

Internet Archive

One of the World's Largest Digital Libraries Deploys Open Storage and the Sun™ Modular Datacenter to Reduce Costs and Energy Use

Customer Success Story

Industry

- Technology

Business Issues

- Build a server infrastructure to support massive amounts of web data—2 PB of storage, growing by 1 PB per year
- Provide an efficient, reliable, and scalable datacenter
- Keep space, energy, management and maintenance costs low

Solution

Internet Archive chose the Sun™ Modular Datacenter, Sun Fire Open Storage Servers and the Solaris™ 10 Operating System for its massive data storage needs—2 PB and growing—and to support up to 500 queries per second. The organization is investing in a Sun solution to gain reliability and scalability by building a datacenter that can be set up quickly and that significantly reduces space, energy, management, and maintenance needs.

Business Results

- Gained a reliable and flexible datacenter that supports multiple PB of storage
- Increased storage capacity of its servers
- Reduced space and energy needs for lower costs
- Superior data integrity to guard against data loss

Products/Services/Solutions

- Sun Modular Datacenter (Sun MD)
- Sun Fire X4500 Open Storage Server
- Solaris 10 Operating System
- Sun™ Managed Services (Remote Operations Monitoring of the Sun MD)

URL Reference

sun.com/customers

Success at a glance

Internet Archive, a San Francisco, California-based organization, maintains what is probably the world's largest free digital archive. Since 1996, IA has begun collecting “snapshots” of the Web every two months. Today, it also records content from TV channels worldwide, as well as movies, music, and books. With every Web snapshot alone, the archive collects about 100 TB of data—approximately 4 billion Web pages. It has amassed about 2 PB of data in its database and receives up to 500 user queries per second.

The current rate of data acquisition is about 1 PB a year and growing. To deal with the explosive growth, Internet Archive had replaced its datacenter every three years, a costly proposition. Internet Archive either had to retrofit an existing building—upgrading the power, heating, and cooling systems and stringing miles of cable—or purchase land and build a new datacenter. And to ensure that it had sufficient server capacity for three years, Internet Archive overinvested in hardware for the early years, but often found the infrastructure outdated by the time that actually needed the maximum capacity.

So when it needed a new facility at the end of 2008, Internet Archive chose the Sun Modular Datacenter, which is completely self-contained—eliminating the need for a separate housing structure. It offers four times the density per rack of a typical datacenter. The Sun Modular Datacenter is also up to ten times faster to deploy than a traditional datacenter. And when it needs more capacity, Internet Archive can add to its current solution or even add another complete modular datacenter. In support

of the vision of the Internet Archive Project, Sun Microsystems has provided a secure site for deployment of the Sun MD, where it snapped into the existing power, cooling and connectivity infrastructure of Sun's highly efficient, modular Santa Clara datacenter.

For Internet Archive, the portable datacenter was equipped with 60 Sun Fire™ X4500 Open Storage Systems, which is a highest density integrated solution with server and storage in one 48TB, 4RU formfactor. These systems will also deliver 30 to 50% less power consumption compared with competitive configurations. The systems run the Solaris 10 Operating System, which uses the Solaris ZFS file system—a preliminary move from Internet Archive's Linux-based system. That system used server mirroring to provide backup, but with Sun's systems and Solaris ZFS™ storage pools, Internet Archive can get double the storage capacity of its old servers.

To further simplify management of the new environment, Internet Archive has elected to have Sun's Managed Services organization monitor the environmental control systems critical to the function of the Sun Modular Datacenter. This service, developed specifically for the Sun MD, provides remote monitoring of power, heating and cooling, fire/smoke detection and suppression, water detection and physical access points (door open/close). In the event of an alarm, Sun alerts Internet Archive, and then oversees the dispatch of proper repair technicians. Monitoring is preformed by engineers in the Sun Remote Operations ControlCenter in Ashburn, VA, and is enabled by Sun's ControlTower appliance. These services can be expanded to include payload management of the systems within the Sun MD should Internet Archive require it.

When the deployment is complete in March 2009, Internet Archive also expects to significantly reduce its huge power bill and maintenance costs. What's more, it now has more protection against data loss. In the past, if a disk failed, Internet Archive could lose 1 TB of data. "By using the Sun Fire X4500 and Solaris 10 with ZFS, we can have two disk failures and not lose any data," says Brewster Kahle, founder and digital librarian, Internet Archive. Solaris ZFS performs constant checks

of the data to help protect against data loss that is often undetected in normal use.

Kahle believes that Internet Archive has found the datacenter solution it was seeking. "With the Sun Modular Datacenter, we have a reliable, larger-scale system than before. And we get all sorts of efficiencies, including energy and space efficiency. For our organization, it's a one-stop shop for solving our data storage needs."

"The Sun Modular Datacenter is a very well engineered datacenter, and we didn't have to hire the engineers—we could just buy it from Sun. When we need more capacity, we can just move in another and put it in the parking lot."

Brewster Kahle

Founder and Digital Librarian, Internet Archive