SUCCESS STORY

How Open Source Software-defined Storage is containing the Data Explosion in Universities

The Academic Data Growth Challenge

Research and Medical Science

"We are always looking for ways to help our researchers work more effectively, which means giving them access to powerful and flexible IT resources."

— Simon Naughton
Director of Infrastructure & Operations
Swinburne University

Enterprises across the world are facing an uphill struggle managing their increasing data growth, and academic institutions are no exception. Indeed, many universities face unique challenges in this area.

Physicists have been using massive data sets for years. The famous CERN facility aptly demonstrates this with a data center that recently reached the staggering milestone of 200PB—kept in, of all things, a tape repository. The scale of data they generate is difficult to imagine. Particles collide in the Large Hadron Collider (LHC) detectors approximately 1 billion times per second, generating about one petabyte of collision data per second. Unsurprisingly, CERN doesn't even try to store it all.

Since the advent of high throughput genomics, biologists have joined the big data club, too. Life scientists require access to supercomputers to do their work. When you are engineering new proteins from genetic data and you want to see how and if those proteins will work in a living organism, you can't do it on a laptop—no matter how good your PhD or how ground-breaking your thoughts. These days, science is computer-aided, is run on huge data sets, and comes with a built-in, high-volume storage requirement. Universities regularly boast about their latest supercomputers as a way to attract the best researchers. Powerful compute and heavy-duty storage capability is a pressing concern. Research scientists are in the business of manufacturing information—they literally make knowledge.

Medical data also brings steep challenges. University hospitals need to store huge volumes of structured and unstructured data. This includes ultra-sensitive and strictly-regulated patient data and the huge data footprint of modern medicine created by MRI scans, X-Ray images, and ultrasound. With every improvement in technology, more data is generated, stored, compared, and analysed. Surgery is often routinely filmed, both as a teaching aid for medical students and for the value to researchers as new techniques are explored. Privacy concerns with sensitive patient data are a serious consideration, and levels of access control, security, and controls for information sharing have to be in place.

The Digitisation of Learning

University data storage challenges aren't all about science. Increasingly, universities are digitising teaching, recording lectures both for the students who miss them and to empower distance learning initiatives for off-campus students—so a year's worth of lectures as video files quickly becomes an enormous object store.

Students themselves are also creating huge volumes of data in the course of their studies—recording performances from dance to theatre, working on audio files from music to performance poetry, and even making feature-length films for journalism and media courses. Every year the cameras become more powerful, so the same level of recording produces more data.

The Acceleration of New Technology Accelerates Storage Requirements

Advances in technology across the board—from medical scanners to cameras to microscopes—bring increasing pressure to storage environments. New equipment invariably arrives every year, and invariably increases the storage strain. A new MRI scanner in a teaching hospital generates more data, upgrading ultrasound scanners from older black and white imaging to full colour generates more data, and the same thing happens when the Film and TV Studies team gets a new SLR camera: more data.

"Storage requirements are increasing all the time. As medical technology improves, so too does the size and quality of medical imaging records, while many departments now also produce surgery videos to support teaching and research. We needed to be able to store more and more large multimedia files."

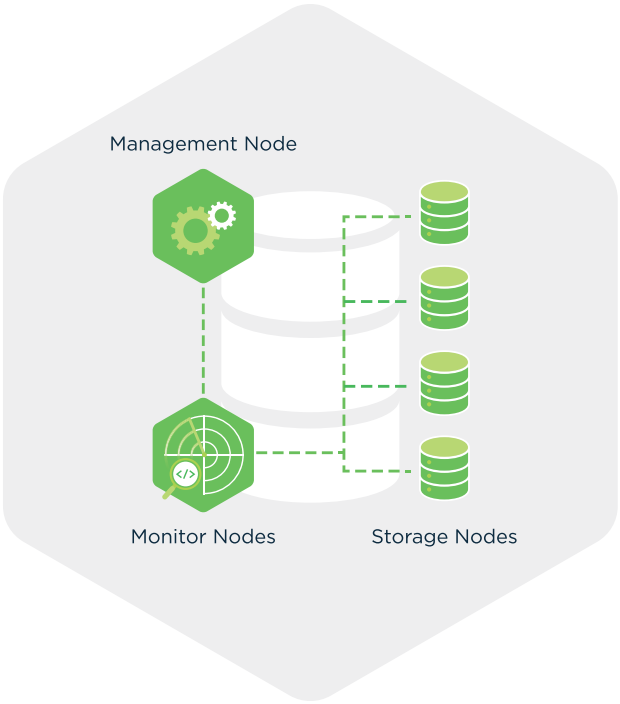
— Michael Nieporte
Head of IT Infrastructure
University Hospital Essen

Universities are about knowledge. This knowledge is being digitised, and storing all that digital knowledge is a bigger problem every year.

When the 'Traditional' Approach Stops Working

Simply adding array after array of proprietary storage rapidly becomes financially unsustainable. The ongoing annual budget increases to do this don't exist, which forces IT teams to find new approaches. The search is on for robust, scalable solutions that deliver the required capacity on a long-term basis without breaking the bank.

Open source software-defined storage software from SUSE provides universities with a robust, scalable, and cost-effective approach that can reduce costs by as much as 70 percent compared to traditional storage, without compromising performance.



- *Unified Cluster*
- *Resilient & Self-healing*
- *Open Source Software on x86*
- *High Performance*
- *Public Cloud-like Pricing*
- *Massively Scalable*
- *Hardware flexibility*
- *Reduced IT Costs*

What is SDS—Software-Defined Storage?

Unlike older-style storage approaches, software-defined storage (SDS) enables administrators to operate the software separately from the hardware. Just as server virtualization empowered the pooling of compute resources, SDS aggregates existing hardware resources, pooling storage space previously allocated to separate applications. SDS can be managed through a single interface, allowing for simplicity. SDS can be scaled out across a cluster, and policies can be set to control features and functionality. Best of all, SDS can be run on commodity hardware and proven open source software, generating major cost savings compared to traditional arrays.

While a software-defined approach might seem new, the logic within enterprise storage devices has always been written in software. It has only been in the last few years that hardware has progressed enough that enterprise storage software and hardware can now be separated. Software-defined storage isn't a new concept, but the ability to separate the storage software from the storage hardware is a new value proposition.

Taking the right approach to SDS will give you a highly scalable solution that radically reduces your storage costs in terms of both capital cost and operation expenditure. The solution must also be simple, robust, and flexible—all while delivering industry-leading functionality.

SUSE Enterprise Storage is based on Ceph, the leading open source software-defined storage technology. It is designed as a distributed storage cluster to provide virtually unlimited scalability, from terabytes to petabytes and beyond. It is a truly unified block, object, and file solution—ideal for Backup, Compliance, Bulk Storage, Large Data, and Active Archive applications.

SUSE Enterprise Storage is self-managing and self-healing. Its intelligent algorithms store data in a highly distributed manner. It continuously monitors data utilization and re-balances data placement. This means it can optimize system performance without the need for teams of administrators to constantly monitor and manage the storage.

How Universities are Benefitting from SUSE Enterprise Storage

"With SUSE Enterprise Storage, we have built a very high-performing and scalable storage landscape at a fraction of the cost than we would have been able to with traditional storage systems."

— Simon Naughton
Director of Infrastructure & Operations
Swinburne University

Swinburne University of Technology

Scientists at Australia's Swinburne University of Technology carry out research that is transforming industries, broadening our understanding of the world, and improving lives. Information technology is central to how modern research works, and the academics need ever-increasing compute power and data storage.

The university uses NeCTAR (the National eResearch Collaboration Tools and Resources). These shared cloud resources for Australian researchers are built on OpenStack for compute. However, storage is kept local due to latency concerns and data-sharing restrictions.

Swinburne recognised that software-defined storage was an ideal fit for its needs, because it could deliver high performance and scalability at significantly lower costs than traditional storage infrastructure.

Swinburne wanted to use Ceph because it is the most commonly used storage for OpenStack, but they lacked the skilled staff resources necessary for a community platform. In addition, most of the vendor-supported versions they looked at were licensed on a capacity basis, so license costs would increase in line with data volumes, making the total cost unaffordable.

After conducting short trial comparing SUSE Enterprise Storage to other vendors, the results confirmed SUSE as the preferred approach. Swinburne is now running two clusters, which support requirements for NeCTAR and a data protection initiative.

The new storage clusters are distributed and provide virtually unlimited storage, giving Swinburne the ability to add capacity as needed.

"SUSE Enterprise Storage offered us the best of both worlds—an enterprise-supported solution that is licensed per node, instead of on a capacity basis. This was a differentiator that made the product commercially viable for us."

— Simon Naughton
Director of Infrastructure & Operations
Swinburne University

"In our experience, SUSE Enterprise Storage offers much lower TCO. Not only have we cut administration time and effort, but also the solution enables us to practically eliminate downtime, which represents a huge cost saving and is vital for our students' productivity and satisfaction."

— Emile Bijk
Head of Network and Information Systems
HKU

HKU University of the Arts

At HKU University of the Arts in Utrecht, Netherlands, students create and share growing amounts of digital information. As camera picture resolution and frame rates have increased, so too have the storage requirements. With storage fragmented across a number of different sites, the university was developing a significant storage headache; a new approach was needed to ensure that they could continue to attract and serve students.

HKU wanted to consolidate their storage to make it more robust, more scalable, and easier to manage. Several isolated storage systems served the needs of multiple application servers, which meant that capacity wasn't optimised. It also presented other technical difficulties related to maintenance.

"If we needed to take a storage system down for maintenance or to increase its capacity, the hosted applications would also suffer downtime."

— Emile Bijk
Head of Network and Information Systems
HKU

By adopting a single, centralized storage environment—clustered for high availability—HKU aimed to increase performance and reduce downtime. This, in turn, would help students make the most of their time at HKU and increase the university's attractiveness to new applicants.

HKU decided to trial SUSE Enterprise Storage. It proved to be a great fit, providing extreme scalability at a lower TCO (Total Cost of Ownership). Centralising their storage created a single point of failure, with a single outage potentially bringing down all storage, so it was very important that the solution was resilient and that a high level of protection and support was delivered—so HKU worked with SUSE to get its deployment right.

HKU initially ran three separate clusters, but has now added a fourth. There has been a significant reduction in the time it takes to manage storage. There is no longer any application downtime when the system is upgraded. Maintenance can be done on any individual node without downtime as well: applications and users are automatically redirected to one of the remaining live nodes.

University of Valencia

Today's students have plenty of choice when it comes to deciding where to study. They tend to be attracted to academic institutions that not only boast a prestigious history, but are also tech-savvy and forward-thinking. Established over 500 years ago and the fourth-ranked university in Spain, the University of Valencia is both. Driven by the need to keep student services 'always on,' the university needed to achieve the flexibility and scalability required to manage terabytes of data—with peace of mind that the solution would be able to keep up with rapidly scaling data volumes.

"To provide the digital services that our students and staff require and expect in a reliable, timely fashion, we needed to be faster and more nimble. Our legacy storage systems were no longer suitable to support us with this goal."

— Israel Ribot
System Engineer
University of Valencia

The university was impressed by Ceph, an open source solution offering virtually unlimited scalability for unified file, block and object storage, all running on low-cost commodity hardware. SUSE Enterprise Storage was selected because of its ease of deployment, flexibility and cost-effectiveness—delivering an affordable way to keep up with surging data demands.

At present, the solution is running in the university's main data center, where they run 200 virtual machines and handle over 50TB of data—with the volume set to increase in the near future as the project scales.

Comparative testing between SUSE Enterprise Storage and the prior NFS system revealed both higher and more consistent performance, and three times faster speeds than the traditional NAS storage previously used for virtualisation.

"Installing 30 virtual machines with NFS took 2.4 times longer than installing a single virtual machine. With SES, the same test took only 1.5 times longer. And as we scaled from 10 to 20 to 30 to 40 machines, the SUSE solution's performance remained steady, while the NFS performance degraded in a linear fashion."

— Israel Ribot
System Engineer
University of Valencia

Costs have been slashed by over 40 percent.

"It usually takes three months to procure, install and configure a new mid-range storage system and migrate data over. Now, we can simply order a new customized and tested system from Thomas-Krenn, and add its capacity to the storage pool within days—at 70 to 80 percent lower expenditure."

— Michael Nieporte
Head of IT Infrastructure
University Hospital Essen

University Hospital Essen

With more and more departments going digital, Essen University Hospital has to store increasing volumes of data, such as radiology scans, pathology slides, and even surgery videos. The University Hospital's medical research scientists collaborate with front-line healthcare professionals, with the shared aim of improving outcomes for patients. To work as a team, scientists and doctors need to be able to share huge volumes of highly sensitive data securely. The nature of the data (patient data) left the university hospital uncomfortable with using the public cloud.

Secure collaboration isn't the only challenge: storing the results of diagnostic radiology exams (such as X-rays, ultrasounds, and MRI scans) from hundreds of thousands of patients admitted annually was causing costs to sky-rocket.

"Our storage environment was under constant pressure. We were storing all patient data on mid-range storage devices, which would quickly reach full capacity. This meant that we had to frequently expand or replace our storage systems, which was costly and time-consuming."

— Michael Nieporte
Head of IT Infrastructure
University Hospital Essen

With the university hospital's institute of pathology set to digitise its workflows, it was expected that another 300 to 400 TB of data per year would be added—growth on a scale that would quickly consume the entire storage budget in a single department.

Recognising that the traditional appliance route was not going to work, University Hospital Essen sought to avoid expensive proprietary arrays and expand its capacity by moving to open source software running on commodity hardware.

University Hospital Essen turned to SUSE Enterprise Storage. Designed as a distributed storage cluster, SUSE Enterprise Storage provides a single, software-defined storage solution that separates the storage management from the underlying hardware. Regardless of the actual infrastructure, all capacity is combined across multiple storage arrays, and can be grouped into one or several storage pools according to the specific requirements. By expanding a cluster with additional nodes when extra capacity is needed, the virtual storage pools can be scaled essentially without limit, and SUSE Enterprise Storage will automatically redistribute the data to make the best use of the additional nodes.

Being able to add storage capacity quickly and easily will enable University Hospital Essen to accommodate ever-growing volumes of data within budget.

For more information, contact your local SUSE Solutions Provider, visit us online or call SUSE at:

1-800-796-3700 (U.S. and Canada)
1-801-861-4500 (Worldwide)

SUSE
Maxfeldstrasse 5
90409 Nuremberg
Germany

www.suse.com