University of Illinois at Urbana-Champaign lowers storage TCO by 15%

Institutional Need
The University of Illinois at Urbana-Champaign library is digitizing many of its collections and receiving new publications electronically. It needed the right storage platform to support archiving of unstructured data and replicating data off site.

Solution
The university deployed the Dell™ DX Object Storage Platform in conjunction with Dell Consulting Services. Digital archives and associated metadata can now grow indefinitely, while built-in replication ensures safety of the data.

Benefits
• 15% lower total cost of ownership for storage
• Anytime, anywhere access to archived data
• Ongoing scalability accommodates growth of archives without forklift upgrades
• Off-site replication of digital objects enhances business continuity
• Faster time to value than building system in-house

Application areas
• Services
• Storage Solutions

“Electronic access is our main interface now. We need to make search and retrieval of electronic digital objects as user-friendly as possible. The Dell DX Object Storage Platform was a match made in heaven.”

Robert Ferrer, Research Programmer, University of Illinois at Urbana-Champaign

Customer profile
Company: University of Illinois at Urbana-Champaign
Industry: Education
Country: United States
Students: 42,600
Faculty/Staff: 10,800
Web site: illinois.edu
The concept and definition of the library has remained relatively unchanged for thousands of years: a place in which literary, musical, artistic or reference materials are kept for use but not for sale.

However, the creation, delivery and access of library media has changed dramatically with the advent of the digital age. It is, therefore, imperative that the methods for preserving and archiving this information change as well.

Libraries around the world, especially in academia, are embracing digital curation as a means to preserve and protect their collections using minimal space while offering enhanced access to patrons. The University of Illinois at Urbana-Champaign has one of the largest public university libraries in the world, including a rare book and manuscript library. More than a million people visit the online catalog every week.

"Electronic access is our main interface now," says Robert Ferrer, research programmer. "We’re shipping a lot of our volumes to cold storage, and most journals don’t even send us paper copies anymore. We need to make search and retrieval of electronic digital objects as user-friendly as possible."

Taking control of digital chaos

Until recently, the library had no centralized archiving platform to store and protect digital objects and associated metadata. Data was often stored on expensive SAN storage and backed up to tape, or burned to DVD or CD-ROM. "We had a lot of ad hoc systems that proliferated over the past 10 to 15 years," says Thomas Habing, manager, Library Software Development Group. "We needed to get all the burgeoning digital assets that we are creating and acquiring under better control."

The team set out to develop a digital preservation archive using the open-source Fedora Commons repository software and Hydra shared code base. A major consideration in the development of the archive was the selection of a storage platform. "Although we settled on the software early in the project, when we looked at the DX Object Storage, we saw that it had a variety of features built into it that we didn’t have to implement ourselves, such as replication, self-healing and MD5 checksums to validate data integrity."

Thomas Habing, Manager, Library Software Development Group, University of Illinois at Urbana-Champaign

When we looked at the DX Object Storage, we saw that it had a variety of features built into it that we didn’t have to implement ourselves, such as replication, self-healing and MD5 checksums to validate data integrity.”

Technology at work

<table>
<thead>
<tr>
<th>Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dell™ Consulting Services</td>
</tr>
<tr>
<td>Dell Support Services – Dell ProSupport™ with Next Business Day onsite service</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hardware</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dell DX Object Storage Platform with DX6012S Storage Nodes and DX6000 Cluster Services Nodes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apache Solr full-text indexing software</td>
</tr>
<tr>
<td>Dell DX Object Storage software</td>
</tr>
<tr>
<td>Fedora Akubra Storage Plugin</td>
</tr>
<tr>
<td>Fedora Commons repository layer</td>
</tr>
<tr>
<td>Hydra Framework shared code base</td>
</tr>
<tr>
<td>Red Hat® Linux®</td>
</tr>
</tbody>
</table>
there was significant uncertainty with regard to the storage layer,” says Habing.

Cost, scalability and the long-term viability of the storage solution were major concerns. Simplicity of implementation, deployment and ongoing management were also important considerations. “We thought about building a system in-house using a traditional file system approach, but with a small team and limited resources, that may not have given us the functionality or flexibility we needed,” says Habing. “We probably saved three to six months of development time, and we still wouldn’t have had feature equivalence with the Dell solution. A homegrown solution also would have been time consuming to manage. We’re looking at the possibility of opening the repository to the campus at large, not just the library, in which case scalability would become an even greater concern.”

Another primary requirement was that all data be replicated off site for disaster recovery and business continuity purposes, with the ability to monitor the integrity of all copies. “We needed multiple, continuously validated copies in different locations to ensure the safety of the data,” says Ferrer.

A match made in heaven
Ferrer found the answer while attending a Dell roadmap meeting. “The Dell DX Object Storage Platform was being introduced, and I immediately recognized it as a potential solution to our archiving challenges,” says Ferrer. “I brought it to Tom and his group, and pretty soon we realized it was a match made in heaven.”

Designed to intelligently access, store, protect and distribute fixed digital content, the Dell DX Object Storage Platform differs from the traditional file storage model in that data is stored in addressable objects.

“We have extremely rich metadata for each digital object that we put in storage so that we can identify it,” says Habing.

“When we looked at the DX Object Storage, we saw that it had a variety of features built into it that we didn’t have to implement ourselves, such as replication, self-healing and MD5 checksums to validate data integrity. The scalability was also appealing—we can just plug in another storage node, PXE boot it, and we’re up and running.”

The library deployed four DX6012S Storage Nodes and two DX6000 Cluster Services Nodes using a single cluster architecture across disparate networks. The DX Solution is logically one cluster but contains sub-clusters for each library and automatically replicates a copy of the data objects, maintaining at least one copy in each physical location at all times. “Right now we have a sub-cluster in our main library and a sub-cluster in the engineering library,” says Habing. “Although we did the networking deployment ourselves, it was helpful to have Dell Consulting Services vet our ideas and make sure they were sound.”

The campus cluster architecture enables high availability for the digital archives. “The DX Object Storage will continue to run with just a single sub-cluster available,” says Ferrer. “Basically we’re looking at no downtime for our digital archives. Even if a Cluster Service Node or Storage Node goes down, we can still run with no downtime until we get it repaired. This is because a single sub-cluster can service a request for any file in the overall cluster, and the Storage Nodes are not dependent on the Cluster Service Nodes for regular operations. Right now the distance between sub-clusters is only a half mile apart, but soon we plan to deploy an additional sub-cluster at a site in Chicago as well, which will give us three validated copies of the data automatically replicated between all sites.”

Keep it forever, find it fast
The university formed a strategic alliance with Dell to develop a Fedora Akubra Storage Plugin for the Dell DX Object Storage Platform. Dell has completed

“Our Medusa system in combination with the Dell DX Object Storage Platform will provide for more unified workflows and give us a path for growth and scalability as the needs of the library change.”

Thomas Habing, Manager, Library Software Development Group, University of Illinois at Urbana-Champaign
Certification of the University of Illinois Akubra software connector and has jointly released it to the open source community for free use and distribution. “The Dell DX Object Storage Platform works well with the Fedora Commons repository system,” says Habing. “What we did was customize Fedora to manage objects on the DX Object Storage.”

In keeping with Hydra’s mythological theme, the team named the digital preservation system Medusa. The workflow engine calls on a series of disparate micro-services via the Advanced Message Queuing Protocol (AMQP), while moving an information package along the process from digitization to archival storage, generating metadata along the way. Each object has its rich metadata inextricably linked to it, enabling long-term preservation while ensuring data remains safe and accessible.

“We’re hoping to avoid forklift upgrades and grow the DX Object Storage over time in a modular fashion,” says Habing. “Retention requirements for libraries are a lot different than for your average business. We need to keep these objects for decades or even longer.”

**Lowering TCO for storage by 15% or more**
The centralized archive will allow the library to avoid instances of data loss that might occur with the continued proliferation of ad hoc systems. “We’ll also be able to free up islands of high-performance disk on our SAN by moving objects into the archive,” says Habing. “We predict that will be a significant cost savings for us, because we will be able to use that high-performance storage for workloads that really need it, such as virtual machines, instead of buying more.”

Policy-based management on the Dell DX Object Storage Platform provides automated, object-level retention and deletion based on metadata. Self-healing functionality continuously manages the integrity of stored objects to detect and repair errors, automatically reconfiguring and regenerating objects without the need for IT intervention. “We’ll be able to lower our total cost of ownership for storage by at least 15 percent by introducing a proper archiving tier,” says Habing. “That’s compared to cloud storage providers, which was an option we considered. Compared with traditional SAN storage, we’re saving between 25 and 50 percent.”

**Improving data protection**
Once digital objects are archived on the DX Object Storage, they will no longer need to be backed up as a part of the library’s routine data protection efforts. “We’re separating the nearline backups from the archive, so our backup system won’t have to deal with any of the archive stuff,” says Ferrer.

Backups will complete overnight instead of running 24x7. “Right now, backups are ongoing,” says Ferrer. “We’ll be able to fit them into a seven-hour window once we’ve moved all the archived data to the DX Object Storage. Replication is automated, so we won’t have to worry about it. With all this automation, we’ll be like Maytag® repairmen.”

**Faster time to value for cutting-edge technology**
The library achieved faster time to value for Medusa by using the Dell DX Object Storage Platform. “We may have never gotten to feature equivalence with what the DX Object Storage is capable of with our own custom development, and it would have taken a lot longer to build,” says Habing.

The university’s partnership with Dell is helping it stay on the leading edge of technology while managing risk appropriately. “We’re used to working with Dell to roll out state-of-the-art technologies before they are widely used,” says Ferrer.

Habing agrees: “We’re interested in pushing the envelope a little bit, but we want to have confidence that we are not going off the deep end. Dell offers us a good balance of cutting-edge technology that’s still well vetted and supported.”

End users are no longer reliant on heterogeneous and ad hoc systems for management of primary digital assets kept by the library. “Our Medusa system in combination with the Dell DX Object Storage Platform will provide more unified workflows and give us a path for growth and scalability as the needs of the library change,” says Habing.

**View all Dell case studies at dell.com/casestudies**

---

Availability and terms of Dell Services vary by region. For more information, visit dell.com/servicedescriptions © August 2012. This case study is for informational purposes only. Dell makes no warranties—express or implied—in this case study. Reference number: 1001072