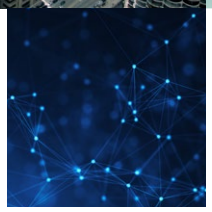

WHITE PAPER

DevOps with Azure NetApp Files and Cloud Volumes Service

NetApp

Executive Summary

One of a DevOps engineer's most important responsibilities is to spin up continuous integration (CI) and continuous delivery (CD) and developer test environments. Automating this process to include a copy of production data can increase the output quality of the software development process by an order of magnitude. However, this can be a difficult requirement to fulfil, especially for a large, agile team of software developers who regularly need to integrate their code and perform releases.

Before integration and release, software developers need test environments to catch bugs and other problems early in the development lifecycle. The ability to repeatedly test the new features they are creating with a quickly available, up-to-date, and isolated copy of production data hugely improves their speed and the quality of the code they deliver. However, with production datasets that are always rapidly growing in size and can't afford any downtime, DevOps engineers need to quickly create temporary copies of this data, which may be needed for only a few minutes to run a set of test suites and then torn down and re-created.

NetApp offers on-demand, fault-tolerant file services that can be created instantly and scaled to many terabytes in AWS and Google Cloud Platform with NetApp® Cloud Volumes Service and via a first-party, Microsoft Azure service with Azure NetApp Files. The advanced NetApp Snapshot™ and cloning capabilities provided by each service removes the complexity of provisioning data for DevOps test environments. This white paper examines the challenges that DevOps engineers face and then considers how easily these challenges can be met by using Azure NetApp Files or Cloud Volumes Service.

DevOps Challenges

The traditional approach to creating a copy of a live production dataset, such as a file share or a database, is to back up and restore the dataset. This solution, however, produces a multitude of issues and does not scale well when large numbers of environments are required, and/or the environments need to be refreshed frequently. Here are the issues with this solution.

- Data consistency. It's difficult to perform hot backups of a file share and get a consistent point-in-time copy. That's because the system is still being actively used, so the data is liable to change during the backup operation. Therefore, files copied at the start of the backup may have changed by the time the backup completes. This can lead to false positives during software testing and an unnecessary loss of developer time to debug these issues, because most enterprise systems can't tolerate the downtime required to perform a cold backup.
- Time taken to perform. Backup and restore processes typically revolve around file copy operations and scale linearly with the size of the source dataset. This means that the larger the production dataset, the longer it takes to process. Using data compression techniques can improve the time required, but not ultimately solve the problem. The delay in creating these environments is costly in terms of developer time and productivity, especially when the environments need to be refreshed repeatedly — for example, to run tests.

These problems can lead DevOps engineers and development teams to try to work around the issue of data provisioning. One solution might be to manually fabricate a test dataset. This approach allows tests to be run without needing to access the production system at all or to perform a backup and restore. The problem with this approach, however, is that creating test datasets of this kind can be difficult and time consuming, as well as requiring ongoing maintenance as the software platform evolves.

Fabricated test data that does not exhibit all of the variations found in real production use cases is of limited value. This is because inadequate test data reduces the effectiveness of software test suites and opens the way for bugs and other issues to make it into production systems. A variation of this approach could be to use a test dataset for developer environments and backup and restore for CI and CD. This approach may be viable in some scenarios; however, it limits the developer's ability to catch problems early, and is affected by all of the issues described earlier for backup and restore.

After creating test copies of production data, DevOps engineers need to manage and automate the connection of these test storage volumes to their counterpart test environment servers, which might be containerized to work in Docker or Kubernetes. Ideally, this should be achievable without introducing a lot of extra complexity.

Docker

Used by developers to write code that can be run on many different platforms. Developers and admins use this tool to enhance the DevOps pipeline by giving them a cost-effective way to work with multiple programs.

Kubernetes

A platform for container usage and microservices, and a portable cloud in itself, Kubernetes offers management capabilities for containerized workloads. It also provides network and storage infrastructure for ease of use.

What if DevOps engineers could get around all of these issues with a single, enterprise-ready solution? The ability to instantly create multiple, up-to-date, temporary working environments is a game changer for agile software development teams. As described in the next section, this is all possible with the NetApp Snapshot and cloning technologies.

Service Capabilities

Azure NetApp Files and NetApp Cloud Volumes Service offer ready-to-use, highly available, and high-performance file services in the AWS, Google Cloud Platform, and Azure environments, respectively. In true cloud form, these file services require no setup of the underlying cloud resources or ongoing server-side maintenance. This allows users to simply choose the size of the allocation they need, mount the files to the client hosts that need access, and start using their storage.

Users can change the performance characteristics of each volume they create by choosing one of three service levels: Standard, Premium, or Extreme. The service levels provide increasing levels of IOPS and I/O throughput, and they can be instantly changed on the fly after a volume has been created. Service levels provide flexibility to control the balance between performance and capacity storage.

Here are the IOPS and throughput for each service level:

- **Standard service level.** Up to 1000 IOPS per TB (16k I/O) and 16MB of throughput per TB.
- **Premium service level**. Up to 4000 IOPS per TB (16k I/O) and 64MB of throughput per TB.
- **Extreme service level.** Up to 8000 IOPS per TB (16k I/O) and 128MB of throughput per TB.

As well as enabling terabytes of file storage to be rapidly provisioned and accessed, Azure NetApp Files and Cloud Volumes Service allow highly storage-efficient Snapshot copies to be created of any volume. These copies essentially consume no additional space, because the volume and the copy share the same blocks. Only new blocks or changes consume new space and this also means that they can be created instantly, regardless of the source data size.

Snapshot copies

- Created instantly
- Storage efficient
- Protect the data

To ensure that the data in the Snapshot copy stays consistent with the point in time at which it was created, NetApp redirects updates made to blocks held within the copy over to new blocks. This redirect-on-write mechanism means that the additional storage cost of each copy is constrained by the level of update and delete activity performed on the volume after the copy was created. This makes the Snapshot copies very space efficient, and any allocated blocks used to support the copy are automatically freed up when the copy is deleted.

Another major advantage of these Snapshot copies is that they can be accessed as read-only versions of the source file system, reflecting the state of the data at the point in time at which they were created. This functionality is readily available to end users and can be accessed without performing a restore or allocating any new storage. Snapshot copies can be mounted to client hosts just like any other file share, making access to historical versions of data both fast and easy. This ability can be especially useful when test suites need to be executed against earlier versions of data files.

Although these Snapshot copies are created instantly and are space efficient, they only provide read-only access to the source data. To get a writable copy, you get NetApp cloning technology to create volume clones based on the Snapshot data. These clones are created very quickly from any size of source data. These cloning features allow Azure NetApp Files and Cloud Volumes Service to address all of the DevOps challenges related to environment setup described previously.

Clones

- Created quickly
- Entire data sets
- Easy to provision in the UI
- API creation coming soon

Because the space and time efficiencies of Snapshot copies and the time efficiencies of clones apply regardless of the size of the source data, the benefits just described apply to small volumes of a few gigabytes just as well as they do to larger, multi-terabyte volumes. In both cases, the Snapshot and cloning processes are still easy to perform. Each volume can also support many copies and clones concurrently, enabling support for as many environments as necessary and thereby guaranteeing faster TTM.

Connecting to a clone from a client host or application is exactly the same as connecting to a regular volume. DevOps engineers can connect regular and/or cloned volumes hosted in Azure NetApp Files or Cloud Volumes Service to Docker containers by using native Docker commands, and they can provision persistent NFS storage in Kubernetes environments.

Clones are currently created through the user interface. However, with the RESTful API interface, DevOps users will be able to integrate volume Snapshot and clone creation with their wider pipeline processes. Through programmatic access to the service, setting up development and test environments can be automated, which makes these processes consistently repeatable, as well as saving time and manual effort.

Conclusion

DevOps engineers who face the serious challenge of provisioning multiple, up-to-date, and easily refreshable test copies of large production datasets can use the NetApp Snapshot and cloning features built into Azure NetApp Files and Cloud Volumes Service to great effect. Creating test environments with data can greatly improve the quality of the output produced by software development teams, as well as increasing the efficiency with which they can produce results.

Many large organizations are moving to cloud-based NetApp technologies that combine enterprise-grade storage management with AWS, Azure, or Google Cloud Platform resources and services. This enables these organizations to manage their cloud storage with a much greater degree of flexibility in any cloud environment they choose.

Visit us here to find out more about how Azure NetApp Files and Cloud Volumes Service can help you with your DevOps workflow.

About NetApp

NetApp is the data authority for hybrid cloud. We provide a full range of hybrid cloud data services that simplify management of applications and data across cloud and on-premises environments to accelerate digital transformation. Together with our partners, we empower global organizations to unleash the full potential of their data to expand customer touchpoints, foster greater innovation, and optimize their operations. For more information, visit www.netapp.com. #DataDriven

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

WP-7282-1018

NetApp®