Silverton Consulting Lab Validation Report: Multi-Cloud Data Mobility and Performance

Executive Summary

To evaluate Cloud Data Mobility, Silverton Consulting (SCI) utilized four different methods to transfer data in and out of Amazon Web Services (AWS) and Microsoft Azure public cloud services. Two methods used standard cloud services, and two used NetApp Private Storage (NPS) for Cloud services. The data corpus employed for this lab validation was similar to that which any software development or data analytics team would use. We provided elapsed time projections for each method using configurations that could support up to 1 TB of data and gigabit networking services.

Cloud data transfer elapsed time test results, shown in Figure 1, varied significantly, from weeks for physical disk import/exports, to hours for SSH (universal web) file transfer protocol (SFTP) data transfers, to minutes for NPS iSCSI data transfers, and to mere seconds for NPS iSCSI cloud redirects. By far, the fastest way to import or export data to public cloud services or transfer data between AWS and Azure environments was to use NPS.

![Cloud Lab Data Mobility Test Results - Hours](image)

*Figure 1 Elapsed time (hours) for data import/export activities*

NPS is part of the Data Fabric approach offered by NetApp, consisting of a number of capabilities and solutions that allow customers to move data easily between on-premises data centers and multiple cloud-hosted storage service providers. The Data Fabric approach provides a quick and simple way to satisfy the many data residency requirements for hybrid cloud deployments.
Methods of moving data to and from the cloud

Hybrid cloud is a growing trend among organizations that want to take advantage of cloud services to transfer data to and from cloud service providers, but also retain some data on-premises. The main challenge with moving data to and from the cloud is size. Application data can be quite large (TBs or more), and transmitting this data can take a long time, even with high-speed gigabit networking.

There are many options for moving data from on-premises systems to the cloud and back, including:

<table>
<thead>
<tr>
<th>Import/Export Data Transfers</th>
<th>Cloud Gateways File or Block</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWS/Azure disk</td>
<td>AWS storage gateways</td>
</tr>
<tr>
<td>AWS Instance/VM</td>
<td>Cloud storage gateways</td>
</tr>
<tr>
<td>SFTP solutions</td>
<td></td>
</tr>
</tbody>
</table>

However, most of these methods are limited to moving data to and from a single cloud and do not permit data mobility between clouds. For that reason, SCI’s lab testing exercised four methods to import and export data to and from AWS and Azure public cloud services: (1) physical disk import/export, (2) SFTP get/put networking, (3) NPS copy to/from and (4) NPS cloud redirect. SFTP was used because it is universally available to any Internet-accessible data center today. Similarly, physical disk import/export is available for most data centers today. NPS was made use of to demonstrate the ease in which cloud data mobility can be accomplished in contrast to the other methods.

Test environment

SCI’s testing gathered large, public data sets downloaded from the Internet from sources like AWS, Microsoft, Microsoft Research and European Union Software Development Research. The mix of data was intended to mimic the information a software team would use as part of its development activities. The data corpus consisted of roughly 200 GB of data, with more than 8,400 files in about 300 directories.

AWS testing resided in the us-east-1a region for SFTP and the us-east-1d region for NPS direct connect (both located in Northern Virginia). SCI used an m4.4xlarge compute instance for AWS SFTP testing and NPS testing with S3 storage for disk import/export located in Oregon. We used non-encrypted, magnetic AWS Elastic Block Store (EBS) storage to hold the data corpus for SFTP and NPS import/export testing.
Azure testing resided in the Central US and used a Standard_A1 compute instance size for SFTP testing and a Standard_D14_v2 for NPS testing. Azure Blob storage was also in the Central US and was locally redundant. Azure compute data disks were magnetic.

We used NetApp’s NPS Proof of Concept (POC) cloud lab\(^1\) for the NPS tests. NPS is NetApp storage located in colocation facilities with direct networking between the collocated storage and public cloud compute infrastructure. This allows very high-speed data transfer between the NPS storage and cloud applications.

NPS Import/export activities used NetApp storage that had Direct Connect networking to AWS and NetApp storage that had ExpressRoute connections to Azure. NPS cloud redirect testing required a single NetApp Storage Virtual Machine (SVM) that was Direct Connected to AWS and ExpressRoute connected to Azure. The NPS SVM, Azure and AWS compute instances were all on the same virtual LAN (VLAN). As such, we could redirect iSCSI volume access from one cloud provider to the other by simply mounting and unmounting the volumes.

**Elapsed time analysis**

In Figure 1 above, we show the elapsed time in *hours:minutes:seconds* for the various data import/export activities undertaken in SCI’s lab testing. Note that the vertical axis is a time scale. The actual durations are shown below the horizontal axis, as some timeframes are almost imperceptible on the graph.

The duration of the **disk import** testing included preparation time to create the disk with data on it, shipment time to transport the disk to AWS or Azure import facilities and time for AWS and Azure to copy the data to AWS S3 and Azure Blob storage. Similarly, the **disk export** elapsed time included preparing the blank export disks, shipping the disks, copying the data to the disks and shipping the disks back to SCI’s lab. SCI used ground transport UPS for AWS and FedEx for Azure. Part of the AWS and Azure disk import/export data copying activities took place over holidays or weekends, so elapsed times in Figure 1 consist of only business days vs. elapsed calendar time. Although the data corpus was only ~200 GB, a TB of data would not have taken much more time to copy. Even if copying a TB had taken a few more hours, that time would be nothing compared to the six to twelve days it took to ship and process the disks.

For **SFTP data transfers**, we used SCI’s lab networking capabilities, which supported only 200Mbps download and 20Mbps upload data transfers. We project that SFTP exports (transferring data from cloud to lab) using gigabit networking services and TB of data would have taken roughly the same amount of time (five to ten hours) since there would be five times more data but five times faster network speeds. SFTP imports (transferring data from lab to

---

\(^1\) Available at https://poc.netapp.com/ as of 14Apr2016
cloud) would have been about four times faster (five times more data with twenty times faster network speeds) – roughly ten hours for AWS and eleven hours for Azure.

For **NPS POC lab import and export**, networking bandwidth (a function of the compute instance size being used in the cloud) was the limiting factor. NPS data transfers took 40 minutes for AWS and 18 minutes for Azure. As described above, NPS is NetApp ONTAP storage residing in a colocation facility with high-speed network connections to public cloud providers.

For **NPS cloud redirect testing**, data transfer speed wasn’t a limiting factor. Rather, for NPS redirect testing, how quickly the compute instances could mount and unmount iSCSI volumes were the main determinant of elapsed time. NPS redirect testing took 2 seconds for one volume and 4 seconds for 4 volumes for both transferring access from AWS to Azure and from Azure to AWS. NPS cloud redirect testing required the building of a VLAN that included the NetApp SVM storage and AWS and Azure compute infrastructure. That is, a single SVM had Direct Connect networking to AWS while at the same time was also ExpressRoute connected to Azure.

**Lessons learned**

Elapsed time lessons learned during SCI testing include the following:

- **Physical disk import and export** – Overnight shipping vs. ground transport would have reduced elapsed time. Azure was more difficult to set up due to a lack of support for Windows 10 (now rectified), and a more user-friendly process (a new management portal now supports disk import/export job creation). AWS setup and use were straightforward in comparison. AWS emailed SCI when the import/export data transfers had completed, whereas Azure only updated the import/export job status. Even with these improvements, we estimate a best case time for physical disk cloud data import or export to be approximately three to five business days for one TB of data.

- **SFTP import and export** – We used native SFTP services, which were cumbersome. Other available SFTP tools could have simplified this activity but they wouldn’t have reduced the elapsed time that much. The Linux-oriented AWS access security is based on public key encryption certificates that it creates for the user. Windows-oriented Azure uses security mostly based on an Active Directory that the user has to create. Our best case time estimate for SFTP import and export of one TB of data is between five and eleven hours, assuming gigabit networking services.

- **NPS import and export** – AWS-to-NPS and Azure-to-NPS networking bandwidth is based on the size of a compute instance purchased from the cloud provider. During early AWS-to-NPS testing with a smaller compute instance, *t2.micro*, we encountered low networking bandwidth. After we switched to a larger AWS compute instance, bandwidth...
was much faster, but costlier. Because of the bandwidth restrictions we found with AWS small compute instances, we only used a large instance with Azure for NPS activities. SCI’s Azure instance had 1Gbps bandwidth between it and the NPS POC lab. The elapsed time in Figure 1 resulted from using the larger compute instances for AWS and Azure. Using large-instance compute services, our best case time estimate for NPS import and export ranges from one to six hours for one TB of data.

- **NPS cloud redirects** – For redirect testing, bandwidth was not as important as networking configuration. A VLAN was configured, which spanned the NPS POC lab, AWS and Azure cloud compute instances so we could access the same iSCSI volume(s) from both AWS and Azure. Lab redirect testing elapsed time consisted of mounting the volume(s), doing a directory listing and then unmounting the volumes on each cloud. It is important to note that since no data was moved during the cloud redirect testing, our best case time estimate remains constant at two to four seconds, regardless of the amount of data per volume redirected between AWS and Azure.

![Cloud Lab Data Mobility Test Results - Best Case Hours](image)

*Figure 2 Comparison of SCI Best-Time Estimates*

A Note About Using Flash to Improve Performance

Although we did not include all-flash storage in our testing, we can envision a configuration in which the on-premises storage would be all-flash, NPS would use All-Flash or Flash/HDD hybrid
storage, and the cloud storage would be HDD-based. In this scenario, each location would offer a different cost/performance model, which would enable IT teams to match data performance requirements to workloads as needed, throughout the hybrid cloud environment. NetApp refers to this idea as Flash-to-Disk-to-Cloud. This concept is a departure from the traditional IT practice of keeping application data on the same storage platform throughout its entire lifecycle, which typically leads to over-provisioning of more expensive storage tiers and higher costs.

**Summary**

There are many ways today to transfer on-premises data to the public cloud. However, SCI testing results show that if your data resides on NPS, moving data to and from the public cloud is extremely fast and easy to perform. An additional bonus is that NPS also permits data transfer *between* clouds, achieving true Hybrid Cloud data mobility.

Any copying of data to and from clouds, whether through physical shipments or command line instructions, requires planning, manual intervention, and potential downtime. The cloud redirect facility offered by NPS reduces this effort from days and hours to mere seconds. For organizations embracing multi-cloud hybrid environments, we recommend considering NetApp’s Data Fabric approach and NPS.

**Disclosures**

- The study was commissioned by NetApp and delivered by SCI.
- The testing activities occurred mostly during the first quarter of 2016 in SCI lab facilities located in Colorado with cloud services residing in regions throughout the United States.

*Silverton Consulting, Inc., is a U.S.-based Storage, Strategy & Systems consulting firm offering products and services to the data storage community.*