Technical Report

NetApp and Microsoft Hyper-V in Software Development and Testing Environments

Pavel Lobanov, NetApp
July 2011 | TR-3922

ABSTRACT

This document provides technical insight into the unique value propositions of virtual storage solutions from NetApp® for a Microsoft® Hyper-V™ Server in software build-and-test automation environments that can help reduce costs, increase storage and server use, and reduce the amount of time for software build-and-test cycles. It also provides guidance on how to configure solutions on NetApp storage for the most common scenarios and configurations.

TABLE OF CONTENTS

LIST OF FIGURES

1 INTRODUCTION

A cost-effective way for software development companies and teams to speed up their software build-development cycles, increase the quality of their software, and reduce the time required to produce and test current software builds is to reuse the same build-and-test infrastructure more efficiently. The main challenge is preparing build-and-test infrastructures for each build cycle because in most cases such infrastructures should be in a clean state. Traditionally, build-and-test infrastructure consists of physical machines with operating system (OS) drives on direct-attached storage (DAS), and preparing those machines requires a time-consuming process of reimaging build-and-test machines. In most cases, build-and-test infrastructures are available only once a day, and to increase the number of daily build-test runs, the number of dedicated physical machines in build-and-test infrastructures must be doubled or tripled.

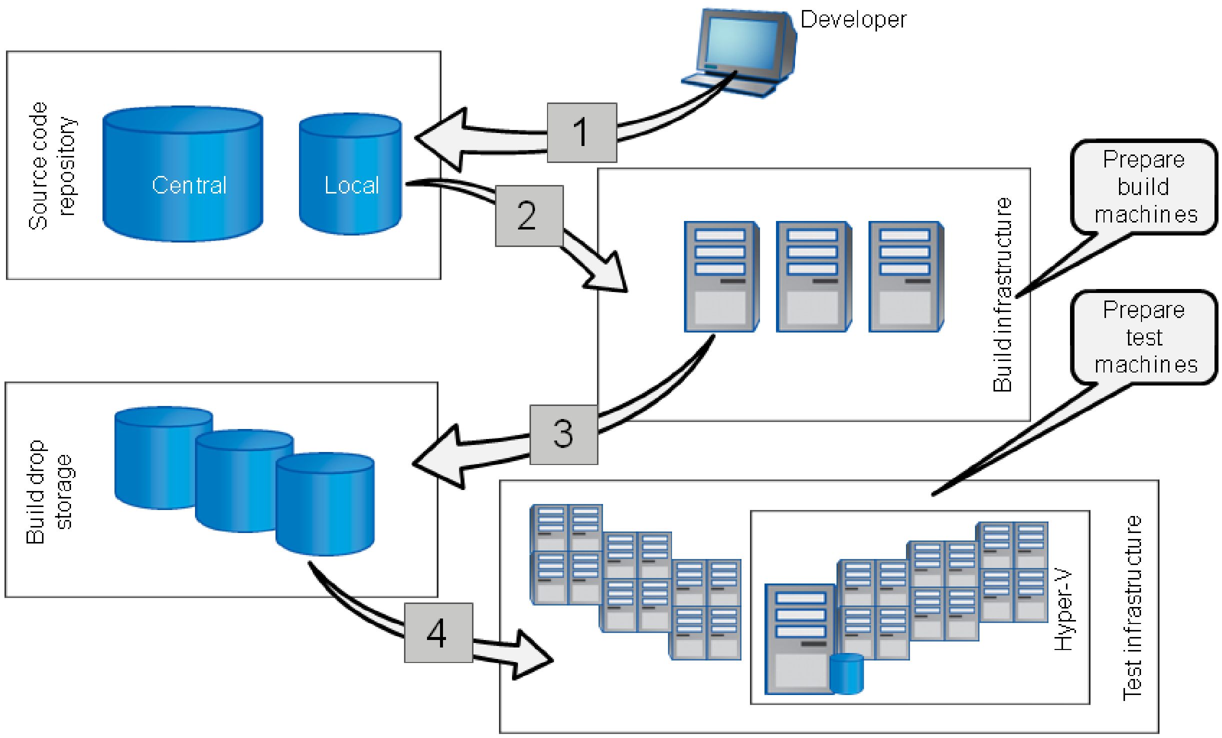
Microsoft Hyper-V Server and storage-area network (SAN) boot technologies enable businesses to use build-and-test infrastructure machines more efficiently by consolidating storage and processor power, and through faster provisioning of clean physical and virtual machines, avoiding manual intervention. Also, virtualized infrastructures require less operational expense by minimizing data center floor space, power, and cooling usage.

2 BUILD-AND-TEST PROCESS DESCRIPTION

Every build cycle is a series of steps within a well-defined build-test process:

1. Developers submit code changes to a central source code repository, such as Microsoft Team Foundation Server. Depending on each specific environment, a local copy of the source tree might be replicated to a local copy specific to each software product.
2. The build process takes place:
   a. In a typical environment, depending on build infrastructure availability, the build process is triggered manually or automatically by the software build, tracking, and management system.
   b. The build infrastructure allocates/provisions prepared and clean build machines set up for a specific type (processor architecture and language) of software. Depending on the number of flavors, the number of build machines might be significant.
   c. Build machines pull source code from the source code repository (central or local) and run the build process.
3. When the build process is complete, each build machine drops compiled code into the build folder on storage hosting builds, sometimes known as build drop shares.
4. The build is tested:
   a. After the build is fully completed and made available on the build drop share, the test automation system triggers predefined test passes (build verification tests [BVTs] or more intense software-specific tests).
   b. As part of the test run, the test automation system allocates/provisions prepared and clean test machines set up with OS defined by the test platform.
   c. As part of the test run, each test machine pulls the build from the build drop share and installs it.
5. After each test machine is fully ready with an installed build, it runs predefined test scenarios, as shown in Figure 1.

Figure 1) Software build-and-test process.



2.1 CHALLENGES

PREPARING BUILD-AND-TEST MACHINES

Most organizations and teams use a semi-automatic or automatic way to prepare machines that might include reapplying the OS image using disk ghost software or similar utilities or reinstalling the OS by using Microsoft Windows® Deployment Services (WDS) or a similar mechanism. These approaches provide a poor success rate of preparing machines and require human resources to perform manual intervention to fix failed machines during this process. Installing an OS over a network in an acceptable timeframe requires additional investments in faster and more expensive network infrastructure.

INSTALLING BUILD-AND-TEST MACHINES

During test-machine preparation, each test machine must install the current build. Taking into consideration the size of the build and the number of test machines, the amount of network traffic and the load on the server providing the build drop share might be significant. To speed up the process, dedicated file share hardware and faster and more expensive network infrastructure can be deployed.

ADDITIONAL CONSIDERATIONS

To provide code quality even before changes are submitted into the central source code repository, it is common practice to validate code changes. This precheckin validation (PCV) process requires a build-and-test process to be run against each change submission on a small scale. But even this process needs a much smaller set of build-and-test machines. Taking into consideration the number of submitted code changes, the total number of clean build-and-test machines needed to perform PCVs might be significant.

3 SAN BOOT TECHNOLOGIES

Most SAN arrays provide the ability for servers to boot from logical unit numbers (LUNs) on shared SAN arrays instead of local DAS drives over FC, iSCSI, and FCoE protocols. This creates opportunities for consolidating bootable disk storage in a single location, reducing the amount of raw disk space, reducing the size of servers, and as a result reducing the data center footprint and power consumption.

3.1 MICROSOFT HYPER-V TECHNOLOGIES

Microsoft Hyper-V technology provides cost-effective IT infrastructure through hardware consolidation, better hardware utilization, and a smaller data center footprint. Hyper-V technology enables opportunities of creating dynamic environments in which a number of virtual machines are provisioned and later reprovisioned for temporary use on a constant basis. Such dynamic environments are the base of build-and-test infrastructures.

3.2 NETAPP TECHNOLOGIES

NetApp offers a rich set of features along with consolidation and sharing of storage. Some features that are applicable for deployments of build-and-test infrastructures include the following:

- Volume and LUN FlexClone® technology
- Deduplication
- Flash Cache (formerly known as Performance Acceleration Module)
- SnapMirror® replication

3.3 FLEXCLONE TECHNOLOGY

NetApp FlexClone technology instantly replicates data volumes and datasets as transparent, virtual copies. NetApp clones increase productivity and save storage space without compromising performance.

NetApp FlexClone technology creates true clones—instantly replicated data volumes and datasets—without requiring additional storage space.

Each cloned volume is a transparent virtual copy that can be used for essential business operations:

- Testing and bug fixing
- Fast, efficient desktop and server provisioning
- Platform and upgrade sets
- Multiple simulations against large datasets
- Remote office testing/offshore testing
- Market-specific product variations

FlexClone saves space with minimal overhead to improve productivity.

For more information, refer to TR-3742: Using FlexClone to Clone Files and LUNs.

3.4 DEDUPLICATION

NetApp data deduplication and data compression are fundamental components of the core operating architecture of Data ONTAP®. These innovative data reduction features can be used broadly across multiple applications and storage tiers, including primary data, backup data, and archival data.

NetApp deduplication combines the benefits of granularity, performance, and resiliency to provide a significant advantage in the race to accommodate ever-increasing storage capacity demands.

NetApp data compression is a new feature that compresses data as it is written to NetApp fabric-attached storage (FAS) and V-Series storage systems. Like deduplication, NetApp data compression works in both SAN and network-attached storage (NAS) environments and is application and storage-tier agnostic.

Data deduplication and data compression help in the struggle to control data proliferation. The average UNIX® or Windows disk volume contains thousands or even millions of duplicate or compressible data objects. As data is created, distributed, backed up, and archived, this data is stored unabated across all storage tiers. The end result is inefficient use of data storage resources.

For more information, refer to TR-3702: NetApp Storage Best Practices for Microsoft Virtualization.

3.5 FLASH CACHE

Optimize the performance of storage systems without adding high-performance disk drives. Grow while conserving power, cooling, and space.

Use Flash Cache, formerly the Performance Acceleration Module II (PAM II), to optimize the performance of random read–intensive workloads such as file services, messaging, virtual infrastructure, and OLTP databases without using more high-performance disk drives. This intelligent read cache speeds access to data, reducing latency by a factor of 10 or more compared to disk drives. Faster response times can translate into higher throughput for random input/output (I/O) workloads.

NetApp Flash Cache enables performance that is comparable to that of solid-state disks (SSDs) without the complexity of another storage tier. There is no need to move data from tier to tier to optimize performance and cost. Active data automatically flows into Flash Cache because every volume and LUN behind the storage controller is subject to caching.

Caching priority can be given to most important volumes and LUNs when the load is heaviest by using our FlexShare® quality-of-service software in combination with Flash Cache cards. Flash Cache can be tuned to match specific workloads with software settings that allow only metadata (or new data when it is written to disk) to be cached.

NetApp Flash Cache can reduce costs for storage, power, and rack space. We have demonstrated that Flash Cache can eliminate up to 75% of the high-performance disk drives in a storage system while providing better response times across the same I/O throughput range.

Our intelligent cache can be used in combination with Serial Advanced Technology Attachment (SATA) drives for many workloads to increase storage capacity and reduce costs while maintaining a high level of performance.

For more information, refer to TR-3832: Flash Cache and PAM Best Practices Guide.

3.6 SNAPMIRROR

NetApp SnapMirror is our flagship data replication solution that not only can provide disaster recovery (DR) protection for business-critical data, but also enables a DR site for other business activities so that a DR solution can be turned into a business accelerator.

Leveraging the NetApp Unified Storage Architecture, SnapMirror simplifies the management of data replication, so it can be used as single solution across all NetApp storage arrays and protocols for any application in both virtual and traditional environments in a variety of configurations. Also, SnapMirror can be tuned to meet a full range of recovery-point objectives (RPOs), from zero seconds to hours. In addition, the recovery process supports low recovery time objectives.

Although protecting data is critical, doing it cost effectively is just as important. With the new SnapMirror network-compression technology, bandwidth utilization can be reduced by 70%, accelerating data transfers to lower RPO and network costs. Using NetApp deduplication in a virtual environment improves storage efficiency by 90%.

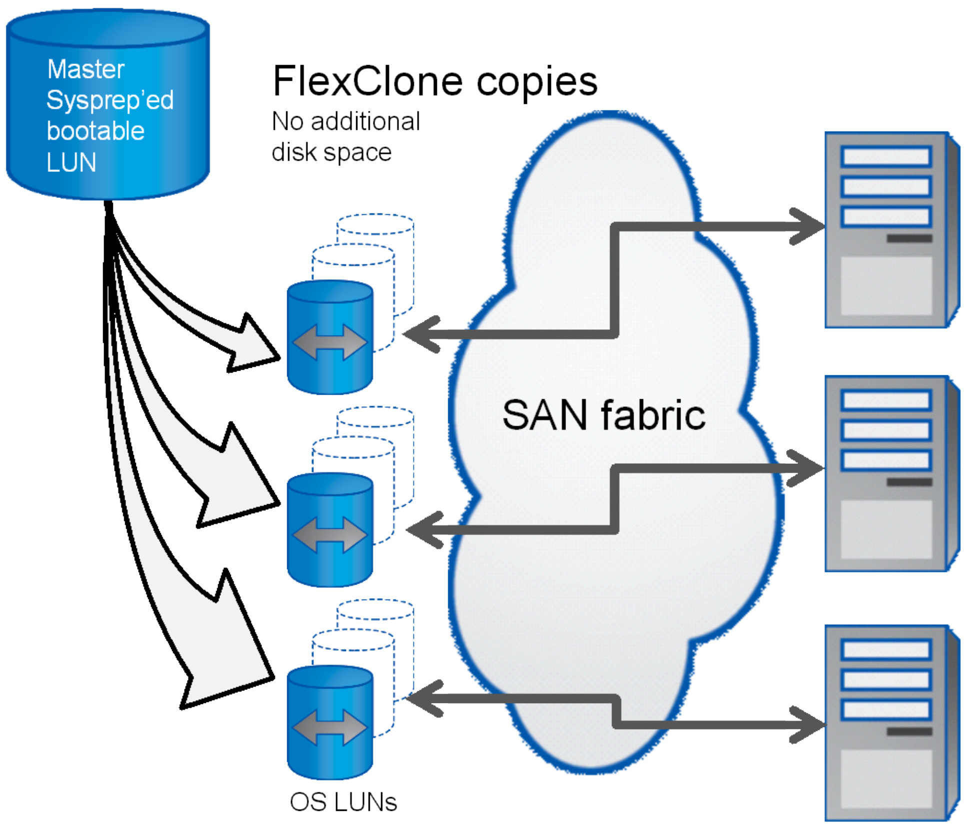
Integrated with FlexClone, which provides instantaneous, space-efficient copies, SnapMirror accelerates business by allowing the use of replicated data for DR testing, business intelligence, and development and test without business interruptions. SnapMirror is also a component of NetApp DataMotion for vFiler®, which provides always-on data migration in upgrading technology, adding more capacity, or load-balancing workload scenarios. To address specific business service-level agreements (SLAs), SnapMirror, together with NetApp MultiStore® and Provisioning Manager, can provide application transparency and, more importantly, provide it with minimal planned downtime during any of these activities.

4 APPLICATION OF NETAPP TECHNOLOGIES IN BUILD-AND-TEST ENVIRONMENTS

4.1 OS LUNS AND LUNS CONTAINING HYPER-V VHD IMAGES

The main challenge in build-and-test environments is to provision the bulk of clean physical or virtual machines in a short amount of time. A SAN array can hold a set of clean boot LUNs on which Sysprep has been run or LUNs containing boot virtual hard disk (VHD) with the OS ready for minisetup to be run. Using NetApp FlexClone and SAN boot technologies, such LUNs, or even volumes holding multiple LUNs, can be cloned instantly, and multiple LUN clones can be attached to physical or virtual machines, making clean machines available for the build or test process in a matter of minutes or even seconds. Figure 2 shows the LUN cloning process that takes place in the physical or virtual machine.

Figure 2) Instantly bootable LUN cloning process.



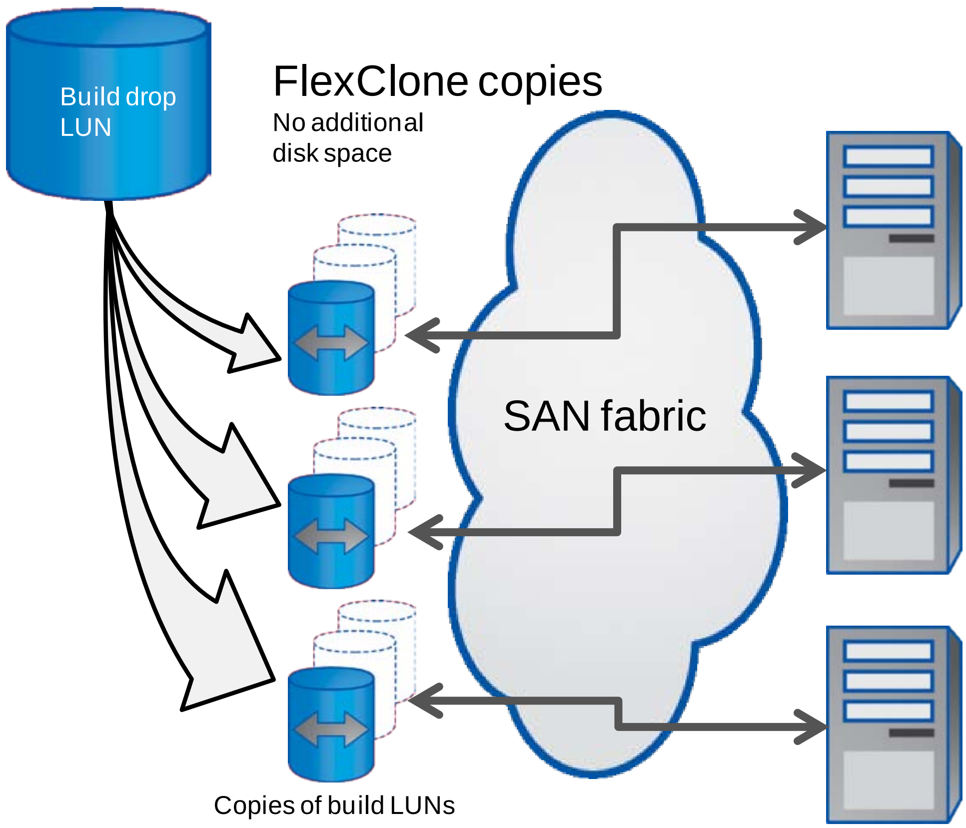
After the build-and-test machines complete the requested build or test runs, LUN clones can be destroyed instantly, and the physical hardware is ready to be reused for the next round of runs.

4.2 BUILD DROP LUN DELIVERY TO TEST MACHINES

As part of preparing the test machines, the software build to be tested must be installed. In typical environments with DAS storage, build is available on storage hosting builds, and a number of test machines (which can be high as several hundreds or even thousands) must pull build installation bits over the network. Taking into consideration the amount of network traffic and network throughput, this process might take an extended period of time, slowing down test machine provisioning and preparation, and it might require faster and more expensive network infrastructure, increasing capital expenses.

Using NetApp SAN arrays as shared storage for builds and using NetApp FlexClone technologies, LUNs containing builds to be tested can be cloned instantly and attached to multiple test virtual machines or to a Hyper-V host serving the build as a Common Internet File System (CIFS) protocol share to virtual machines hosted on each server, as shown in Figure 3.

Figure 3) Instant-build LUN cloning process.



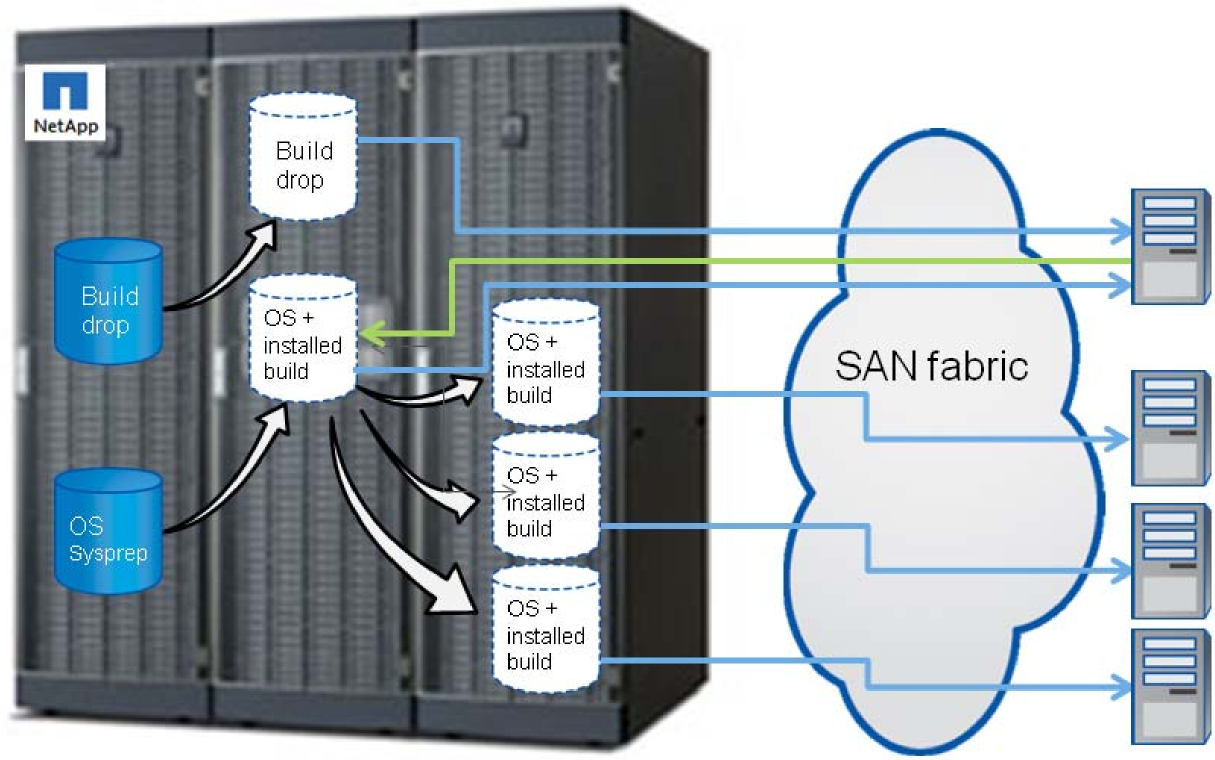
After all test machines have installed the build, the LUN clones are destroyed instantly.

4.3 ADDITIONAL IMPROVEMENTS IN TEST MACHINE PREPARATION PROCESS

Taking advantage of the instant LUN cloning and clone destruction provided by NetApp FlexClone technologies, you can implement additional improvements during test machine preparation. To minimize data transfer over SAN infrastructure during test machine installation of the build and to provide a better build-installation rate and minimize the number of test machines failing because of unsuccessful build installations, you can take a different approach.

As the first step in preparing test machines, only a small subset of requested test machines are provisioned and build installed. After that, Sysprep can be run on the OS of such test machines, and they can be detached from the original test machine. (For more information on the Microsoft Sysprep utility, see section 6, "References.") As the next step, LUNs with bootable OS with preinstalled build can be cloned again and attached to all requested test machines, effectively making all test machines not only booted from those LUNs but also fully prepared with the build installed, as shown in Figure 4.

Figure 4) Instant cloning of bootable LUN with installed build.



4.4 SAN BOOT STORM PERFORMANCE AND DISK SPACE UTILIZATION CHALLENGES

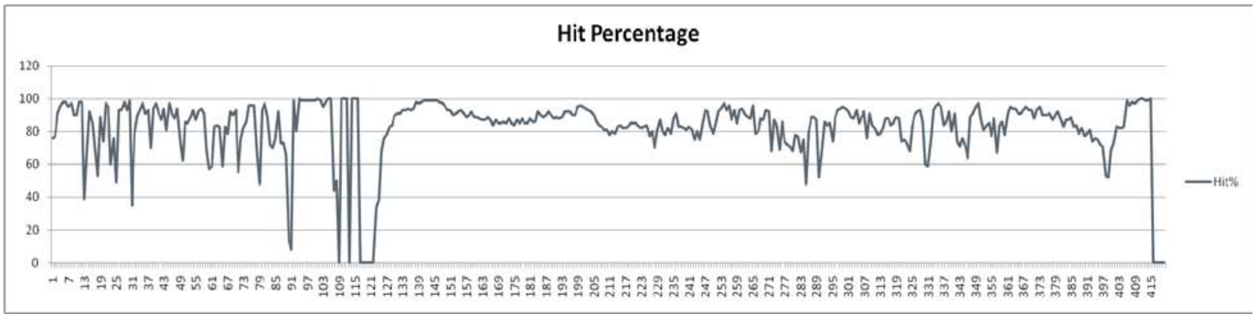
BOOT STORM PERFORMANCE

During the build-and-test machine preparation process, machines go through OS minisetup, changing OS settings such as machine name, domain joined, and so forth. Such changes, along with the initial boot of the machines, put an increased load on the SAN array. This process, known as "boot storm," requires not only intensive front-end SAN data traffic but also intensive disk I/O on the back-end of SAN arrays.

In most cases, back-end disk I/O performance can be addressed by using read-caching technologies, but traditional caching technologies might not solve the entire problem because of the limited amount of cache, as a result forcing you to use faster and more expensive disk drives such as solid-state drives (SSDs). NetApp technologies such as NetApp WAFL® (Write Anywhere File Layout) disk system and Flash Cache offer improved performance for cases of boot storm where I/O requests are served from cache instead of from physical disks on the SAN array's back end. Also, taking into consideration that data blocks of LUN clones are the same blocks on physical disks, the number of blocks to be cached is drastically reduced in comparison with traditional SAN array caching.

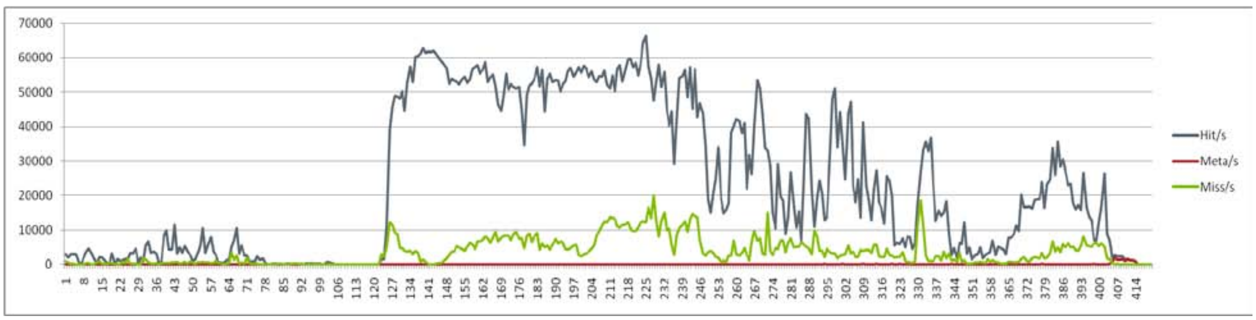
Flash Cache performance results have been captured during the blast of more than 200 VMs in one customer's software testing environment. The graph in Figure 5 shows the hit percentage rate, which is the percentage of read I/O requests served from Flash Cache versus from disks.

Figure 5) Percentage of requests served from Flash Cache instead of disks.



The graph in Figure 6 shows the per-second rate of I/O requests served from Flash Cache (hit/s), I/O requests served from disks (miss/s), and metadata-specific I/O requests.

Figure 6) Example of per-second rate of requests served from Flash Cache and disks.



These performance graphs show that the majority of I/O requests are served from Flash Cache, eliminating expensive back-end I/O on physical disks.

DISK SPACE UTILIZATION

Constant growth in the demand to store more data is driving more demand for disk space. But the amount of needed disk space is not necessarily equal to or greater than the amount of stored data. One of the most disk space–consuming areas in build-and-test environments is storage hosting builds, where multiple builds are stored during software development cycles. From the first look, each build is a unique piece of data and in some cases might occupy up to several hundred gigabytes on disks. But a closer look at and analysis of the data in each build shows that the usual change rate between consecutive builds is 5% to 10%. This means that at least 90% of data in such builds is absolutely redundant. Let's look at each build: it's a collection of flavors in different languages and platforms. And in most cases, data between all flavors, especially binary files, is redundant, which brings opportunities for additional disk space saving.

NetApp provides a free utility Storage Savings Estimation Tool (SSET) that utilizes the same algorithm used by Data ONTAP deduplication and compression to analyze data. Using this tool showed up to 97% in disk space saving on some sets of build data tested with different customers.

In addition to disk space savings provided by NetApp deduplication for build storage, FlexClone technology provides hundreds of LUN clones without additional storage requirements.

GEOGRAPHICALLY DISPERSED SOFTWARE DEVELOPMENT AND TEST ORGANIZATIONS

In today's software development industry, it's not unusual to have parts of organizations or teams in different geographical locations. Timely collaboration between them is crucial to reduce the time of each build cycle. As part of such collaboration, a significant amount of data should be shared between teams.

NetApp SnapMirror deduplication technologies offer faster replication of data between NetApp SAN arrays across the globe and enable software builds developed and built in one location to be tested in another location.

5 CONCLUSION

Dynamic and fast-provisioning virtual machines for build and test purposes introduce better hardware utilization in the Microsoft Hyper-V environment running on NetApp attached SAN arrays along with lower cost of ownership and operations.

NetApp offers great benefits to software development companies, organizations, and teams. Itenables them to “do more with less, faster” and allows them to produce better quality software at lower cost.

6 REFERENCES

- Microsoft White Paper: Boot from SAN
  www.microsoft.com/windowsserversystem/wss2003/techinfo/plandeploy/bootfromsaninwindows.mspx
- Microsoft Hyper-V
  http://en.wikipedia.org/wiki/Hyper-V
- Microsoft Sysprep utility
  http://technet.microsoft.com/en-us/library/cc783215(WS.10).aspx
- Technical Report 3701: NetApp and Microsoft Virtualization: Solution and Implementation Guide
  http://media.netapp.com/documents/tr-3701.pdf
- Technical Report 3742: Using FlexClone to Clone Files and LUNs
  www.netapp.com/us/library/technical-reports/tr-3742.html
- Technical Report 3702: NetApp Storage Best Practices for Microsoft Virtualization
  www.netapp.com/us/library/technical-reports/tr-3702.html
- Technical Report 3832: Flash Cache and PAM Best Practices Guide
  www.netapp.com/us/library/technical-reports/tr-3832.html