NetApp Tech OnTap

Josh Thorstad
Infrastructure Architect, Enterprise Architecture-Information Technology
Be The Match

Thousands of patients with life-threatening diseases such as leukemia, lymphoma, or sickle cell disease need a marrow transplant, but don't have a match in their family. More than 70% of patients depend on Be The Match® to find a match. Operated by the nonprofit National Marrow Donor Program® (NMDP), Be The Match helps connect these patients with donors. Over the past 25 years, we have helped provide more than 50,000 transplants for patients by identifying suitable matches based on HLA type. (HLA serves as a kind of fingerprint that provides a much closer match than a simple blood type.)

A patient's doctor can contact us to search our Be The Match Registry®, which provides access to more than 15 million members and 400,000 umbilical cord blood units worldwide.. Our complex algorithms quickly identify the closest matches for patients. Within minutes, we identify potential donors. With that information in hand, we are able to start the process of contacting people and requesting additional medical screening to find the best possible donor for a patient. Ultimately, the best donor is selected and we work with them to schedule a donation at a collection center. A volunteer courier transports the collected marrow or peripheral blood stem cells (PBSCs) to the hospital where the patient is being prepared to receive the transplant.

Explore

About Be The Match

Be The Match engages a growing community of people inspired to help patients who need a marrow or umbilical cord blood transplant from an unrelated donor. The National Marrow Donor Program (NMDP), a leader in the field of marrow and cord blood transplantation, operates Be The Match, which provides opportunities for the public to become involved in saving the lives of people with leukemia, lymphoma, and other life-threatening diseases. Volunteers can join the Be The Match Registry—the world's largest and most diverse listing of potential marrow donors and donated cord blood units—as well as contribute financially to Be The Match Foundation or give their time.

More

As you can imagine, the more data we can store and transmit electronically and the faster data flows, the faster critical decisions can be made and the more patients we can help. Beginning in 2007, we set out to rethink the role of technology at Be The Match. Our goal by 2015 is to double the number of annual transplants we facilitate, while cutting the time needed to complete a transplant from 96 days to 45 days.

Like many small and fast-growing organizations, Be The Match was faced with a disparate collection of homegrown systems, multiple data sources, and isolated pools of data. Manual processes created additional complexities and inefficiencies that hampered our ability to meet our goals. In this article, I describe how we transformed our IT infrastructure starting from the ground up with storage in order to achieve the availability, scalability, security, and efficiency we need to move forward. I also discuss our private and public cloud plans.

While our IT transformation is still very much a work in progress, we're excited by the results we've achieved already:

- At 96 days, the time between donor/patient matching and the actual transplant has already been reduced by 15%.
- The number of annual transplants facilitated has grown steadily and stood at 5,500 per year in 2010 (up from 4,800 in 2009).
- The ability to more quickly integrate outcome data has accelerated the evolution of our proprietary patient/donor matching algorithm. Greater accuracy means improved survival rates.
- With the deployment of our public-facing portal for donor recruitment, annual member registrations grew by 136%, from an average of 360K/year (2000–2008) to an average of 850K newly registered members per year (2009 and 2010).

We've achieved these critical milestones in large part because of our IT efforts, which have allowed us to increase availability and improve efficiency. Using NetApp® storage has allowed us to grow our total deployed storage without adding admin head count and to reduce our total storage requirement by about 35% through thin provisioning and deduplication for significant capital and operational savings.

Shared IT Infrastructure

In 2007, when we started our transformation, our IT infrastructure consisted of about 90 servers, each with direct-attached storage and tape and no server virtualization. Finding time to do routine maintenance like firmware upgrades was a problem, backup and disaster recovery (DR) were difficult, and we had experienced some very serious outages. This created a pretty compelling need for us to look at a shared infrastructure with centralized storage that would give us better disk performance, lower admin costs, and easier backup and DR.

Today, we've standardized on NetApp unified storage. In 2007, we installed two FAS3020 clusters to allow SnapMirror® replication between our primary data center and a secondary facility. At that time, data really began to explode, and in 2008 we did a data-in-place head swap to upgrade both clusters to FAS3040 controllers. Just this year we upgraded one of those clusters again to a FAS3270 configuration to keep up with increased demand from our busy development and test (dev/test) environment. The ability to easily upgrade our central storage to get more horsepower as we've grown without having to do any complicated data migrations has been key to us.

In 2010 we added a FAS6080 cluster, which we use specifically to support our high-end databases. We've also got a FAS2040 in place to support our backup environment, so we've got a total of four NetApp arrays in production now.

For our server environment we chose IBM xSeries servers running Linux® and Windows®. We're currently about 50% virtualized overall, with some legacy applications still running on Solaris. The transition to greater virtualization is still ongoing. As part of our business transformation we are porting applications from Solaris to Linux so they can run in our VMware® environment or we are deploying packaged applications that can perform those functions. If you look at the mix of applications we run, only about 10% of them are highly strategic to our business; the other 90% should be able to be serviced by off-the-shelf solutions. A company our size—780 total employees—simply can't continue to keep expending development effort on noncore functions that can be handled by standard software.

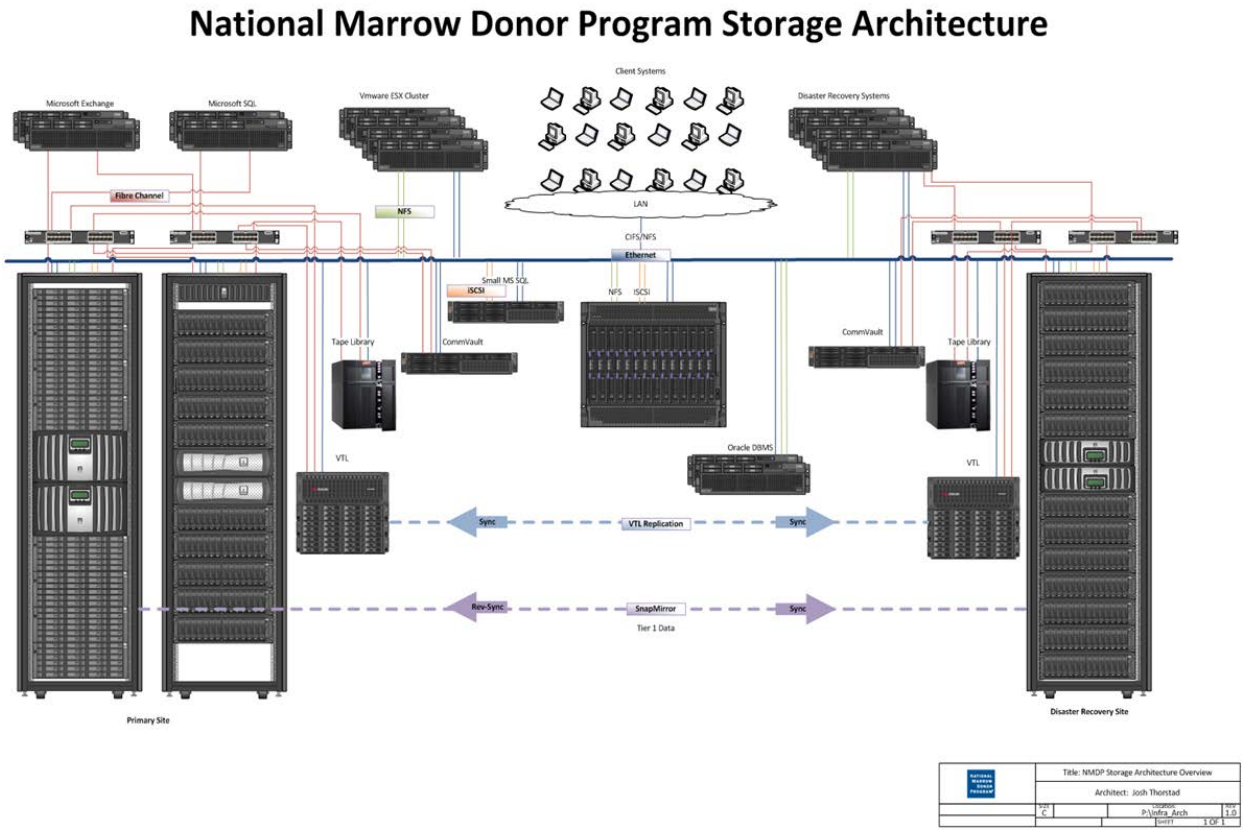


Figure 1) Be TheMatch storage architecture.

Our key applications (in-house and off-the-shelf) include:

- Traxis: the application that transplant centers use to find matches in our registry
- STAR LINK®: the database that manages our donor registry
- Oracle® E-Business Suite: recently deployed to handle financials and other business-related functions
- Microsoft® Exchange: essential to help coordinate our various locations

Leveraging IT to meet Aggressive Business Goals

The shared IT infrastructure I described in the preceding section gives us the tools we need to meet our aggressive business goals with improved performance, availability, security, data protection, and disaster recovery. Our NetApp storage has served as the foundation that has allowed us to move forward on many fronts.

Unified Storage

Because we're a small nonprofit organization, Be The Match didn't have the luxury of dedicating one storage system to Fibre Channel, another to NFS, and so on. Initially, we just had a single NetApp storage cluster in production, so we ended up running all four storage protocols—FC, NFS, CIFS, and iSCSI—on that single cluster. We immediately eliminated nine Windows file servers and moved all that CIFS data to NetApp. If you look at the applications I mentioned above, our Traxis application uses NFS, while STAR LINK is a .net application with a Microsoft SQL Server® back end, so it uses Fibre Channel. Microsoft Exchange, of course, also uses Fibre Channel while Oracle E-Business Suite runs over NFS. A few applications with smaller datasets use iSCSI. For the FAS6080 we brought in last year we decided to dedicate it for use by our bigger databases so it just uses NFS and FC, but all our other systems run all four protocols with no conflicts and no problems.

Because these storage systems are clustered, they deliver great availability (even without the DR capabilities I talk about shortly). As I already mentioned, the ability to upgrade storage controllers in place without having to migrate data has allowed us to keep up with our performance growth needs without any disruption and without having to purchase more performance capability than we need up front.

Backup and Recovery

The Be The Match Registry is the largest and most racially and ethnically diverse registry of its kind in the world. Before NetApp, backups of our STAR LINK database—which contains this data—were taking more than 24 hours to complete, so one day's backup was running into and interfering with the next day's backup. By adopting NetApp SnapManager® for Microsoft SQL Server (SMSQL), we've been able to cut the total time to perform that backup to less than 4 hours, and without disrupting database access.

In addition to SMSQL we also use SnapManager for Exchange (SME) to facilitate Exchange backups in a similar fashion. We added the Single Mailbox Recovery feature to SME to allow us to recover individual lost messages without having to go back to VTL or tape. With 18 remote offices, our business relies heavily on e-mail, and we probably use this capability on a weekly basis to recover messages that were accidentally deleted for whatever reason. We'll also take advantage of SnapManager for Oracle going forward once we educate our DBAs on its advantages for Oracle environments.

While the SnapManager tools have been invaluable to our backup strategy, there are also applications that aren't covered by those tools. For example, we have some older applications that use Sybase. While there isn't a SnapManager for Sybase, we were able to use NetApp SnapDrive® and the APIs that it provides to provide a similar level of backup functionality for this data as well. You can find more details on NetApp SnapManager tools and SnapDrive in a previous Tech OnTap® article.(NetApp recently released its Snap Creator™ Framework to further facilitate application integration.)

The snapshot schedule that we use in each case depends on the requirements of the particular dataset. Some applications are done hourly while others are only daily or weekly.

We rely on CommVault to round out our backup strategy. For data that's not being replicated or otherwise protected by NetApp Snapshot™ in some fashion, we use CommVault to back up to a virtual tape library (VTL). We have a VTL in each location with replication between them. At our DR site, we also back up our VTLs and NetApp storage to tape for archival purposes and off-site storage as needed for compliance.

Disaster Recovery

With our old direct-attached storage configuration, online disaster recovery was pretty much impossible for us, so we had to rely on recovery from off-site tape. Having all our data consolidated on centralized NetApp storage dramatically simplified the picture and allowed us to greatly improve our DR capabilities.

We use NetApp SnapMirror® software to asynchronously replicate all tier-1 data to our DR site. As with our Snapshot copy schedules, our replication schedules are tied to the needs of the dataset. Some data is replicated as often as every half hour.

Our SnapManager tools allow us to configure and manage replication for SQL Server and Exchange data. During each replication cycle, a consistent Snapshot copy is automatically made of the data on the source system, so that, after replication, the data at the DR site is in a

state in which we can restart immediately if needed.

We're just getting started with VMware Site Recovery Manager (SRM), and we'll be using that to automate the disaster recovery process for our VMware environment. SRM will help us execute the steps necessary to connect, inventory, reconfigure, and power up virtual machines at our DR site. Manual execution of these tasks can be complicated, especially when you've got dependencies that require one VM to start before another. SRM simplifies the management of the entire DR process, including discovery and configuration, failover, and DR testing.

Storage Efficiency

Another advantage that we get from NetApp is great improvements in our overall storage efficiency. In addition to the use of space-efficient Snapshot copies and replication that I described above, we also rely heavily on NetApp thin provisioning and deduplication.

Almost all storage provisioned on our NetApp systems is thin provisioned. This allows many volumes to share a single pool of free storage that saves us a tremendous amount of space. We also implemented deduplication very early on—NetApp was still approving each installation before it went live at the time. Because of the type of data we have, we use deduplication primarily for our VMware datastores and CIFS shares.

Improvements So Far

While our IT infrastructure is still very much a work in progress, we've already made significant improvements. I already described the significant business improvements in the introduction. Purely from an IT perspective, we've gone through significant evolution and growth while:

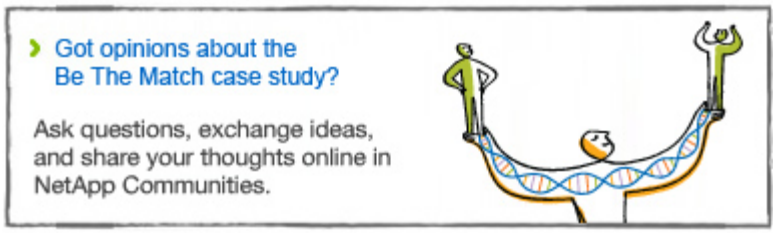
- **Increasing reliability and availability.** We've significantly reduced both unplanned and planned downtime. Our clustered NetApp systems allow us to take one storage controller down for any needed maintenance without disrupting ongoing operations. VMware VMotion™ gives us the same capability for servers.
- **Improving efficiency**. Thin provisioning and deduplication yield significant savings. We see space savings of up to 58% on deduplicated volumes. We currently have about 260TB of allocated capacity. If it weren't for thin provisioning and deduplication, we'd be using about 350TB of allocated capacity; that's about 35% more storage that we'd have to rack, power, and manage.
- **Increasing storage without increasing head count.** We've added storage and storage systems, but all our storage management needs continue to be handled by just 1.5 full-time-equivalent (FTE) staff members. This is in large part because all of our NetApp systems—whether small or large and regardless of protocol—can be managed using the same interfaces, tools, and commands. We can upgrade or add a new system and our trained staff can manage it with no surprises.
- **Simplifying management and fine-tuning performance.** We use NetApp management tools such NetApp Operations Manager (now a part of NetApp OnCommand™) and Performance Advisor to manage and understand what's going on with our storage. With these tools we can see immediately when the storage load on our existing controllers is reaching the level at which a head swap (as I described above) is needed. When a DBA reports a performance problem, we also use these tools to determine if the problem is storage related and then isolate the source.

Cloud Plans

In the coming months, Be The Match will continue to grow and evolve our infrastructure with the goal of creating a fully virtualized, shared infrastructure that uses the best available technologies to serve our patient and donor needs. From a software standpoint, we are focusing development efforts on our core needs (matching donors and patients) while leveraging the best available off-the-shelf software to satisfy other needs.

One of the next steps in our evolution will be the creation of a private cloud to allow our internal developers to request and receive fully provisioned virtual servers with no administrative intervention. We expect technology from both NetApp and VMware to be key elements of that initiative.

We are also looking at how we can evolve our NetApp based infrastructure into a public cloud. This would allow the network of affiliates in 40 countries to leverage the advances pioneered at Be The Match in support of a truly global mission to give a growing number of terminally ill patients a second chance at life.



By **Josh Thorstad, Infrastructure Architect, Enterprise Architecture-Information Technology**

Josh has worked in IT for the past 14 years, working his way up from an entry-level help-desk position through various roles in UNIX, network, and storage administration to his current position as infrastructure architect. After seven years at Be The Match, he is now responsible for all infrastructure planning, including servers, networks, telecom, storage, and middleware. Based on his experiences using NetApp storage at a previous company, Josh was instrumental in bringing NetApp on board at Be The Match.

Quick Links

- Tech OnTap Community
- Archive
- User Groups
- PDF