Dundee University supports world-class research with IBM data storage solution

Overview

- ***Challenge***
  Provide scalable storage infrastructure for College of Life Sciences, able to meet significant growth requirements; enable cost-effective long-term storage of experimental data; provide high-performance storage to support microscopy image processing
- ***Solution***
  Implemented IBM Tivoli Storage Manager (TSM) on two IBM System p servers, managing data on an IBM TotalStorage DS4400 with 11 Terabyte (TB) capacity, a DS4500 with 50TB capacity and two IBM System Storage TS3500 Tape Library systems
- ***Key Benefits***
  Microscopy images can be easily and cost-effectively stored and accessed, enabling new research projects to exploit historical data; high-performance cluster can access data at 400 Megabyte (MB)/s, boosting system performance; Tivoli software simplifies management, helping to keep staffing costs low.

The University of Dundee, one of the UK's leading universities, has an international reputation for excellence in a range of subjects. Home to 18,000 students and 3,000 staff, it was named Scottish University of the Year in 2004/5 by the Sunday Times. In 2006, Dundee received the Queen's Anniversary Prize – the most distinguished award that can be made to a UK University – for work in drug discovery and development.

The University's College of Life Sciences is a world ranking research centre and the only five-star biological sciences grouping in Scotland. The College currently employs 670 staff from 50 countries, all working in state-of-the-art facilities on the city campus.

Capacity and performance

Researchers at the College of Life Sciences make heavy – and growing – use of advanced microscopy techniques to investigate the inner workings of cells and tissues. Images from an increasing number of advanced microscopes are stored digitally at high resolution, resulting in very large data volumes. As the College grows in stature as a research centre, more researchers and more projects require

more images to be stored – the demand for storage capacity is currently growing at an estimated 40 percent per year and this percentage is expected to rise in the future.

The lifecycle of this data is considerable too, as Dr Jason Swedlow, Wellcome Trust Senior Research Fellow and Reader at the University of Dundee, explains: 'Most projects run for between two and five years and the value of the data they generate is often only fully realised when the project is complete. In most cases, data from a completed project is then useful for at least another year, as new work makes reference to previous studies.

'Beyond that, data will be accessed less frequently – but since much scientific progress depends on the ability to build on previous discoveries, its value may remain high. We needed a way to keep growing volumes of experimental data accessible at low cost and with minimal management overheads. At the time, we were storing historical data offline and in a way that effectively rendered it inaccessible.'

The existing system relied on CD- or DVD-based backups, which used media inefficiently and were, in practical terms, impossible to restore successfully – it was too difficult to find the right information, too slow to copy it back onto the main system and too often corrupted as CDs were scratched or warped. Physical space was a major constraint: at an average utilisation rate of 90 percent, 1TB of data requires a stack of DVDs more than four feet tall.

'Another issue is image processing,' adds Dr Jonathan Monk, Director of High-Performance Computing. 'Microscope images must be put through a deconvolution algorithm to remove optical distortion, which is a computationally intensive task. We have a high-performance computing (HPC) cluster handling the deconvolution, which needs high input/output (I/O) speeds from the storage environment to work effectively. In choosing a new solution, we not only needed cost-effective, easily-restorable offline storage but also a high-performance online storage environment.'

A long-term solution

The College of Life Sciences needed a storage environment that could be delivered in economical portions yet scale to provide a true enterprise-level solution. 'We had a clear vision for the future and needed a vendor that could be flexible in helping us to achieve that vision,' says Dr Swedlow. 'Choosing IBM gave us not only a single-source vendor for both hardware and software but also access to people who understood what we were aiming to do and were willing always to go the extra mile to solve issues.'

IBM helped the College of Life Sciences to design and deploy the planned storage architecture and IBM Business Partner (BP) Tectrade supported the project, providing implementation, integration, tuning and technical support services. At the heart of the new environment is IBM TSM software, which runs on two IBM System p servers. This manages a storage area network (SAN) based on a fibre channel (FC)-connected IBM TotalStorage DS4400 with 11TB of disk, a DS4500 with 50TB of serial advanced technology attachment (SATA) disk and two IBM System Storage TS3500 Tape Library systems.

> *'The IBM DS4500 enables the HPC cluster to access microscopy images at 400MB/s; the performance is excellent.'*
>
> – *Dr. Jonathan Monk, Director of HPC*

'The DS4500 enables the HPC cluster to access microscopy images at 400MB/s; the performance is excellent,' says Dr Monk. 'Demand on microscopy resources is increasing rapidly and this performance increase means we can accomplish more research work, in greater detail and at greater speed. Tivoli tiers the storage environment, so that images which are frequently accessed by the cluster are stored on disks, while files which are less commonly used move to tape.'

The tape backup system makes offline storage simple and transparent to users – files can be found and restored rapidly when required. Compared with the CDs used previously, tapes also provide far greater reliability and longevity – keeping valuable scientific data safe for longer periods.

Says Dr Jonathan Monk, 'By using multiple TSM instances on two System p servers, we have gained not only the ability to load-balance clients but also a very capable Disaster Recovery option. The time to recovery for our TSM environment has dropped from over a day to under an hour.'

Virtualised, hierarchical storage

The College of Life Sciences now has a storage environment that makes cost-effective use of both disk and tape storage, while providing a high-performance environment for image processing in the HPC cluster.

'The advantage of the solution's hierarchical storage functionality is the ability to keep the right information stored on the right media without requiring much manual intervention,' says Dr Monk. 'As far as users are concerned, the solution is almost invisible – it simply moves files away and empowers the user to retrieve them as required. It saves a lot of time for administrators too: a two-person team can run our entire storage infrastructure.'

With storage resources virtualised into a single pool, the solution can be extended simply by adding more disk systems to the SAN – without multiplying management costs, causing disruption or requiring any restructuring of the storage architecture.

Designing the architecture to scale easily meant relatively high up-front costs – but expansion costs are low. Says Dr Swedlow, 'By pooling grants from a number of different sources, including the Wellcome Trust and using them to invest in the right foundations, we have gained a storage infrastructure that can grow with our research organisation.

'Initially, the costs per TB were high, because we were investing in the scalable fabric of the new architecture. Now that the foundations are in place, we are reaping the rewards of low-cost expansion – and the costs per TB are far lower than for our previous SAN.'

The IBM storage solution not only keeps staffing and expansion costs low – it saves on hidden costs as well. The high-speed I/O of data for the HPC cluster has dramatically accelerated microscopy image processing, so researchers are no longer waiting for the IT to catch up with them.

> *'The storage solution from IBM and Tectrade enables our researchers to make optimal use of their time and resources.'*
>
> – *Dr. Jason Swedlow, Wellcome Trust Senior Research Fellow and Reader at the University of Dundee*

'We are engaged in world-class research and have to work to very strict deadlines if we are to compete with other institutions,' explains Dr Swedlow. 'The storage solution from IBM and Tectrade enables our researchers to make optimal use of their time and resources.'

Playing a part in scientific progress

Working with IBM has helped the University to achieve its goal of cost-effective, high-performance storage that can scale to meet long-term requirements. 'Now that we have the IT infrastructure to support our research, it is easier to attract funding and house new projects – helping the University maintain its position as a world-class institution, and helping the College of Life Sciences play its part in the progress of science,' says Dr Swedlow.

He concludes, 'Science is a cumulative endeavor, where each new researcher stands on the shoulders of those who have gone before. By ensuring that data from old experiments remains accessible, our new IBM storage infrastructure helps us to accelerate the development of new ideas.'

For more information

Contact your IBM sales representative or IBM BP. Visit us at:

ibm.com/storage

IBM United Kingdom Limited
emea marketing and publishing services (emaps)
Normandy House
PO Box 32
Bunnian Place
Basingstoke
RG21 7EJ
United Kingdom

TSC03016-GBEN-00
07-STR-000027