Technical Report

Data movement with E-Series and BeeGFS for AI and analytics workflows

Cody Harryman and Ryan Rodine, NetApp
June 2022 | TR-4915

Abstract

This document describes how to move data from any data repository into a BeeGFS file system backed by NetApp® E-Series® SAN storage. For artificial intelligence (AI) and machine learning (ML) applications, customers might routinely need to move large data sets exceeding many petabytes of data into their BeeGFS clusters for model development. This document explores how to accomplish this by using NetApp XCP and NetApp Cloud Sync tools.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Solution overview

For customers developing AI and ML, it's common to ingest large datasets for model building purposes. For example, to build a facial expression identification model you might want to import a massive dataset of facial expression images. Datasets can contain images, video frames, audio, and text, and can be found in a variety of sources including NFS, CIFS, Amazon Simple Storage Service (AWS S3), Azure Blob, and more.

This document describes how to move large datasets into a BeeGFS file system backed by E-Series block storage using two different NetApp tools running on a data mover node: NetApp XCP and NetApp Cloud Sync. The data mover node resides alongside the BeeGFS cluster and facilitates data movement from a generic source.

Target audience

The target audience of this paper are system administrators putting together solutions for data scientist. The paper describes how to build a data mover node to assist with ingest and egress of large datasets into a BeeGFS cluster.

Solution technology

BeeGFS file system

BeeGFS is a POSIX-based parallel file system supported by NetApp. This file system can distribute files and metadata across multiple servers and storage systems providing a storage solution with low latency and high throughput along with incredible scalability. BeeGFS also supports remote direct memory access over InfiniBand and RoCE, reducing CPU utilization on client nodes that would otherwise be incurred by high-speed networking. These capabilities make BeeGFS an ideal and cost-effective storage behind various AI, analytics, and high performance computing (HPC) workloads, which is the focus of this document. This is not the only practical use case for BeeGFS—it can also be used as a scratch workspace for reviewing and preparing unstructured data.

If you don't have BeeGFS, and you want to try it, you can easily set it up on one or more nodes. If you're deploying with NetApp E-Series, end-to-end deployment is automated using Ansible.

NetApp XCP

NetApp XCP is a client-based software for any-to-NetApp and NetApp-to-NetApp data migrations and file system insights. XCP is designed to scale and achieve greater performance by utilizing all the available system resources to handle high-volume data sets and high-performance migrations. XCP helps you to get complete visibility into the file system with the option to generate customer reports.

NetApp XCP is command-line software that's available in a single package supporting file protocols such as NFS and SMB. XCP is available as a Linux binary for NFS datasets and is available as a Windows executable for SMB datasets. Best practices guide for XCP can be found here.

NetApp Cloud Sync

Cloud Sync is a hybrid data replication Software-as-a-Service (SaaS) that transfers and synchronizes data between multiple protocols including NFS, S3, and CIFS seamlessly and securely between on-premises storage and cloud storage. This software is used for data migration, archiving, collaboration, analytics, and more. After the data is transferred, Cloud Sync continuously syncs the data between the source and destination. Going forward, it then transfers the delta. It also secures the data within your own

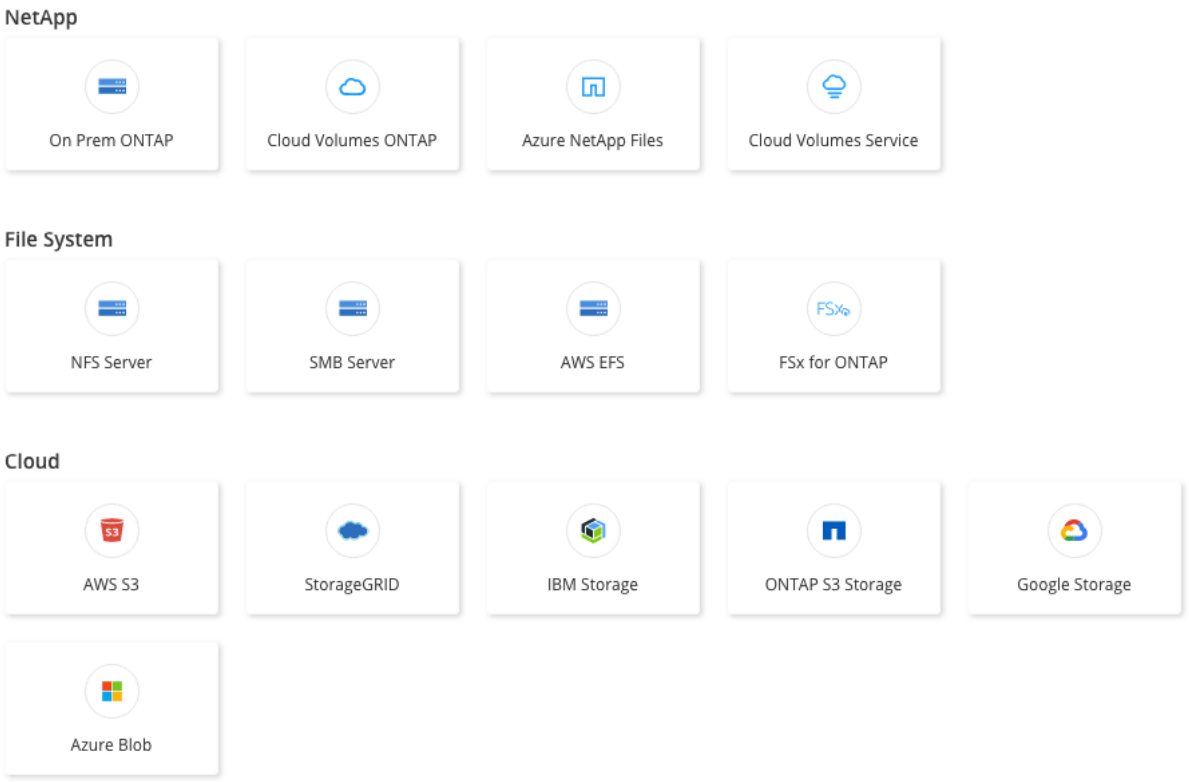
network, in the cloud, or on premises. This software is based on a pay-as-you-go model, which provides a cost-effective solution and provides monitoring and reporting capabilities for your data transfer.

Data mover solutions for BeeGFS and E-Series

Customers implementing a data mover node will be attaching to one of any number of data source types. These could be ONTAP NFS servers, AWS S3 servers, Azure Blob storage, and many more. NetApp has two solutions available depending on your needs: NetApp XCP and NetApp Cloud Sync.

Although XCP only supports NFS, SMB, and POSIX sources, Cloud Sync supports all the data sources shown in Figure 1.

Figure 1) Cloud Sync data source types.

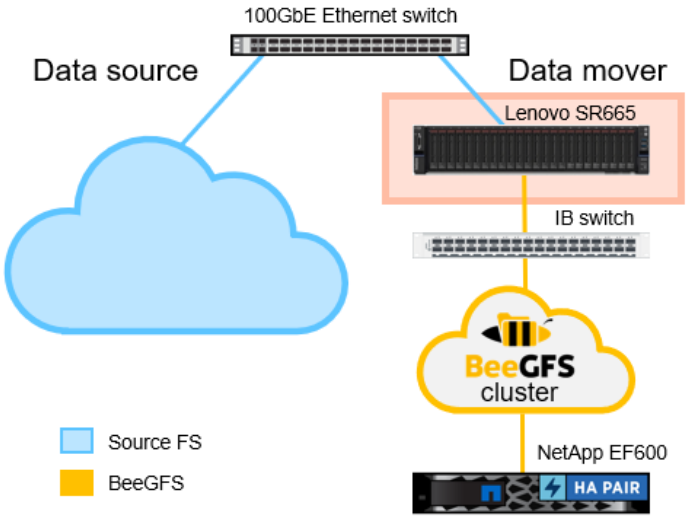


Test architecture

To test the solution and benchmark the performance of each tool, we built two test configurations. The first, consisting of one BeeGFS cluster and one XFS host connected through Ethernet, each backed with a NetApp EF600 system. The second, consisting of one BeeGFS cluster and one Lustre FS cluster connected through Ethernet, each backed with a NetApp EF600 system.

Figure 2 shows the test configuration architecture built for the solution validation.

Figure 2) Test configuration architecture.



Test setup

Sample datasets were preloaded on the data source. The source client and the data movement node were connected through Ethernet where the data movement is to take place moving the data set from source to target.

To facilitate the data movement, NFS mounts were created on both the source client and data movement node. For the XCP data movement solution, you only need an NFS mount on the source—for Cloud Sync, an NFS mount is required on both the source and target.

The data movement node itself was a Red Hat Enterprise Linux (RHEL) 8.3 server configured and connected to the target BeeGFS cluster as usual.

NFS versions

NFS export to NFS export copy is not supported with XCP on BeeGFS because it uses NFS version 3 and BeeGFS requires at least version 4. For more information, see NFS Export on the BeeGFS Documentation page. As such, we'll use the file mount point feature in XCP.

For Cloud Sync, NFS 4 should be specified during the setup wizard.

Installing XCP

For instructions on downloading, installing, and licensing XCP on your data mover node, see the NetApp XCP website.

For more information on XCP setup, tuning, and commands, see Appendix D: XCP tuning and Appendix E: XCP copy commands.

Installing Cloud Sync

For instructions on downloading, installing, and licensing Cloud Sync on your data mover node, see the NetApp Cloud Sync website. Make sure to specify NFS 4 when following the installation wizard. For testing, we used NFS exports with the data mover node selected as the onsite data broker.

For more information about the Cloud Sync setup, see the link in Appendix F: Cloud Sync setup.

Solution benchmarks

The solution was tested using two datasets of varying file size, as described in Table 1 and is compared to rsync, the standard Linux data movement tool. Additional information on the datasets used can be found in Appendix A: Data set information. Performance for both test configurations was the same.

Table 1) Solution benchmarks.

ImageNet dataset	NMDA dataset
Approx. 1,000 image files ~140MB each	Approx. 80 files ~1.1GB each
144GiB total	87.4GiB total
XCP: • Average speed: ~2.95GiBps • Transfer time: 48 seconds	XCP: • Average speed: ~2.19GiBps • Transfer time: 37 seconds
Cloud Sync: • Average speed: ~398MiBps • Transfer time: 6 minutes	Cloud Sync: • Average speed: ~181MiBps • Transfer time: 8 minutes
rsync: • Average speed: ~296MiBps • Transfer time: 6 minutes 26seconds	rsync: • Average speed: ~275 MiBps • Transfer time: 5 minutes 24 seconds

Scale out

Benchmark speeds for both NetApp XCP and NetApp Cloud Sync were achieved running only one data mover node. Higher speeds can be achieved by adding instances of XCP and Cloud Sync services on additional nodes. An example for scale out on XCP can be found here. Accelerate your Cloud Sync performance by following the guide here.

Permissions

To manage permissions for your data, create and mount directories from the BeeGFS cluster with the permissions you want and move the data into the appropriate directories.

Additionally, NetApp's BeeGFS CSI driver supports permissions as of Version 1.1. See the blog post from more details.

Conclusion

The fidelity and flexibility of any AI inferencing or ML model depends greatly on the training material used to build it. Having the ability to pull in vast amounts of training data can make the difference between a functional and an exceptional model. At the end of the day, your data movement solution needs to be fast, easy to use, and easy to integrate into your current BeeGFS configuration.

NetApp XCP and NetApp Cloud Sync are both excellent options to meet your data movement needs. Our results show that XCP is the faster solution but requires more management and setup and is limited to certain types of data sources. NetApp Cloud Sync was shown to be far more flexible, with many options for data sources, and far easier to set up.

Appendix A: Data set information

The first set tested was derived from ImageNet, an image database organized according to the WordNet hierarchy, in which each node of the hierarchy is depicted by hundreds and thousands of images. For testing, we used a preprocessed TFRecords file set based on the ImageNet 2012 dataset.

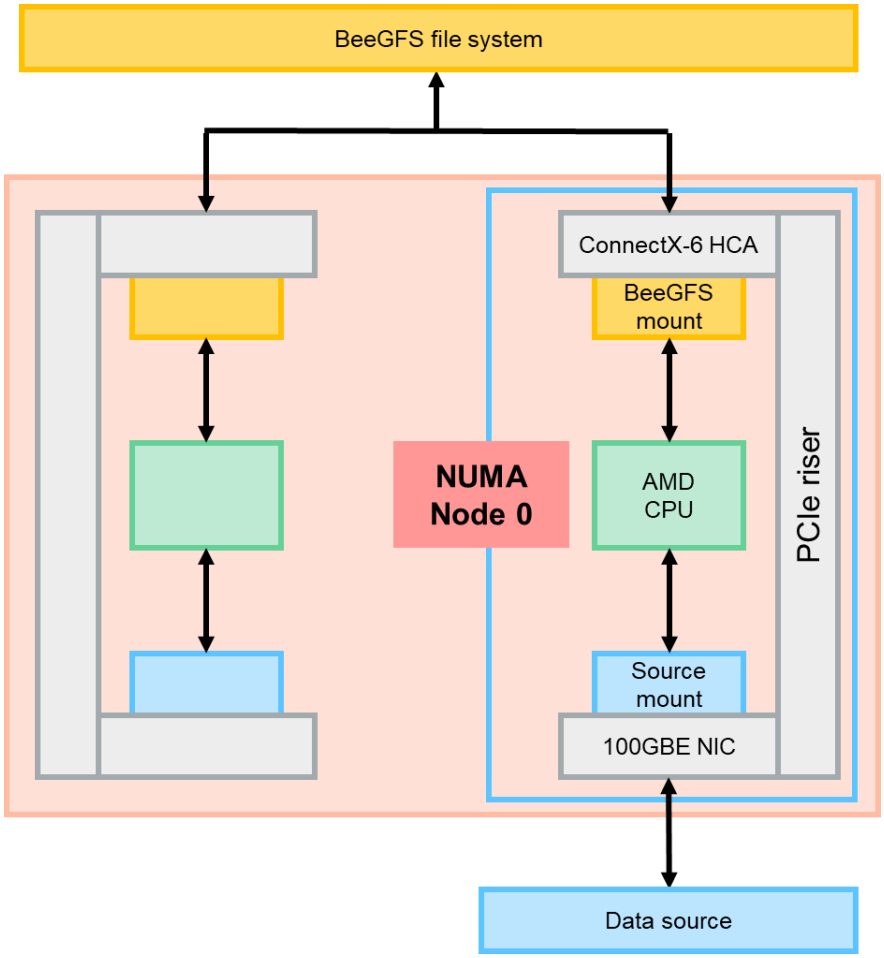
The second set tested was the Single Neurons as Deep Nets- NMDA test data set containing ~10 hours of simulation time of a cortical Layer 5 Pyramidal Neurons (L5PC) under in-vivo like input conditions (in-vivo is what biologists call things that happen "inside a living organism"). This file set had fewer, larger files than the ImageNet data set.

Appendix B: Test configuration hardware used

The data mover node was built using a Lenovo SR665 rack server with Dual AMD 7343 3.2Ghz CPUs and 256GB DDR4 RAM populated with four network adapters. Two Mellanox ConnectX-6 200Gb InfiniBand Adapters and two Mellanox ConnectX-5 100GbE Ethernet adapters.

The ConnectX-6 adapters are connected to the BeeGFS file system through a Mellanox MSB7800 IB Switch. The Ethernet adapters are connected to the data source through a Mellanox Sn2700 Ethernet Switch.

Figure 3) Hardware component arrangement.



Appendix C: NFS setup

1. Install NFS on the client server if needed.

```
sudo yum install nfs-utils
```

2. Export the NFS file system to be copied.

```
sudo vi /etc/exports
/mnt/source/imagenet 192.168.100.0/24(rw,async,fsid=0,crossmnt,no_subtree_check,no_root_squash)

sudo systemctl restart nfs-server
```

3. Mount the remote NFS export on the data mover node.

```
mount -t nfs -o vers=4,minorversion=2,noatime,nodiratime 192.168.100.25:/ /mnt/nfs_source
```

Appendix D: XCP tuning

Depending on the speed of your connection, you can tune some XCP settings. On Linux, set the following in `/etc/sysctl.conf` and run `sysctl -p`.

This is optimizing for 100Gbe:

```
net.ipv4.tcp_rmem = 4096 1342177 134217728
net.ipv4.tcp_wmem = 4096 1342177 134217728
net.core.rmem_max=268435456
net.core.rmem_default=268435456
net.core.optmem_max=268435456
net.core.wmem_max = 268435456
net.core.wmem_default = 268435456
net.core.netdev_max_backlog = 300000
net.ipv4.tcp_fin_timeout = 7
net.ipv4.tcp_slow_start_after_idle = 0
net.ipv4.tcp_window_scaling = 1
net.ipv4.tcp_sack = 1
net.ipv4.tcp_no_metrics_save = 1
net.ipv4.tcp_mtu_probing=1
```

Appendix E: XCP copy commands

Use the following XCP copy command for the image-net data set:

```
./xcp copy -parallel 24 file:///mnt/nfs_source/imagenet file:///mnt/beegfs
```

Here is the command feedback:

```
Stats : 1,153 scanned, 1,152 copied, 1,153 indexed
Speed : 144 GiB in (2.95 GiB/s), 144 GiB out (2.95 GiB/s)
Total Time : 48s.
STATUS : PASSED
```

Use the XCP copy command for the NMDA data set:

```
./xcp copy -parallel 24 file:///mnt/nfs_source/NMDA file:///mnt/beegfs
```

Here is the command feedback:

```
Stats : 103 scanned, 102 copied, 103 indexed, 81 giants
Speed : 87.4 GiB in (2.19 GiB/s), 87.4 GiB out (2.19 GiB/s)
Total Time : 39s.
STATUS : PASSED
```

The number of parallel processes prescribed in the copy command is related to the size of the files being copied. In general, larger files would warrant fewer parallel processes and smaller files would warrant more. You can experiment with the -parallel flag, which increases the number of XCP processes and can improve performance.

We used the POSIX based copy feature with XCP on the data mover node. You can also specify a POSIX directory for the XCP catalog file if you would like:

```
cat /opt/NetApp/xFiles/xcp/xcp.ini

catalog = file:///catalog_directory
```

Appendix F: Cloud Sync setup

For instructions on creating a sync relationship, see the Cloud Manager documentation.

Where to find additional information

To learn more about the information that is described in this document, review the following documents and/or websites:

- BeeGFS Documentation
  https://doc.beegfs.io/latest/
- NetApp XCP Documentation
  http://docs.netapp.com/us-en/xcp/home.html
- NetApp Cloud Sync Documentation
  https://docs.netapp.com/us-en/occm/concept_cloud_sync.html
- NetApp AI Solutions
  https://www.netapp.com/artificial-intelligence/
- NetApp Product Documentation
  https://www.netapp.com/support-and-training/documentation/

Version history

Version	Date	Document version history
Version 1.0	January 2022	Initial release.
Version 1.1	June 2022	Added reference to Lustre as a data source.

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.

Copyright information

Trademark information

TR-4915-0622

NetApp