One of the main functions of a digital library is the storage of various digital objects and metadata.

PSNC has been developing the dLibra software for building digital libraries and the dMuseion software for building digital museums. PSNC also maintains the PIONIER Network Digital Libraries Federation (DLF) that is the integration point for majority of Polish digital libraries.

dMuseion software will soon be available for the public.

Digital Library of Wielkopolska contains 110 000 various digital objects, 2 800 000 presentation files and 780 GB presentation files. The digital library can be accessed via the Europeana portal. The detailed IT infrastructure and the digitalisation workflow can be found on the slides. The backend storage currently is: SUN StorageTek 2540 FC 2540 Array (12 x 300GB SAS) plus extension array (12 x 300GB SAS). 1 TB of 4.5 TB dedicated for WBC and currently used in 70%.

The descriptive metadata is either created manually or imported from MARC (library catalogues). DublinCORE metadata set is used. The master copies (in size of 10 TB with 1.3TB annual growth) are mostly stored in TIFF format. The presentation files are created fully automatically from the master copies. Master copies are stored at the local libraries and the digital library stores only the presentation copies.

Slides:

<http://www.terena.org/activities/tf-storage/ws9/slides/10092010-7tf-storage-mielnicki.pdf>

5. National updates

- Szabolcs Szekelyi (NIIF) gave a summary of NIIF/HUNGARNET current storage developments and deployments. There is an established national funding to reengineer the current NIIF infrastructure. Total financial range is €15M of which €5M is on distributed computing and data storage.

The national storage network has three sites with 500 TB raw data capacity in total. The plan is to upgrade to 1.2 PB. The procured hardware is Fujitsu-Siemens Eternus DX80 using redundant controllers and iSCSI interface. The development of the self-service management interface is ongoing. It is going to be web-based, implemented in Python/Django with the functions of RAID array management, volume management, and iSCSI target management.

There is a plan to build an IaaS cloud on top of that storage. The beta release is expected in October 2010. The service is going to be public for NIIF users and member institutions. KVM will be used as virtualization technology and OpenNebula will be used as cloud management system. The initial configuration will cover ~60 nodes (4 CPU cores and 12 GB RAM each) distributed to ~8 sites in Hungary. Its features will include self-service management, live migration of virtual machines between sites, and QoS provisioning (guaranteed minimum CPU performance for virtual machines).

Slides:

<http://www.terena.org/activities/tf-storage/ws9/slides/10092010-7tf-storage-szekelyi.pdf>

Zsombor Nagy (NIIF) talked about Chelonia that is a lightweight self-healing distributed storage. It was developed by the KnowARC project and maintained by the NorduGrid Collaboration. In Norway, Chelonia is deployed on 10+ (desktop) machines, one terabyte each. In Hungary the lack of manpower and resources significantly slowed down the deployment and also the development of Chelonia. EMI project is not interested in the

further developments.

Slides:

<http://www.terena.org/activities/tf-storage/ws9/slides/10092010-7tf-storage-nagy.pdf>

It was mentioned that NIIF's CluserGrid solution uses GANETI cloud management system: <https://code.google.com/p/ganeti/> Faidon Liambotis (GRNET) noted that they are currently using 3 clusters with 2, 12 and 32 nodes respectively. It is also used internally at Google. Live migration is initially supported only via DRBD (which it automatically manages) but GRNET developed a plug in for SAN (shared block) and NAS (shared filesystem) support. GRNET is currently writing a web frontend for clients.

- Martin Kämpf (SWITCH) spoke about various Swiss university projects. There is a local storage working group at SWITCH. One of its project-ideas is to crossover remote-storage between Universities, since there is no traffic cost within the SWITCH network and a commercial remote-storage solution can be rather expensive, anyway. The nanoo.tv project, initiated by Zurich University of Arts in order to broadcast-record educational content, is looking for searchable metadata solutions, centrally accessible data store, and stable archive or some sort of a backup. The DocExchange application has no publicly available source yet and only deployed at University of Basel. Clearly defined licensing model and special focus on usability for the community are essential, the lesson they have learnt.

Slides:

<http://www.terena.org/activities/tf-storage/ws9/slides/10092010-7tf-storage-kampf.pdf>

- Peter Vercimak (CESNET) talked about storage services' vision and plans in the Czech Republic. There are two national projects ongoing in order to build CESNET e-Infrastructure. The goal of both projects regarding storage is common; to build up and to put into service a storage system of three distributed large scale repositories for saving and sharing of large volume of data including archiving. The main purpose of CESNET's storage system is to provide an easily accessible and redundant data repository for academic and scientific community. From the user (students) point of view it should provide "unlimited storage capacities".

Technical concept is to have a HSM system composed of disk arrays and tape libraries. The data access method and communication protocol will be a combined NAS/SAN solution. Disk array (Tier1, Tier2) will be about 300 TB, tape library (Tier3) is planned to be about 3 PB capacity. More details about the architecture can be found on the slides:

<http://www.terena.org/activities/tf-storage/ws9/slides/10092010-7tf-storage-vercimak.pdf>

- Michal Jankowski (PSNC) gave a presentation about the Polish national data storage architecture and mechanisms. PSNC maintains the service platform for e-science, called PLATON, in Poland. Actually data storage is too complex and expensive to be done by universities and they have no expertise. Outsourcing the process may be the only or at least the most reasonable solution for them.

At PSNC there is a national data storage research and development project called KMD and a popular backup/archival service deployment project called PLATON-U4. The primary aim is

to support scientific and academic community in protecting and archiving their data. The potential customers are digital libraries, virtual laboratories, academic computer centres and network operators, research institutions, universities, and clinical hospitals. The details of the design assumptions, overall architecture, and use cases can be found in the presentation:

<http://www.terena.org/activities/tf-storage/ws9/slides/10092010-7tf-storage-jankowski.pdf>

There was a question about the real benefits of the PSNC solution compared to a commercial solution (e.g., 2 SUNs with iRODS on top). Actually, PSNC's solution is reliable and independent from commercial products such as iRODS. It is simpler and more user friendly. It is implemented in simple C++ code and not in Java. The access part is implemented in lower layers not to create bottleneck there. It is developed by a consortium consists of 10 universities. The actual development staff is about 8-10 people.

6. FileSender update

Jan Meijer (UNINETT) gave an update on FileSender <http://www.filesender.org/>, as usual. The value proposition of FileSender is to provide easy sharing of arbitrarily large files through a trusted intermediary. The history, the open source development community, and the service architecture were explained by Jan.

Because of the latest code security audit results the planned first release date has been postponed. The new target date is the 1st of December, 2010. The roadmap after 1.0 release will include the move of all admin UI to PHP, as well as the support for HTML5 UI, multi language, multiple storage back-ends, encryption, virus scan hooks, and scalability. The work plan is available here:

http://www.assembla.com/wiki/show/file_sender/Workplan2010

The developer community is still looking for beta testers, additional coder, release tester/release manager, and last but not least money for 2011.

Slides:

<http://www.terena.org/activities/tf-storage/ws9/slides/10092010-7tf-storage-meijer.pdf>

7. EC project calls and funding sources

At the end of the meeting, Peter Szegedi (TERENA) reported about the recent discussion with the European Commission's scientific officers. The officers have been asked to inform TF-Storage about the storage and cloud related EC policy and potential project calls.

Under the Research Infrastructures programme there is no specific part dedicated to storage but it is intrinsic to the activities undertaken in the distributed computing, supercomputing and data parts. Under the ICT programme there is neither a specific part dedicated to storage but it is included under the Software and services activities handled by Unit INFSO/D3. On the other hand, there are several activities on the subject of clouds and there are more to come. The

ongoing activities include the funding of research projects (in both, the ICT and RI programmes) and the ongoing definition of a standardisation roadmap. In the framework of the implementation of the Digital Agenda for Europe (DAE) there is an action item explicitly referring to clouds for government and science. It says: "Ensure sufficient financial support to joint ICT research infrastructures and innovation clusters, develop further e-Infrastructures and establish an EU strategy for cloud computing notably for government and science."

In return, the EC officers are wondering what the current motivations of the TF-Storage participants are, from which strategic perspective the Task force looks at storage and clouds, how it positions vis-à-vis the activities mentioned above and if it has plans to support these activities?

Slides:

<http://www.terena.org/activities/tf-storage/ws9/slides/10092010-7tf-storage-szegedi.pdf>

TF-Storage group has agreed to act together and inform EC officers on these questions. First, we need to agree on our strategic perspective as a group. Each TF-Storage participants should have their own strategic perspective, so the first step would be to collect those strategic views and see if we can get something out from those plans as a common strategy. The benefit of this can be twofold: on the one hand, the TF-Storage group itself can have a complete a view of storage and cloud related national perspectives, on the other hand, the EC officers can be informed about these strategic views that may help the policy makers to take the right assumptions. There is a danger if the policy makers relay on wrong assumptions, those policies cannot be easily followed by anyone of us.

TF-Storage group has agreed to start collecting the national strategic perspectives on storage and clouds.

ACTION 1 on Peter (TERENA) to approach the TF-Storage participants and ask their strategic perspective on storage and clouds (if any).

The answers will be discussed in small groups. Video Conference(s) will be scheduled in November 2010 in order to clarify the details and define a common strategy. EC officers can be informed on the results to be harmonised with the TERENA TF-MSP group, as well.

8. Next meeting, AOB, and close

The next coming TF-Storage meeting will be in January/February 2011, most likely hosted by NIIF/Hungarnet in Budapest, Hungary. The exact date will be agreed later on. The TERENA TF-EMC2 group mentioned that they are willing to co-locate their meeting with us.

List of participants

Name	Affiliation
Adam Zawada	PSNC
Auke Kuiper	DataDirect Networks
Brian Boyle	HEAnet
David Ranson	BlueArc
Faidon Liambotis	GRNET
Frode Sten	BlueArc
Guido Aben	AARNet
Iva Cabric	CARNet
Jan Furman	CESNET
Jan Meijer	UNINETT/eCampus
Josip Rodin	CARNet
Łukasz Heldt	9LivesData
Maciej Brzezniak	PSNC
Mario Vandaele	Belnet
Martin Kämpf	SWITCH
Michal Jankowski	PSNC
Paul Boven	JIVE
Pawel Sychala	PSNC
Peter Szegedi	TERENA
Peter Vercimak	CESNET
Rafał Mikołajczak	PSNC
Rosend Llurba	NCF
Szabolcs Székelyi	NIIF
Toine Beckers	DataDirect Networks
Zsombor Nagy	NIIF