



Big Data Analytics Data to Artificial Intelligence

NetApp Solutions

NetApp
January 10, 2023

This PDF was generated from <https://docs.netapp.com/us-en/netapp-solutions/data-analytics/bda-ai-introduction.html> on January 10, 2023. Always check docs.netapp.com for the latest.

Table of Contents

- Big Data Analytics Data to Artificial Intelligence 1
 - TR-4732: Big data analytics data to artificial intelligence..... 1
 - Customer challenges 3
 - Data mover solution 3
 - Data mover solution for AI 4
 - GPFS to NetApp ONTAP NFS 5
 - HDFS and MapR-FS to ONTAP NFS..... 8
 - Business benefits 9
 - GPFS to NFS-Detailed steps 9
 - MapR-FS to ONTAP NFS..... 33
 - Where to find additional information..... 43

Big Data Analytics Data to Artificial Intelligence

TR-4732: Big data analytics data to artificial intelligence

Karthikeyan Nagalingam, NetApp

This document describes how to move big-data analytics data and HPC data to AI. AI processes NFS data through NFS exports, whereas customers often have their AI data in a big-data analytics platform, such as HDFS, Blob, or S3 storage as well as HPC platforms such as GPFS. This paper provides guidelines for moving big-data-analytics data and HPC data to AI by using NetApp XCP and NIPAM. We also discuss the business benefits of moving data from big data and HPC to AI.

Concepts and components

Big data analytics storage

Big data analytics is the major storage provider for HDFS. A customer often uses a Hadoop-compatible file system (HCFS) such as Windows Azure Blob Storage, MapR File System (MapR-FS), and S3 object storage.

General parallel file system

IBM's GPFS is an enterprise file system that provides an alternative to HDFS. GPFS provides flexibility for applications to decide the block size and replication layout, which provide good performance and efficiency.

NetApp In-Place Analytics Module

The NetApp In-Place Analytics Module (NIPAM) serves as a driver for Hadoop clusters to access NFS data. It has four components: a connection pool, an NFS InputStream, a file handle cache, and an NFS OutputStream. For more information, see [TR-4382: NetApp In-Place Analytics Module](#).

Hadoop Distributed Copy

Hadoop Distributed Copy (DistCp) is a distributed copy tool used for large inter-cluster and intra-cluster copying tasks. This tool uses MapReduce for data distribution, error handling, and reporting. It expands the list of files and directories and inputs them to map tasks to copy the data from the source list. The image below shows the DistCp operation in HDFS and nonHDFS.



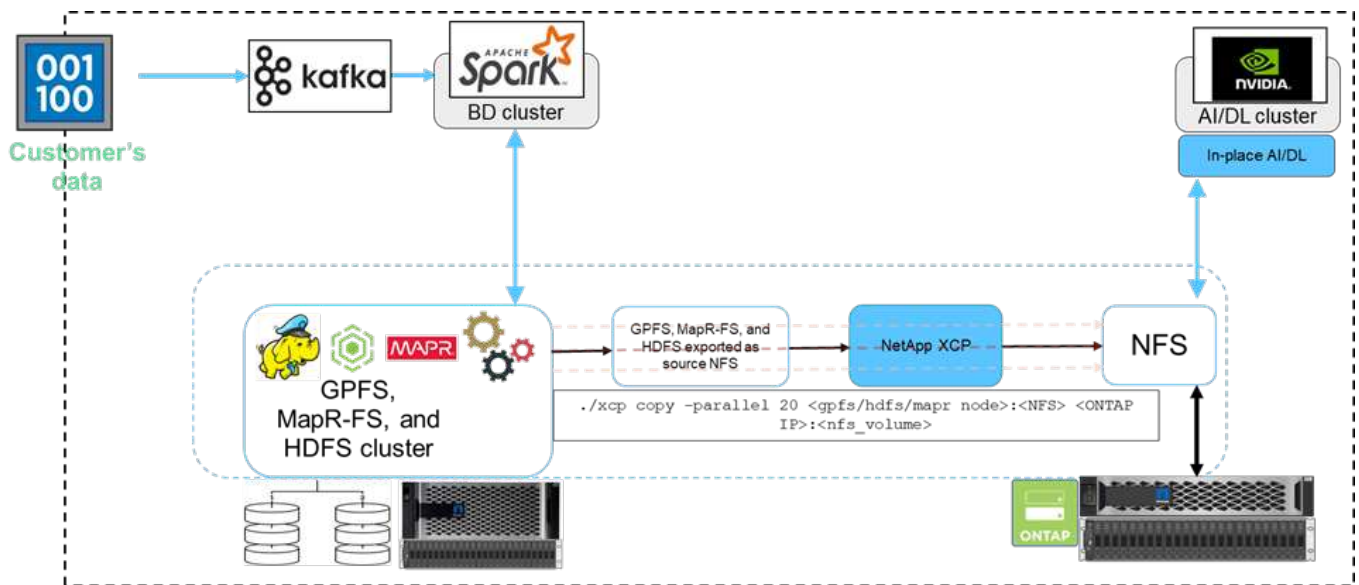
Hadoop DistCp moves data between the two HDFS systems without using an additional driver. NetApp provides the driver for non-HDFS systems. For an NFS destination, NIPAM provides the driver to copy data that Hadoop DistCp uses to communicate with NFS destinations when copying data.

NetApp Cloud Volumes Service

The NetApp Cloud Volumes Service is a cloud-native file service with extreme performance. This service helps customers accelerate their time-to-market by rapidly spinning resources up and down and using NetApp features to improve productivity and reduce staff downtime. The Cloud Volumes Service is the right alternative for disaster recovery and back up to cloud because it reduces the overall data-center footprint and consumes less native public cloud storage.

NetApp XCP

NetApp XCP is client software that enables fast and reliable any-to-NetApp and NetApp-to-NetApp data migration. This tool is designed to copy a large amount of unstructured NAS data from any NAS system to a NetApp storage controller. The XCP Migration Tool uses a multicore, multichannel I/O streaming engine that can process many requests in parallel, such as data migration, file or directory listings, and space reporting. This is the default NetApp data Migration Tool. You can use XCP to copy data from a Hadoop cluster and HPC to NetApp NFS storage. The diagram below shows data transfer from a Hadoop and HPC cluster to a NetApp NFS volume using XCP.



NetApp Cloud Sync

NetApp Cloud Sync is a hybrid data replication software-as-a-service that transfers and synchronizes NFS, S3, and CIFS data seamlessly and securely between on-premises storage and cloud storage. This software is used for data migration, archiving, collaboration, analytics, and more. After 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.

Next: [Customer challenges.](#)

Customer challenges

[Previous: Introduction.](#)

Customers might face the following challenges when trying to access data from big-data analytics for AI operations:

- Customer data is in a data lake repository. The data lake can contain different types of data such as structured, unstructured, semi-structured, logs, and machine-to-machine data. All these data types must be processed in AI systems.
- AI is not compatible with Hadoop file systems. A typical AI architecture is not able to directly access HDFS and HCFS data, which must be moved to an AI-understandable file system (NFS).
- Moving data lake data to AI typically requires specialized processes. The amount of data in the data lake can be very large. A customer must have an efficient, high-throughput, and cost-effective way to move data into AI systems.
- Syncing data. If a customer wants to sync data between the big-data platform and AI, sometimes the data processed through AI can be used with big data for analytical processing.

[Next: Data mover solution.](#)

Data mover solution

[Previous: Customer challenges.](#)

In a big-data cluster, data is stored in HDFS or HCFS, such as MapR-FS, the Windows Azure Storage Blob, S3, or the Google file system. We performed testing with HDFS, MapR-FS, and S3 as the source to copy data to NetApp ONTAP NFS export with the help of NIPAM by using the `hadoop distcp` command from the source.

The following diagram illustrates the typical data movement from a Spark cluster running with HDFS storage to a NetApp ONTAP NFS volume so that NVIDIA can process AI operations.



The `hadoop distcp` command uses the MapReduce program to copy the data. NIPAM works with MapReduce to act as a driver for the Hadoop cluster when copying data. NIPAM can distribute a load across

multiple network interfaces for a single export. This process maximizes the network throughput by distributing the data across multiple network interfaces when you copy the data from HDFS or HCFS to NFS.



NIPAM is not supported or certified with MapR.

Next: [Data mover solution for AI.](#)

Data mover solution for AI

Previous: [Data mover solution.](#)

The data mover solution for AI is based on customers' needs to process Hadoop data from AI operations. NetApp moves data from HDFS to NFS by using the NIPAM. In one use case, the customer needed to move data to NFS on the premises and another customer needed to move data from the Windows Azure Storage Blob to Cloud Volumes Service in order to process the data from the GPU cloud instances in the cloud.

The following diagram illustrates the data mover solution details.



The following steps are required to build the data mover solution:

1. ONTAP SAN provides HDFS, and NAS provides the NFS volume through NIPAM to the production data lake cluster.
2. The customer's data is in HDFS and NFS. The NFS data can be production data from other applications that is used for big data analytics and AI operations.
3. NetApp FlexClone technology creates a clone of the production NFS volume and provisions it to the AI cluster on premises.
4. Data from an HDFS SAN LUN is copied into an NFS volume with NIPAM and the `hadoop distcp` command. NIPAM uses the bandwidth of multiple network interfaces to transfer data. This process reduces the data copy time so that more data can be transferred.
5. Both NFS volumes are provisioned to the AI cluster for AI operations.
6. To process on-the-premises NFS data with GPUs in the cloud, the NFS volumes are mirrored to NetApp Private Storage (NPS) with NetApp SnapMirror technology and mounted to cloud service providers for GPUs.

7. The customer wants to process data in EC2/EMR, HDInsight, or DataProc services in GPUs from cloud service providers. The Hadoop data mover moves the data from Hadoop services to the Cloud Volumes Services with NIPAM and the `hadoop distcp` command.
8. The Cloud Volumes Service data is provisioned to AI through the NFS protocol. Data that is processed through AI can be sent on an on-premises location for big data analytics in addition to the NVIDIA cluster through NIPAM, SnapMirror, and NPS.

In this scenario, the customer has large file-count data in the NAS system at a remote location that is required for AI processing on the NetApp storage controller on premises. In this scenario, it's better to use the XCP Migration Tool to migrate the data at a faster speed.

The hybrid-use-case customer can use Cloud Sync to migrate on-premises data from NFS, CIFS, and S3 data to the cloud and vice versa for AI processing by using GPUs such as those in an NVIDIA cluster. Both Cloud Sync and the XCP Migration Tool are used for the NFS data migration to NetApp ONTAP NFS.

[Next: GPFS to NetApp ONTAP NFS.](#)

GPFS to NetApp ONTAP NFS

[Previous: Data mover solution for AI.](#)

In this validation, we used four servers as Network Shared Disk (NSD) servers to provide physical disks for GPFS. GPFS is created on top of the NSD disks to export them as NFS exports so that NFS clients can access them, as shown in the figure below. We used XCP to copy the data from GPFS-exported NFS to a NetApp NFS volume.



GPFS essentials

The following node types are used in GPFS:

- **Admin node.** Specifies an optional field containing a node name used by the administration commands to communicate between nodes. For example, the admin node `mastr-51.netapp.com` could pass a network check to all other nodes in the cluster.
- **Quorum node.** Determines whether a node is included in the pool of nodes from which quorum is derived. You need at least one node as a quorum node.
- **Manager Node.** Indicates whether a node is part of the node pool from which file system managers and token managers can be selected. It is a good idea to define more than one node as a manager node. How many nodes you designate as manager depends on the workload and the number of GPFS server licenses you have. If you are running large parallel jobs, you might need more manager nodes than in a four-node cluster supporting a web application.
- **NSD Server.** The server that prepares each physical disk for use with GPFS.
- **Protocol node.** The node that shares GPFS data directly through any Secure Shell (SSH) protocol with the NFS. This node requires a GPFS server license.

List of operations for GPFS, NFS, and XCP

This section provides the list of operations that create GPFS, export GPFS as an NFS export, and transfer the data by using XCP.

Create GPFS

To create GPFS, complete the following steps:

1. Download and install spectrum-scale data access for the Linux version on one of the servers.
2. Install the prerequisite package (chef for example) in all nodes and disable Security-Enhanced Linux (SELinux) in all nodes.
3. Set up the install node and add the admin node and the GPFS node to the cluster definition file.
4. Add the manager node, the quorum node, the NSD servers, and the GPFS node.
5. Add the GUI, admin, and GPFS nodes, and add an additional GUI server if required.
6. Add another GPFS node and check the list of all nodes.
7. Specify a cluster name, profile, remote shell binary, remote file copy binary, and port range to be set on all the GPFS nodes in the cluster definition file.
8. View the GPFS configuration settings and add an additional admin node.
9. Disable the data collection and upload the data package to the IBM Support Center.
10. Enable NTP and precheck the configurations before install.
11. Configure, create, and check the NSD disks.
12. Create the GPFS.
13. Mount the GPFS.
14. Verify and provide the required permissions to the GPFS.
15. Verify the GPFS read and write by running the `dd` command.

Export GPFS into NFS

To export the GPFS into NFS, complete the following steps:

1. Export GPFS as NFS through the `/etc/exports` file.
2. Install the required NFS server packages.
3. Start the NFS service.
4. List the files in the GPFS to validate the NFS client.

Configure NFS client

To configure the NFS client, complete the following steps:

1. Export the GPFS as NFS through the `/etc/exports` file.
2. Start the NFS client services.
3. Mount the GPFS through the NFS protocol on the NFS client.
4. Validate the list of GPFS files in the NFS mounted folder.
5. Move the data from GPFS exported NFS to NetApp NFS by using XCP.
6. Validate the GPFS files on the NFS client.

[Next: HDFS and MapR-FS to ONTAP NFS.](#)

HDFS and MapR-FS to ONTAP NFS

[Previous: GPFS to NetApp ONTAP NFS.](#)

For this solution, NetApp validated the migration of data from data lake (HDFS) and MapR cluster data to ONTAP NFS. The data resided in MapR-FS and HDFS. NetApp XCP introduced a new feature that directly migrates the data from a distributed file system such as HDFS and MapR-FS to ONTAP NFS. XCP uses async threads and HDFS C API calls to communicate and transfer data from MapR-FS as well as HDFS. The below figure shows the data migration from data lake (HDFS) and MapR-FS to ONTAP NFS. With this new feature, you don't have to export the source as an NFS share.



Why are customers moving from HDFS and MapR-FS to NFS?

Most of the Hadoop distributions such as Cloudera and Hortonworks use HDFS and MapR distributions uses their own filesystem called MapR-FS to store data. HDFS and MapR-FS data provides the valuable insights to data scientists that can be leveraged in machine learning (ML) and deep learning (DL). The data in HDFS and MapR-FS is not shared, which means it cannot be used by other applications. Customers are looking for shared data, specifically in the banking sector where customers' sensitive data is used by multiple applications. The latest version of Hadoop (3.x or later) supports NFS data source, which can be accessed without additional third-party software. With the new NetApp XCP feature, data can be moved directly from HDFS and MapR-FS to NetApp NFS in order to provide access to multiple applications

Testing was done in Amazon Web Services (AWS) to transfer the data from MapR-FS to NFS for the initial performance test with 12 MAPR nodes and 4 NFS servers.

| | Quantity | Size | vCPU | Memory | Storage | Network |
|------------|----------|---------------|------|--------|------------------|---------|
| NFS server | 4 | i3en.24xlarge | 96 | 488GiB | 8x 7500 NVMe SSD | 100 |
| MapR nodes | 12 | I3en.12xlarge | 48 | 384GiB | 4x 7500 NVMe SSD | 50 |

Based on initial testing, we obtained 20GBps throughput and were able to transfer 2PB per day of data.

For more information about HDFS data migration without exporting HDFS to NFS, see the “Deployment steps - NAS” section in [TR-4863: Best-Practice Guidelines for NetApp XCP - Data Mover, File Migration, and Analytics](#).

Next: [Business benefits](#).

Business benefits

Previous: [HDFS and MapR-FS to ONTAP NFS](#).

Moving data from big data analytics to AI provides the following benefits:

- The ability to extract data from different Hadoop file systems and GPFS into a unified NFS storage system
- A Hadoop-integrated and automated way to transfer data
- A reduction in the cost of library development for moving data from Hadoop file systems
- Maximum performance by aggregated throughput of multiple network interfaces from a single source of data by using NIPAM
- Scheduled and on-demand methods to transfer data
- Storage efficiency and enterprise management capability for unified NFS data by using ONTAP data management software
- Zero cost for data movement with the Hadoop method for data transfer

Next: [GPFS to NFS-Detailed steps](#).

GPFS to NFS-Detailed steps

Previous: [Business benefits](#).

This section provides the detailed steps needed to configure GPFS and move data into NFS by using NetApp XCP.

Configure GPFS

1. Download and Install Spectrum Scale Data Access for Linux on one of the servers.

```
[root@mastr-51 Spectrum_Scale_Data_Access-5.0.3.1-x86_64-Linux-
install_folder]# ls
Spectrum_Scale_Data_Access-5.0.3.1-x86_64-Linux-install
[root@mastr-51 Spectrum_Scale_Data_Access-5.0.3.1-x86_64-Linux-
install_folder]# chmod +x Spectrum_Scale_Data_Access-5.0.3.1-x86_64-
Linux-install
[root@mastr-51 Spectrum_Scale_Data_Access-5.0.3.1-x86_64-Linux-
install_folder]# ./Spectrum_Scale_Data_Access-5.0.3.1-x86_64-Linux-
install --manifest
manifest
...
<contents removes to save page space>
...
```

2. Install the prerequisite package (including chef and the kernel headers) on all nodes.

```
[root@mastr-51 5.0.3.1]# for i in 51 53 136 138 140 ; do ssh
10.63.150.$i "hostname; rpm -ivh /gpfs_install/chef* "; done
mastr-51.netapp.com
warning: /gpfs_install/chef-13.6.4-1.el7.x86_64.rpm: Header V4 DSA/SHA1
Signature, key ID 83ef826a: NOKEY
Preparing...
#####
package chef-13.6.4-1.el7.x86_64 is already installed
mastr-53.netapp.com
warning: /gpfs_install/chef-13.6.4-1.el7.x86_64.rpm: Header V4 DSA/SHA1
Signature, key ID 83ef826a: NOKEY
Preparing...
#####
Updating / installing...
chef-13.6.4-1.el7
#####
Thank you for installing Chef!
workr-136.netapp.com
warning: /gpfs_install/chef-13.6.4-1.el7.x86_64.rpm: Header V4 DSA/SHA1
Signature, key ID 83ef826a: NOKEY
Preparing...
#####
```

```

Updating / installing...
chef-13.6.4-1.el7
#####
Thank you for installing Chef!
workr-138.netapp.com
warning: /gpfs_install/chef-13.6.4-1.el7.x86_64.rpm: Header V4 DSA/SHA1
Signature, key ID 83ef826a: NOKEY
Preparing...
#####
Updating / installing...
chef-13.6.4-1.el7
#####
Thank you for installing Chef!
workr-140.netapp.com
warning: /gpfs_install/chef-13.6.4-1.el7.x86_64.rpm: Header V4 DSA/SHA1
Signature, key ID 83ef826a: NOKEY
Preparing...
#####
Updating / installing...
chef-13.6.4-1.el7
#####
Thank you for installing Chef!
[root@mastr-51 5.0.3.1]#
[root@mastr-51 installer]# for i in 51 53 136 138 140 ; do ssh
10.63.150.$i "hostname; yumdownloader kernel-headers-3.10.0-
862.3.2.el7.x86_64 ; rpm -Uvh --oldpackage kernel-headers-3.10.0-
862.3.2.el7.x86_64.rpm"; done
mastr-51.netapp.com
Loaded plugins: priorities, product-id, subscription-manager
Preparing...
#####
Updating / installing...
kernel-headers-3.10.0-862.3.2.el7
#####
Cleaning up / removing...
kernel-headers-3.10.0-957.21.2.el7
#####
mastr-53.netapp.com
Loaded plugins: product-id, subscription-manager
Preparing...
#####
Updating / installing...
kernel-headers-3.10.0-862.3.2.el7
#####
Cleaning up / removing...
kernel-headers-3.10.0-862.11.6.el7

```

```
#####
workr-136.netapp.com
Loaded plugins: product-id, subscription-manager
Repository ambari-2.7.3.0 is listed more than once in the configuration
Preparing...
#####
Updating / installing...
kernel-headers-3.10.0-862.3.2.el7
#####
Cleaning up / removing...
kernel-headers-3.10.0-862.11.6.el7
#####
workr-138.netapp.com
Loaded plugins: product-id, subscription-manager
Preparing...
#####
package kernel-headers-3.10.0-862.3.2.el7.x86_64 is already installed
workr-140.netapp.com
Loaded plugins: product-id, subscription-manager
Preparing...
#####
Updating / installing...
kernel-headers-3.10.0-862.3.2.el7
#####
Cleaning up / removing...
kernel-headers-3.10.0-862.11.6.el7
#####
[root@mastr-51 installer]#
```

3. Disable SELinux in all nodes.

```
[root@mastr-51 5.0.3.1]# for i in 51 53 136 138 140 ; do ssh
10.63.150.$i "hostname; sudo setenforce 0"; done
mastr-51.netapp.com
setenforce: SELinux is disabled
mastr-53.netapp.com
setenforce: SELinux is disabled
workr-136.netapp.com
setenforce: SELinux is disabled
workr-138.netapp.com
setenforce: SELinux is disabled
workr-140.netapp.com
setenforce: SELinux is disabled
[root@mastr-51 5.0.3.1]#
```

4. Set up the install node.

```
[root@mastr-51 installer]# ./spectrumscale setup -s 10.63.150.51
[ INFO ] Installing prerequisites for install node
[ INFO ] Existing Chef installation detected. Ensure the PATH is
configured so that chef-client and knife commands can be run.
[ INFO ] Your control node has been configured to use the IP
10.63.150.51 to communicate with other nodes.
[ INFO ] Port 8889 will be used for chef communication.
[ INFO ] Port 10080 will be used for package distribution.
[ INFO ] Install Toolkit setup type is set to Spectrum Scale (default).
If an ESS is in the cluster, run this command to set ESS mode:
./spectrumscale setup -s server_ip -st ess
[ INFO ] SUCCESS
[ INFO ] Tip : Designate protocol, nsd and admin nodes in your
environment to use during install:./spectrumscale -v node add <node> -p
-a -n
[root@mastr-51 installer]#
```

5. Add the admin node and the GPFS node to the cluster definition file.

```
[root@mastr-51 installer]# ./spectrumscale node add mastr-51 -a
[ INFO ] Adding node mastr-51.netapp.com as a GPFS node.
[ INFO ] Setting mastr-51.netapp.com as an admin node.
[ INFO ] Configuration updated.
[ INFO ] Tip : Designate protocol or nsd nodes in your environment to
use during install:./spectrumscale node add <node> -p -n
[root@mastr-51 installer]#
```

6. Add the manager node and the GPFS node.

```
[root@mastr-51 installer]# ./spectrumscale node add mastr-53 -m
[ INFO ] Adding node mastr-53.netapp.com as a GPFS node.
[ INFO ] Adding node mastr-53.netapp.com as a manager node.
[root@mastr-51 installer]#
```

7. Add the quorum node and the GPFS node.

```
[root@mastr-51 installer]# ./spectrumscale node add workr-136 -q
[ INFO ] Adding node workr-136.netapp.com as a GPFS node.
[ INFO ] Adding node workr-136.netapp.com as a quorum node.
[root@mastr-51 installer]#
```

8. Add the NSD servers and the GPFS node.

```
[root@mastr-51 installer]# ./spectrumscale node add workr-138 -n
[ INFO ] Adding node workr-138.netapp.com as a GPFS node.
[ INFO ] Adding node workr-138.netapp.com as an NSD server.
[ INFO ] Configuration updated.
[ INFO ] Tip :If all node designations are complete, add NSDs to your
cluster definition and define required filesystems:./spectrumscale nsd
add <device> -p <primary node> -s <secondary node> -fs <file system>
[root@mastr-51 installer]#
```

9. Add the GUI, admin, and GPFS nodes.

```
[root@mastr-51 installer]# ./spectrumscale node add workr-136 -g
[ INFO ] Setting workr-136.netapp.com as a GUI server.
[root@mastr-51 installer]# ./spectrumscale node add workr-136 -a
[ INFO ] Setting workr-136.netapp.com as an admin node.
[ INFO ] Configuration updated.
[ INFO ] Tip : Designate protocol or nsd nodes in your environment to
use during install:./spectrumscale node add <node> -p -n
[root@mastr-51 installer]#
```

10. Add another GUI server.

```
[root@mastr-51 installer]# ./spectrumscale node add mastr-53 -g
[ INFO ] Setting mastr-53.netapp.com as a GUI server.
[root@mastr-51 installer]#
```

11. Add another GPFS node.

```
[root@mastr-51 installer]# ./spectrumscale node add workr-140
[ INFO ] Adding node workr-140.netapp.com as a GPFS node.
[root@mastr-51 installer]#
```

12. Verify and list all nodes.


```

[root@mastr-51 installer]# ./spectrumscale node list
[ INFO ] List of nodes in current configuration:
[ INFO ] [Installer Node]
[ INFO ] 10.63.150.51
[ INFO ]
[ INFO ] [Cluster Details]
[ INFO ] No cluster name configured
[ INFO ] Setup Type: Spectrum Scale
[ INFO ]
[ INFO ] [Extended Features]
[ INFO ] File Audit logging      : Disabled
[ INFO ] Watch folder            : Disabled
[ INFO ] Management GUI           : Enabled
[ INFO ] Performance Monitoring  : Disabled
[ INFO ] Callhome                  : Enabled
[ INFO ]
[ INFO ] GPFS                      Admin  Quorum  Manager  NSD    Protocol
GUI   Callhome  OS    Arch
[ INFO ] Node                      Node   Node    Node    Server Node
Server Server
[ INFO ] mastr-51.netapp.com      X
rhel7  x86_64
[ INFO ] mastr-53.netapp.com                      X
X                      rhel7  x86_64
[ INFO ] workr-136.netapp.com    X      X
X                      rhel7  x86_64
[ INFO ] workr-138.netapp.com                      X
rhel7  x86_64
[ INFO ] workr-140.netapp.com
rhel7  x86_64
[ INFO ]
[ INFO ] [Export IP address]
[ INFO ] No export IP addresses configured
[root@mastr-51 installer]#

```

13. Specify a cluster name in the cluster definition file.

```

[root@mastr-51 installer]# ./spectrumscale config gpfs -c mastr-
51.netapp.com
[ INFO ] Setting GPFS cluster name to mastr-51.netapp.com
[root@mastr-51 installer]#

```

14. Specify the profile.

```
[root@mastr-51 installer]# ./spectrumscale config gpfs -p default
[ INFO ] Setting GPFS profile to default
[root@mastr-51 installer]#
Profiles options: default [gpfsProtocolDefaults], random I/O
[gpfsProtocolsRandomIO], sequential I/O [gpfsProtocolDefaults], random
I/O [gpfsProtocolRandomIO]
```

15. Specify the remote shell binary to be used by GPFS; use `-r` argument.

```
[root@mastr-51 installer]# ./spectrumscale config gpfs -r /usr/bin/ssh
[ INFO ] Setting Remote shell command to /usr/bin/ssh
[root@mastr-51 installer]#
```

16. Specify the remote file copy binary to be used by GPFS; use `-rc` argument.

```
[root@mastr-51 installer]# ./spectrumscale config gpfs -rc /usr/bin/scp
[ INFO ] Setting Remote file copy command to /usr/bin/scp
[root@mastr-51 installer]#
```

17. Specify the port range to be set on all GPFS nodes; use `-e` argument.

```
[root@mastr-51 installer]# ./spectrumscale config gpfs -e 60000-65000
[ INFO ] Setting GPFS Daemon communication port range to 60000-65000
[root@mastr-51 installer]#
```

18. View the GPFS config settings.

```
[root@mastr-51 installer]# ./spectrumscale config gpfs --list
[ INFO ] Current settings are as follows:
[ INFO ] GPFS cluster name is mastr-51.netapp.com.
[ INFO ] GPFS profile is default.
[ INFO ] Remote shell command is /usr/bin/ssh.
[ INFO ] Remote file copy command is /usr/bin/scp.
[ INFO ] GPFS Daemon communication port range is 60000-65000.
[root@mastr-51 installer]#
```

19. Add an admin node.

```
[root@mastr-51 installer]# ./spectrumscale node add 10.63.150.53 -a
[ INFO ] Setting mastr-53.netapp.com as an admin node.
[ INFO ] Configuration updated.
[ INFO ] Tip : Designate protocol or nsd nodes in your environment to
use during install:./spectrumscale node add <node> -p -n
[root@mastr-51 installer]#
```

20. Disable the data collection and upload the data package to the IBM Support Center.

```
[root@mastr-51 installer]# ./spectrumscale callhome disable
[ INFO ] Disabling the callhome.
[ INFO ] Configuration updated.
[root@mastr-51 installer]#
```

21. Enable NTP.

```
[root@mastr-51 installer]# ./spectrumscale config ntp -e on
[root@mastr-51 installer]# ./spectrumscale config ntp -l
[ INFO ] Current settings are as follows:
[ WARN ] No value for Upstream NTP Servers(comma separated IP's with NO
space between multiple IPs) in clusterdefinition file.
[root@mastr-51 installer]# ./spectrumscale config ntp -s 10.63.150.51
[ WARN ] The NTP package must already be installed and full
bidirectional access to the UDP port 123 must be allowed.
[ WARN ] If NTP is already running on any of your nodes, NTP setup will
be skipped. To stop NTP run 'service ntpd stop'.
[ WARN ] NTP is already on
[ INFO ] Setting Upstream NTP Servers(comma separated IP's with NO
space between multiple IPs) to 10.63.150.51
[root@mastr-51 installer]# ./spectrumscale config ntp -e on
[ WARN ] NTP is already on
[root@mastr-51 installer]# ./spectrumscale config ntp -l
[ INFO ] Current settings are as follows:
[ INFO ] Upstream NTP Servers(comma separated IP's with NO space
between multiple IPs) is 10.63.150.51.
[root@mastr-51 installer]#

[root@mastr-51 installer]# service ntpd start
Redirecting to /bin/systemctl start ntpd.service
[root@mastr-51 installer]# service ntpd status
Redirecting to /bin/systemctl status ntpd.service
• ntpd.service - Network Time Service
   Loaded: loaded (/usr/lib/systemd/system/ntpd.service; enabled; vendor
  preset: disabled)
```

```

Active: active (running) since Tue 2019-09-10 14:20:34 UTC; 1s ago
Process: 2964 ExecStart=/usr/sbin/ntpd -u ntp:ntp $OPTIONS
(code=exited, status=0/SUCCESS)
Main PID: 2965 (ntpd)
CGroup: /system.slice/ntpd.service
└─2965 /usr/sbin/ntpd -u ntp:ntp -g

Sep 10 14:20:34 mastr-51.netapp.com ntpd[2965]: ntp_io: estimated max
descriptors: 1024, initial socket boundary: 16
Sep 10 14:20:34 mastr-51.netapp.com ntpd[2965]: Listen and drop on 0
v4wildcard 0.0.0.0 UDP 123
Sep 10 14:20:34 mastr-51.netapp.com ntpd[2965]: Listen and drop on 1
v6wildcard :: UDP 123
Sep 10 14:20:34 mastr-51.netapp.com ntpd[2965]: Listen normally on 2 lo
127.0.0.1 UDP 123
Sep 10 14:20:34 mastr-51.netapp.com ntpd[2965]: Listen normally on 3
enp4s0f0 10.63.150.51 UDP 123
Sep 10 14:20:34 mastr-51.netapp.com ntpd[2965]: Listen normally on 4 lo
::1 UDP 123
Sep 10 14:20:34 mastr-51.netapp.com ntpd[2965]: Listen normally on 5
enp4s0f0 fe80::219:99ff:feef:99fa UDP 123
Sep 10 14:20:34 mastr-51.netapp.com ntpd[2965]: Listening on routing
socket on fd #22 for interface updates
Sep 10 14:20:34 mastr-51.netapp.com ntpd[2965]: 0.0.0.0 c016 06 restart
Sep 10 14:20:34 mastr-51.netapp.com ntpd[2965]: 0.0.0.0 c012 02 freq_set
kernel 11.890 PPM
[root@mastr-51 installer]#

```

22. Precheck the configurations before Install.

```

[root@mastr-51 installer]# ./spectrumscale install -pr
[ INFO ] Logging to file: /usr/lpp/mmfs/5.0.3.1/installer/logs/INSTALL-
PRECHECK-10-09-2019_14:51:43.log
[ INFO ] Validating configuration
[ INFO ] Performing Chef (deploy tool) checks.
[ WARN ] NTP is already running on: mastr-51.netapp.com. The install
toolkit will no longer setup NTP.
[ INFO ] Node(s): ['workr-138.netapp.com'] were defined as NSD node(s)
but the toolkit has not been told about any NSDs served by these node(s)
nor has the toolkit been told to create new NSDs on these node(s). The
install will continue and these nodes will be assigned server licenses.
If NSDs are desired, either add them to the toolkit with
<./spectrumscale nsd add> followed by a <./spectrumscale install> or add
them manually afterwards using mmcrnsd.
[ INFO ] Install toolkit will not configure file audit logging as it
has been disabled.
[ INFO ] Install toolkit will not configure watch folder as it has been
disabled.
[ INFO ] Checking for knife bootstrap configuration...
[ INFO ] Performing GPFS checks.
[ INFO ] Running environment checks
[ INFO ] Skipping license validation as no existing GPFS cluster
detected.
[ INFO ] Checking pre-requisites for portability layer.
[ INFO ] GPFS precheck OK
[ INFO ] Performing Performance Monitoring checks.
[ INFO ] Running environment checks for Performance Monitoring
[ INFO ] Performing GUI checks.
[ INFO ] Performing FILE AUDIT LOGGING checks.
[ INFO ] Running environment checks for file Audit logging
[ INFO ] Network check from admin node workr-136.netapp.com to all
other nodes in the cluster passed
[ INFO ] Network check from admin node mastr-51.netapp.com to all other
nodes in the cluster passed
[ INFO ] Network check from admin node mastr-53.netapp.com to all other
nodes in the cluster passed
[ INFO ] The install toolkit will not configure call home as it is
disabled. To enable call home, use the following CLI command:
./spectrumscale callhome enable
[ INFO ] Pre-check successful for install.
[ INFO ] Tip : ./spectrumscale install
[root@mastr-51 installer]#

```

23. Configure the NSD disks.

```
[root@mastr-51 cluster-test]# cat disk.1st
%nsd: device=/dev/sdf
nsd=nsd1
servers=workr-136
usage=dataAndMetadata
failureGroup=1

%nsd: device=/dev/sdf
nsd=nsd2
servers=workr-138
usage=dataAndMetadata
failureGroup=1
```

24. Create the NSD disks.

```
[root@mastr-51 cluster-test]# mmcrnsd -F disk.1st -v no
mmcrnsd: Processing disk sdf
mmcrnsd: Processing disk sdf
mmcrnsd: Propagating the cluster configuration data to all
    affected nodes.  This is an asynchronous process.
[root@mastr-51 cluster-test]#
```

25. Check the NSD disk status.

```
[root@mastr-51 cluster-test]# mmlsnsd
```

| File system | Disk name | NSD servers |
|-------------|-----------|----------------------|
| ----- | | |
| --- | | |
| (free disk) | nsd1 | workr-136.netapp.com |
| (free disk) | nsd2 | workr-138.netapp.com |

```
[root@mastr-51 cluster-test]#
```

26. Create the GPFS.

```
[root@mastr-51 cluster-test]# mmcrfs gpfs1 -F disk.1st -B 1M -T /gpfs1

The following disks of gpfs1 will be formatted on node workr-
136.netapp.com:
    nsd1: size 3814912 MB
    nsd2: size 3814912 MB
Formatting file system ...
Disks up to size 33.12 TB can be added to storage pool system.
Creating Inode File
Creating Allocation Maps
Creating Log Files
Clearing Inode Allocation Map
Clearing Block Allocation Map
Formatting Allocation Map for storage pool system
Completed creation of file system /dev/gpfs1.
mmcrfs: Propagating the cluster configuration data to all
    affected nodes.  This is an asynchronous process.
[root@mastr-51 cluster-test]#
```

27. Mount the GPFS.

```
[root@mastr-51 cluster-test]# mmmount all -a
Tue Oct  8 18:05:34 UTC 2019: mmmount: Mounting file systems ...
[root@mastr-51 cluster-test]#
```

28. Check and provide the required permissions to the GPFS.

```
[root@mastr-51 cluster-test]# mmlsdisk gpfs1
disk          driver  sector      failure holds    holds
storage
name          type    size        group metadata data  status
availability pool
-----
nsd1          nsd      512          1 Yes        Yes  ready      up
system
nsd2          nsd      512          1 Yes        Yes  ready      up
system
[root@mastr-51 cluster-test]#

[root@mastr-51 cluster-test]# for i in 51 53 136 138 ; do ssh
10.63.150.$i "hostname; chmod 777 /gpfs1" ; done;
mastr-51.netapp.com
mastr-53.netapp.com
workr-136.netapp.com
workr-138.netapp.com
[root@mastr-51 cluster-test]#
```

29. Check the GPFS read and write by running the dd command.

```
[root@mastr-51 cluster-test]# dd if=/dev/zero of=/gpfs1/testfile
bs=1024M count=5
5+0 records in
5+0 records out
5368709120 bytes (5.4 GB) copied, 8.3981 s, 639 MB/s
[root@mastr-51 cluster-test]# for i in 51 53 136 138 ; do ssh
10.63.150.$i "hostname; ls -ltrh /gpfs1" ; done;
mastr-51.netapp.com
total 5.0G
-rw-r--r-- 1 root root 5.0G Oct  8 18:10 testfile
mastr-53.netapp.com
total 5.0G
-rw-r--r-- 1 root root 5.0G Oct  8 18:10 testfile
workr-136.netapp.com
total 5.0G
-rw-r--r-- 1 root root 5.0G Oct  8 18:10 testfile
workr-138.netapp.com
total 5.0G
-rw-r--r-- 1 root root 5.0G Oct  8 18:10 testfile
[root@mastr-51 cluster-test]#
```


Export GPFS into NFS

To export GPFS into NFS, complete the following steps:

1. Export the GPFS as NFS through the `/etc/exports` file.

```
[root@mastr-51 gpfs1]# cat /etc/exports
/gpfs1      *(rw,fsid=745)
[root@mastr-51 gpfs1]
```

2. Install the required NFS server packages.

```
[root@mastr-51 ~]# yum install rpcbind
Loaded plugins: priorities, product-id, search-disabled-repos,
subscription-manager
Resolving Dependencies
--> Running transaction check
---> Package rpcbind.x86_64 0:0.2.0-47.el7 will be updated
---> Package rpcbind.x86_64 0:0.2.0-48.el7 will be an update
--> Finished Dependency Resolution

Dependencies Resolved

=====
=====
=====
=====
Package                                     Arch
Version                                     Repository
Size
=====
=====
=====
=====
Updating:
  rpcbind                                     x86_64
0.2.0-48.el7                                rhel-7-
server-rpms                                60 k

Transaction Summary
=====
=====
=====
=====
Upgrade  1 Package
```

```
Total download size: 60 k
Is this ok [y/d/N]: y
Downloading packages:
No Presto metadata available for rhel-7-server-rpms
rpcbind-0.2.0-48.el7.x86_64.rpm
| 60 kB 00:00:00
Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
  Updating      : rpcbind-0.2.0-48.el7.x86_64
1/2
  Cleanup      : rpcbind-0.2.0-47.el7.x86_64
2/2
  Verifying    : rpcbind-0.2.0-48.el7.x86_64
1/2
  Verifying    : rpcbind-0.2.0-47.el7.x86_64
2/2

Updated:
  rpcbind.x86_64 0:0.2.0-48.el7

Complete!
[root@mastr-51 ~]#
```

3. Start the NFS service.

```

[root@mastr-51 ~]# service nfs status
Redirecting to /bin/systemctl status nfs.service
• nfs-server.service - NFS server and services
   Loaded: loaded (/usr/lib/systemd/system/nfs-server.service; disabled;
vendor preset: disabled)
   Drop-In: /run/systemd/generator/nfs-server.service.d
            └─order-with-mounts.conf
   Active: inactive (dead)
[root@mastr-51 ~]# service rpcbind start
Redirecting to /bin/systemctl start rpcbind.service
[root@mastr-51 ~]# service nfs start
Redirecting to /bin/systemctl start nfs.service
[root@mastr-51 ~]# service nfs status
Redirecting to /bin/systemctl status nfs.service
• nfs-server.service - NFS server and services
   Loaded: loaded (/usr/lib/systemd/system/nfs-server.service; disabled;
vendor preset: disabled)
   Drop-In: /run/systemd/generator/nfs-server.service.d
            └─order-with-mounts.conf
   Active: active (exited) since Wed 2019-11-06 16:34:50 UTC; 2s ago
   Process: 24402 ExecStartPost=/bin/sh -c if systemctl -q is-active
gssproxy; then systemctl reload gssproxy ; fi (code=exited,
status=0/SUCCESS)
   Process: 24383 ExecStart=/usr/sbin/rpc.nfsd $RPCNFSDARGS (code=exited,
status=0/SUCCESS)
   Process: 24379 ExecStartPre=/usr/sbin/exportfs -r (code=exited,
status=0/SUCCESS)
   Main PID: 24383 (code=exited, status=0/SUCCESS)
   CGroup: /system.slice/nfs-server.service

Nov 06 16:34:50 mastr-51.netapp.com systemd[1]: Starting NFS server and
services...
Nov 06 16:34:50 mastr-51.netapp.com systemd[1]: Started NFS server and
services.
[root@mastr-51 ~]#

```

4. List the files in GPFS to validate the NFS client.

```

[root@mastr-51 gpfs1]# df -Th
Filesystem                                Type      Size  Used Avail
Use% Mounted on
/dev/mapper/rhel_stlrx300s6--22--irmc-root xfs        94G   55G   39G
59% /
devtmpfs                                  devtmpfs   32G     0   32G
0% /dev
tmpfs                                     tmpfs      32G     0   32G
0% /dev/shm
tmpfs                                     tmpfs      32G   3.3G   29G
11% /run
tmpfs                                     tmpfs      32G     0   32G
0% /sys/fs/cgroup
/dev/sda7                                 xfs        9.4G   210M   9.1G
3% /boot
tmpfs                                     tmpfs      6.3G     0   6.3G
0% /run/user/10065
tmpfs                                     tmpfs      6.3G     0   6.3G
0% /run/user/10068
tmpfs                                     tmpfs      6.3G     0   6.3G
0% /run/user/10069
10.63.150.213:/nc_volume3                 nfs4      380G   8.0M  380G
1% /mnt
tmpfs                                     tmpfs      6.3G     0   6.3G
0% /run/user/0
gpfs1                                     gpfs       7.3T   9.1G   7.3T
1% /gpfs1
[root@mastr-51 gpfs1]#
[root@mastr-51 ~]# cd /gpfs1
[root@mastr-51 gpfs1]# ls
catalog ces gpfs-ces ha testfile
[root@mastr-51 gpfs1]#
[root@mastr-51 ~]# cd /gpfs1
[root@mastr-51 gpfs1]# ls
ces gpfs-ces ha testfile
[root@mastr-51 gpfs1]# ls -ltrha
total 5.1G
dr-xr-xr-x  2 root root 8.0K Jan  1 1970 .snapshots
-rw-r--r--  1 root root 5.0G Oct  8 18:10 testfile
dr-xr-xr-x. 30 root root 4.0K Oct  8 18:19 ..
drwxr-xr-x  2 root root 4.0K Nov  5 20:02 gpfs-ces
drwxr-xr-x  2 root root 4.0K Nov  5 20:04 ha
drwxrwxrwx  5 root root 256K Nov  5 20:04 .
drwxr-xr-x  4 root root 4.0K Nov  5 20:35 ces
[root@mastr-51 gpfs1]#

```

Configure the NFS client

To configure the NFS client, complete the following steps:

1. Install packages in the NFS client.

```
[root@hdp2 ~]# yum install nfs-utils rpcbind
Loaded plugins: product-id, search-disabled-repos, subscription-manager
HDP-2.6-GPL-repo-4
| 2.9 kB 00:00:00
HDP-2.6-repo-4
| 2.9 kB 00:00:00
HDP-3.0-GPL-repo-2
| 2.9 kB 00:00:00
HDP-3.0-repo-2
| 2.9 kB 00:00:00
HDP-3.0-repo-3
| 2.9 kB 00:00:00
HDP-3.1-repo-1
| 2.9 kB 00:00:00
HDP-3.1-repo-51
| 2.9 kB 00:00:00
HDP-UTILS-1.1.0.22-repo-1
| 2.9 kB 00:00:00
HDP-UTILS-1.1.0.22-repo-2
| 2.9 kB 00:00:00
HDP-UTILS-1.1.0.22-repo-3
| 2.9 kB 00:00:00
HDP-UTILS-1.1.0.22-repo-4
| 2.9 kB 00:00:00
HDP-UTILS-1.1.0.22-repo-51
| 2.9 kB 00:00:00
ambari-2.7.3.0
| 2.9 kB 00:00:00
epel/x86_64/metalink
| 13 kB 00:00:00
epel
| 5.3 kB 00:00:00
mysql-connectors-community
| 2.5 kB 00:00:00
mysql-tools-community
| 2.5 kB 00:00:00
mysql56-community
| 2.5 kB 00:00:00
rhel-7-server-optional-rpms
| 3.2 kB 00:00:00
```

```

rhel-7-server-rpms
| 3.5 kB 00:00:00
(1/10): mysql-connectors-community/x86_64/primary_db
| 49 kB 00:00:00
(2/10): mysql-tools-community/x86_64/primary_db
| 66 kB 00:00:00
(3/10): epel/x86_64/group_gz
| 90 kB 00:00:00
(4/10): mysql56-community/x86_64/primary_db
| 241 kB 00:00:00
(5/10): rhel-7-server-optional-rpms/7Server/x86_64/updateinfo
| 2.5 MB 00:00:00
(6/10): rhel-7-server-rpms/7Server/x86_64/updateinfo
| 3.4 MB 00:00:00
(7/10): rhel-7-server-optional-rpms/7Server/x86_64/primary_db
| 8.3 MB 00:00:00
(8/10): rhel-7-server-rpms/7Server/x86_64/primary_db
| 62 MB 00:00:01
(9/10): epel/x86_64/primary_db
| 6.9 MB 00:00:08
(10/10): epel/x86_64/updateinfo
| 1.0 MB 00:00:13
Resolving Dependencies
--> Running transaction check
---> Package nfs-utils.x86_64 1:1.3.0-0.61.el7 will be updated
---> Package nfs-utils.x86_64 1:1.3.0-0.65.el7 will be an update
---> Package rpcbind.x86_64 0:0.2.0-47.el7 will be updated
---> Package rpcbind.x86_64 0:0.2.0-48.el7 will be an update
--> Finished Dependency Resolution

```

Dependencies Resolved

```

=====
=====
Package                Arch          Version
Repository              Size
=====
=====
Updating:
nfs-utils                x86_64        1:1.3.0-0.65.el7
rhel-7-server-rpms      412 k
rpcbind                  x86_64        0.2.0-48.el7
rhel-7-server-rpms      60 k

Transaction Summary
=====

```

```
=====
Upgrade 2 Packages
```

```
Total download size: 472 k
```

```
Is this ok [y/d/N]: y
```

```
Downloading packages:
```

```
No Presto metadata available for rhel-7-server-rpms
```

```
(1/2): rpcbind-0.2.0-48.el7.x86_64.rpm
```

```
| 60 kB 00:00:00
```

```
(2/2): nfs-utils-1.3.0-0.65.el7.x86_64.rpm
```

```
| 412 kB 00:00:00
```

```
-----
Total
```

```
1.2 MB/s | 472 kB 00:00:00
```

```
Running transaction check
```

```
Running transaction test
```

```
Transaction test succeeded
```

```
Running transaction
```

```
Updating : rpcbind-0.2.0-48.el7.x86_64
```

```
1/4
```

```
service rpcbind start
```

```
Updating : 1:nfs-utils-1.3.0-0.65.el7.x86_64
```

```
2/4
```

```
Cleanup : 1:nfs-utils-1.3.0-0.61.el7.x86_64
```

```
3/4
```

```
Cleanup : rpcbind-0.2.0-47.el7.x86_64
```

```
4/4
```

```
Verifying : 1:nfs-utils-1.3.0-0.65.el7.x86_64
```

```
1/4
```

```
Verifying : rpcbind-0.2.0-48.el7.x86_64
```

```
2/4
```

```
Verifying : rpcbind-0.2.0-47.el7.x86_64
```

```
3/4
```

```
Verifying : 1:nfs-utils-1.3.0-0.61.el7.x86_64
```

```
4/4
```

```
Updated:
```

```
nfs-utils.x86_64 1:1.3.0-0.65.el7
```

```
rpcbind.x86_64 0:0.2.0-48.el7
```

```
Complete!
```

```
[root@hdp2 ~]#
```

2. Start the NFS client services.

```
[root@hdp2 ~]# service rpcbind start
Redirecting to /bin/systemctl start rpcbind.service
[root@hdp2 ~]#
```

3. Mount the GPFS through the NFS protocol on the NFS client.

```
[root@hdp2 ~]# mkdir /gpfstest
[root@hdp2 ~]# mount 10.63.150.51:/gpfs1 /gpfstest
[root@hdp2 ~]# df -h
```

| Filesystem | Size | Used | Avail | Use% | Mounted on |
|--------------------------------------|------|------|-------|------|-------------|
| /dev/mapper/rhel_stlrx300s6--22-root | 1.1T | 113G | 981G | 11% | / |
| devtmpfs | 126G | 0 | 126G | 0% | /dev |
| tmpfs | 126G | 16K | 126G | 1% | /dev/shm |
| tmpfs | 126G | 510M | 126G | 1% | /run |
| tmpfs | 126G | 0 | 126G | 0% | |
| /sys/fs/cgroup | | | | | |
| /dev/sdd2 | 197M | 191M | 6.6M | 97% | /boot |
| tmpfs | 26G | 0 | 26G | 0% | /run/user/0 |
| 10.63.150.213:/nc_volume2 | 95G | 5.4G | 90G | 6% | /mnt |
| 10.63.150.51:/gpfs1 | 7.3T | 9.1G | 7.3T | 1% | /gpfstest |

```
[root@hdp2 ~]#
```

4. Validate the list of GPFS files in the NFS-mounted folder.

```
[root@hdp2 ~]# cd /gpfstest/
[root@hdp2 gpfstest]# ls
ces gpfs-ces ha testfile
[root@hdp2 gpfstest]# ls -l
total 5242882
drwxr-xr-x 4 root root      4096 Nov  5 15:35 ces
drwxr-xr-x 2 root root      4096 Nov  5 15:02 gpfs-ces
drwxr-xr-x 2 root root      4096 Nov  5 15:04 ha
-rw-r--r-- 1 root root 5368709120 Oct  8 14:10 testfile
[root@hdp2 gpfstest]#
```

5. Move the data from the GPFS- exported NFS to the NetApp NFS by using XCP.


```

[root@hdp2 linux]# ./xcp copy -parallel 20 10.63.150.51:/gpfs1
10.63.150.213:/nc_volume2/
XCP 1.4-17914d6; (c) 2019 NetApp, Inc.; Licensed to Karthikeyan
Nagalingam [NetApp Inc] until Tue Nov  5 12:39:36 2019

xcp: WARNING: your license will expire in less than one week! You can
renew your license at https://xcp.netapp.com
xcp: open or create catalog 'xcp': Creating new catalog in
'10.63.150.51:/gpfs1/catalog'
xcp: WARNING: No index name has been specified, creating one with name:
autoname_copy_2019-11-11_12.14.07.805223
xcp: mount '10.63.150.51:/gpfs1': WARNING: This NFS server only supports
1-second timestamp granularity. This may cause sync to fail because
changes will often be undetectable.
 34 scanned, 32 copied, 32 indexed, 1 giant, 301 MiB in (59.5 MiB/s),
784 KiB out (155 KiB/s), 6s
 34 scanned, 32 copied, 32 indexed, 1 giant, 725 MiB in (84.6 MiB/s),
1.77 MiB out (206 KiB/s), 11s
 34 scanned, 32 copied, 32 indexed, 1 giant, 1.17 GiB in (94.2 MiB/s),
2.90 MiB out (229 KiB/s), 16s
 34 scanned, 32 copied, 32 indexed, 1 giant, 1.56 GiB in (79.8 MiB/s),
3.85 MiB out (194 KiB/s), 21s
 34 scanned, 32 copied, 32 indexed, 1 giant, 1.95 GiB in (78.4 MiB/s),
4.80 MiB out (191 KiB/s), 26s
 34 scanned, 32 copied, 32 indexed, 1 giant, 2.35 GiB in (80.4 MiB/s),
5.77 MiB out (196 KiB/s), 31s
 34 scanned, 32 copied, 32 indexed, 1 giant, 2.79 GiB in (89.6 MiB/s),
6.84 MiB out (218 KiB/s), 36s
 34 scanned, 32 copied, 32 indexed, 1 giant, 3.16 GiB in (75.3 MiB/s),
7.73 MiB out (183 KiB/s), 41s
 34 scanned, 32 copied, 32 indexed, 1 giant, 3.53 GiB in (75.4 MiB/s),
8.64 MiB out (183 KiB/s), 46s
 34 scanned, 32 copied, 32 indexed, 1 giant, 4.00 GiB in (94.4 MiB/s),
9.77 MiB out (230 KiB/s), 51s
 34 scanned, 32 copied, 32 indexed, 1 giant, 4.46 GiB in (94.3 MiB/s),
10.9 MiB out (229 KiB/s), 56s
 34 scanned, 32 copied, 32 indexed, 1 giant, 4.86 GiB in (80.2 MiB/s),
11.9 MiB out (195 KiB/s), 1m1s
Sending statistics...
34 scanned, 33 copied, 34 indexed, 1 giant, 5.01 GiB in (81.8 MiB/s),
12.3 MiB out (201 KiB/s), 1m2s.
[root@hdp2 linux]#

```

6. Validate the GPFS files on the NFS client.

```
[root@hdp2 mnt]# df -Th
```

| Filesystem | Type | Size | Used | Avail | Use% |
|--------------------------------------|----------|------|------|-------|-------|
| Mounted on | | | | | |
| /dev/mapper/rhel_stlrx300s6--22-root | xfs | 1.1T | 113G | 981G | 11% / |
| devtmpfs | devtmpfs | 126G | 0 | 126G | 0% |
| /dev | | | | | |
| tmpfs | tmpfs | 126G | 16K | 126G | 1% |
| /dev/shm | | | | | |
| tmpfs | tmpfs | 126G | 518M | 126G | 1% |
| /run | | | | | |
| tmpfs | tmpfs | 126G | 0 | 126G | 0% |
| /sys/fs/cgroup | | | | | |
| /dev/sdd2 | xfs | 197M | 191M | 6.6M | 97% |
| /boot | | | | | |
| tmpfs | tmpfs | 26G | 0 | 26G | 0% |
| /run/user/0 | | | | | |
| 10.63.150.213:/nc_volume2 | nfs4 | 95G | 5.4G | 90G | 6% |
| /mnt | | | | | |
| 10.63.150.51:/gpfs1 | nfs4 | 7.3T | 9.1G | 7.3T | 1% |
| /gpfstest | | | | | |

```
[root@hdp2 mnt]#
```

```
[root@hdp2 mnt]# ls -ltrha
```

| total | 128K | | | | | | | |
|-------------------|------|------|--------|------|--------|-------|--------|--|
| dr-xr-xr-x | 2 | root | root | 4.0K | Dec 31 | 1969 | | |
| .snapshots | | | | | | | | |
| drwxrwxrwx | 2 | root | root | 4.0K | Feb 14 | 2018 | data | |
| drwxrwxrwx | 3 | root | root | 4.0K | Feb 14 | 2018 | | |
| wcresult | | | | | | | | |
| drwxrwxrwx | 3 | root | root | 4.0K | Feb 14 | 2018 | | |
| wcresult1 | | | | | | | | |
| drwxrwxrwx | 2 | root | root | 4.0K | Feb 14 | 2018 | | |
| wcresult2 | | | | | | | | |
| drwxrwxrwx | 2 | root | root | 4.0K | Feb 16 | 2018 | | |
| wcresult3 | | | | | | | | |
| -rw-r--r-- | 1 | root | root | 2.8K | Feb 20 | 2018 | | |
| READMEdemo | | | | | | | | |
| drwxrwxrwx | 3 | root | root | 4.0K | Jun 28 | 13:38 | scantg | |
| drwxrwxrwx | 3 | root | root | 4.0K | Jun 28 | 13:39 | | |
| scancopyFromLocal | | | | | | | | |
| -rw-r--r-- | 1 | hdfs | hadoop | 1.2K | Jul 3 | 19:28 | f3 | |
| -rw-r--r-- | 1 | hdfs | hadoop | 1.2K | Jul 3 | 19:28 | README | |
| -rw-r--r-- | 1 | hdfs | hadoop | 1.2K | Jul 3 | 19:28 | f9 | |
| -rw-r--r-- | 1 | hdfs | hadoop | 1.2K | Jul 3 | 19:28 | f6 | |
| -rw-r--r-- | 1 | hdfs | hadoop | 1.2K | Jul 3 | 19:28 | f5 | |
| -rw-r--r-- | 1 | hdfs | hadoop | 1.2K | Jul 3 | 19:30 | f4 | |
| -rw-r--r-- | 1 | hdfs | hadoop | 1.2K | Jul 3 | 19:30 | f8 | |

```

-rw-r--r-- 1 hdfs      hadoop      1.2K Jul  3 19:30 f2
-rw-r--r-- 1 hdfs      hadoop      1.2K Jul  3 19:30 f7
drwxrwxrwx 2 root      root        4.0K Jul  9 11:14 test
drwxrwxrwx 3 root      root        4.0K Jul 10 16:35
warehouse
drwxr-xr-x 3          10061 tester1      4.0K Jul 15 14:40 sdd1
drwxrwxrwx 3 testeruser1 hadoopkerberosgroup 4.0K Aug 20 17:00
kermkdir
-rw-r--r-- 1 testeruser1 hadoopkerberosgroup 0 Aug 21 14:20 newfile
drwxrwxrwx 2 testeruser1 hadoopkerberosgroup 4.0K Aug 22 10:13
teragen1copy_3
drwxrwxrwx 2 testeruser1 hadoopkerberosgroup 4.0K Aug 22 10:33
teragen2copy_1
-rw-rwxr-- 1 root      hdfs          1.2K Sep 19 16:38 R1
drwx----- 3 root      root        4.0K Sep 20 17:28 user
-rw-r--r-- 1 root      root        5.0G Oct  8 14:10
testfile
drwxr-xr-x 2 root      root        4.0K Nov  5 15:02 gpfs-
ces
drwxr-xr-x 2 root      root        4.0K Nov  5 15:04 ha
drwxr-xr-x 4 root      root        4.0K Nov  5 15:35 ces
dr-xr-xr-x. 26 root      root        4.0K Nov  6 11:40 ..
drwxrwxrwx 21 root      root        4.0K Nov 11 12:14 .
drwxrwxrwx 7 nobody    nobody      4.0K Nov 11 12:14 catalog
[root@hdp2 mnt]#

```

[Next: MapR-FS to ONTAP NFS.](#)

MapR-FS to ONTAP NFS

[Previous: GPFS to NFS - Detailed steps.](#)

This section provides the detailed steps needed to move MapR-FS data into ONTAP NFS by using NetApp XCP.

1. Provision three LUNs for each MapR node and give the LUNs ownership of all MapR nodes.
2. During installation, choose newly added LUNs for MapR cluster disks that are used for MapR-FS.
3. Install a MapR cluster according to the [MapR 6.1 documentation](#).
4. Check the basic Hadoop operations using MapReduce commands such as `hadoop jar xxx`.
5. Keep customer data in MapR-FS. For example, we generated approximately a terabyte of sample data in MapR-FS by using Teragen.
6. Configure MapR-FS as NFS export.
 - a. Disable the nlockmgr service on all MapR nodes.

```

root@workkr-138: ~$ rpcinfo -p
    program vers proto  port  service
    100000    4   tcp    111   portmapper
    100000    3   tcp    111   portmapper
    100000    2   tcp    111   portmapper
    100000    4   udp    111   portmapper
    100000    3   udp    111   portmapper
    100000    2   udp    111   portmapper
    100003    4   tcp    2049  nfs
    100227    3   tcp    2049  nfs_acl
    100003    4   udp    2049  nfs
    100227    3   udp    2049  nfs_acl
    100021    3   udp    55270 nlockmgr
    100021    4   udp    55270 nlockmgr
    100021    3   tcp    35025 nlockmgr
    100021    4   tcp    35025 nlockmgr
    100003    3   tcp    2049  nfs
    100005    3   tcp    2049  mountd
    100005    1   tcp    2049  mountd
    100005    3   udp    2049  mountd
    100005    1   udp    2049  mountd
root@workkr-138: ~$

root@workkr-138: ~$ rpcinfo -d 100021 3
root@workkr-138: ~$ rpcinfo -d 100021 4

```

- b. Export specific folders from MapR-FS on all MapR nodes in the `/opt/mapr/conf/exports` file. Do not export the parent folder with different permissions when you export sub folders.

```

[mapr@workr-138 ~]$ cat /opt/mapr/conf/exports
# Sample Exports file
# for /mapr exports
# <Path> <exports_control>
#access_control -> order is specific to default
# list the hosts before specifying a default for all
# a.b.c.d,1.2.3.4(ro) d.e.f.g(ro) (rw)
# enforces ro for a.b.c.d & 1.2.3.4 and everybody else is rw
# special path to export clusters in mapr-clusters.conf. To disable
exporting,
# comment it out. to restrict access use the exports_control
#
#/mapr (rw)
#karthik
/mapr/my.cluster.com/tmp/testnfs /maprnfs3 (rw)
#to export only certain clusters, comment out the /mapr & uncomment.
#/mapr/clustername (rw)
#to export /mapr only to certain hosts (using exports_control)
#/mapr a.b.c.d(rw),e.f.g.h(ro)
# export /mapr/cluster1 rw to a.b.c.d & ro to e.f.g.h (denied for
others)
#/mapr/cluster1 a.b.c.d(rw),e.f.g.h(ro)
# export /mapr/cluster2 only to e.f.g.h (denied for others)
#/mapr/cluster2 e.f.g.h(rw)
# export /mapr/cluster3 rw to e.f.g.h & ro to others
#/mapr/cluster2 e.f.g.h(rw) (ro)
#to export a certain cluster, volume or a subdirectory as an alias,
#comment out /mapr & uncomment
#/mapr/clustername /alias1 (rw)
#/mapr/clustername/vol /alias2 (rw)
#/mapr/clustername/vol/dir /alias3 (rw)
#only the alias will be visible/exposed to the nfs client not the
mapr path, host options as before
[mapr@workr-138 ~]$

```

7. Refresh the MapR-FS NFS service.

```

root@workr-138: tmp$ maprccli nfsmgmt refreshexports
ERROR (22) - You do not have a ticket to communicate with
127.0.0.1:9998. Retry after obtaining a new ticket using maprlogin
root@workr-138: tmp$ su - mapr
[mapr@workr-138 ~]$ maprlogin password -cluster my.cluster.com
[Password for user 'mapr' at cluster 'my.cluster.com': ]
MapR credentials of user 'mapr' for cluster 'my.cluster.com' are written
to '/tmp/maprticket_5000'
[mapr@workr-138 ~]$ maprccli nfsmgmt refreshexports

```

8. Assign a virtual IP range to a specific server or a set of servers in the MapR cluster. Then the MapR cluster assigns an IP to a specific server for NFS data access. The IPs enable high availability, which means that, if a server or network with a particular IP experiences failure, the next IP from the range of IPs can be used for NFS access.



If you would like to provide NFS access from all MapR nodes, then you can assign a set of virtual IPs to each server, and you can use the resources from each MapR node for NFS data access.

The screenshot shows the MapR web interface for 'my.cluster.com' with the 'Services / NFS V3 Gateway' page selected. The 'NFS Setup and VIP Assignment' section contains a table with the following data:

| VIP Range | Virtual IP | Node Name | Physical IP | MAC Address |
|--|------------------------------|--|--------------------------------|--------------------------------------|
| <input type="checkbox"/> 10.63.150.92 - 10.63.150.93 | (Pending) -- | -- | -- | -- |
| <input type="checkbox"/> 10.63.150.96 - 10.63.150.97 | 10.63.150.96 10.63.150.97 | workr-138.netapp.com workr-138.netapp.com | 10.63.150.138 10.63.150.138 | 90:1b:0ed1:5d:f9 90:1b:0ed1:5d:f9 |

Page 1 of 1 | Rows 10 | Total Items: 1 - 2 of 2



9. Check the virtual IPs assigned on each MapR node and use them for NFS data access.

```
root@workr-138: ~$ ip a
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN
    group default qlen 1000
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
        valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host
```

```

        valid_lft forever preferred_lft forever
2: ens3f0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 9000 qdisc mq state UP
group default qlen 1000
    link/ether 90:1b:0e:d1:5d:f9 brd ff:ff:ff:ff:ff:ff
    inet 10.63.150.138/24 brd 10.63.150.255 scope global noprefixroute
ens3f0
    valid_lft forever preferred_lft forever
    inet 10.63.150.96/24 scope global secondary ens3f0:~m0
    valid_lft forever preferred_lft forever
    inet 10.63.150.97/24 scope global secondary ens3f0:~m1
    valid_lft forever preferred_lft forever
    inet6 fe80::921b:eff:fed1:5df9/64 scope link
    valid_lft forever preferred_lft forever
3: eno1: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP
group default qlen 1000
    link/ether 90:1b:0e:d1:af:b4 brd ff:ff:ff:ff:ff:ff
4: ens3f1: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP
group default qlen 1000
    link/ether 90:1b:0e:d1:5d:fa brd ff:ff:ff:ff:ff:ff
5: eno2: <NO-CARRIER,BROADCAST,MULTICAST,UP> mtu 1500 qdisc mq state
DOWN group default qlen 1000
    link/ether 90:1b:0e:d1:af:b5 brd ff:ff:ff:ff:ff:ff
[root@workr-138: ~]$
[root@workr-140 ~]# ip a
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN
group default qlen 1000
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
        valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host
        valid_lft forever preferred_lft forever
2: ens3f0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 9000 qdisc mq state UP
group default qlen 1000
    link/ether 90:1b:0e:d1:5e:03 brd ff:ff:ff:ff:ff:ff
    inet 10.63.150.140/24 brd 10.63.150.255 scope global noprefixroute
ens3f0
    valid_lft forever preferred_lft forever
    inet 10.63.150.92/24 scope global secondary ens3f0:~m0
    valid_lft forever preferred_lft forever
    inet6 fe80::921b:eff:fed1:5e03/64 scope link noprefixroute
    valid_lft forever preferred_lft forever
3: eno1: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP
group default qlen 1000
    link/ether 90:1b:0e:d1:af:9a brd ff:ff:ff:ff:ff:ff
4: ens3f1: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP
group default qlen 1000

```



```

link/ether 90:1b:0e:d1:5e:04 brd ff:ff:ff:ff:ff:ff
5: eno2: <NO-CARRIER,BROADCAST,MULTICAST,UP> mtu 1500 qdisc mq state
DOWN group default qlen 1000
link/ether 90:1b:0e:d1:af:9b brd ff:ff:ff:ff:ff:ff
[root@workr-140 ~]#

```

10. Mount the NFS- exported MapR-FS using the assigned virtual IP for checking the NFS operation. However, this step is not required for data transfer using NetApp XCP.

```

root@workr-138: tmp$ mount -v -t nfs 10.63.150.92:/maprnfs3
/tmp/testmount/
mount.nfs: timeout set for Thu Dec  5 15:31:32 2019
mount.nfs: trying text-based options
'vers=4.1,addr=10.63.150.92,clientaddr=10.63.150.138'
mount.nfs: mount(2): Protocol not supported
mount.nfs: trying text-based options
'vers=4.0,addr=10.63.150.92,clientaddr=10.63.150.138'
mount.nfs: mount(2): Protocol not supported
mount.nfs: trying text-based options 'addr=10.63.150.92'
mount.nfs: prog 100003, trying vers=3, prot=6
mount.nfs: trying 10.63.150.92 prog 100003 vers 3 prot TCP port 2049
mount.nfs: prog 100005, trying vers=3, prot=17
mount.nfs: trying 10.63.150.92 prog 100005 vers 3 prot UDP port 2049
mount.nfs: portmap query retrying: RPC: Timed out
mount.nfs: prog 100005, trying vers=3, prot=6
mount.nfs: trying 10.63.150.92 prog 100005 vers 3 prot TCP port 2049
root@workr-138: tmp$ df -h

```

| Filesystem | Size | Used | Avail | Use% | Mounted on |
|------------------------|------|------|-------|------|----------------|
| /dev/sda7 | 84G | 48G | 37G | 57% | / |
| devtmpfs | 126G | 0 | 126G | 0% | /dev |
| tmpfs | 126G | 0 | 126G | 0% | /dev/shm |
| tmpfs | 126G | 19M | 126G | 1% | /run |
| tmpfs | 126G | 0 | 126G | 0% | /sys/fs/cgroup |
| /dev/sdd1 | 3.7T | 201G | 3.5T | 6% | /mnt/sdd1 |
| /dev/sda6 | 946M | 220M | 726M | 24% | /boot |
| tmpfs | 26G | 0 | 26G | 0% | /run/user/5000 |
| gpfs1 | 7.3T | 9.1G | 7.3T | 1% | /gpfs1 |
| tmpfs | 26G | 0 | 26G | 0% | /run/user/0 |
| localhost:/mapr | 100G | 0 | 100G | 0% | /mapr |
| 10.63.150.92:/maprnfs3 | 53T | 8.4G | 53T | 1% | /tmp/testmount |

```

root@workr-138: tmp$

```

11. Configure NetApp XCP to transfer data from the MapR-FS NFS gateway to ONTAP NFS.
 - a. Configure the catalog location for XCP.

```
[root@hdp2 linux]# cat /opt/NetApp/xFiles/xcp/xcp.ini
# Sample xcp config
[xcp]
#catalog = 10.63.150.51:/gpfs1
catalog = 10.63.150.213:/nc_volume1
```

- b. Copy the license file to /opt/NetApp/xFiles/xcp/.

```
root@workr-138: src$ cd /opt/NetApp/xFiles/xcp/
root@workr-138: xcp$ ls -ltrha
total 252K
drwxr-xr-x 3 root root 16 Apr 4 2019 ..
-rw-r--r-- 1 root root 105 Dec 5 19:04 xcp.ini
drwxr-xr-x 2 root root 59 Dec 5 19:04 .
-rw-r--r-- 1 faiz89 faiz89 336 Dec 6 21:12 license
-rw-r--r-- 1 root root 192 Dec 6 21:13 host
-rw-r--r-- 1 root root 236K Dec 17 14:12 xcp.log
root@workr-138: xcp$
```

- c. Activate XCP using the `xcp activate` command.
- d. Check the source for NFS export.

```
[root@hdp2 linux]# ./xcp show 10.63.150.92
XCP 1.4-17914d6; (c) 2019 NetApp, Inc.; Licensed to Karthikeyan
Nagalingam [NetApp Inc] until Wed Feb  5 11:07:27 2020
getting pmap dump from 10.63.150.92 port 111...
getting export list from 10.63.150.92...
sending 1 mount and 4 nfs requests to 10.63.150.92...
== RPC Services ==
'10.63.150.92': TCP rpc services: MNT v1/3, NFS v3/4, NFSACL v3, NLM
v1/3/4, PMAP v2/3/4, STATUS v1
'10.63.150.92': UDP rpc services: MNT v1/3, NFS v4, NFSACL v3, NLM
v1/3/4, PMAP v2/3/4, STATUS v1
== NFS Exports ==
Mounts  Errors  Server
      1      0  10.63.150.92
      Space      Files      Space      Files
      Free      Free      Used      Used Export
  52.3 TiB    53.7B    8.36 GiB    53.7B 10.63.150.92:/maprnfs3
== Attributes of NFS Exports ==
drwxr-xr-x --- root root 2 2 10m51s 10.63.150.92:/maprnfs3
1.77 KiB in (8.68 KiB/s), 3.16 KiB out (15.5 KiB/s), 0s.
[root@hdp2 linux]#
```

- e. Transfer the data using XCP from multiple MapR nodes from multiple source IPs and multiple destination IPs (ONTAP LIFs).

```
root@workr-138: linux$ ./xcp_yatin copy --parallel 20
10.63.150.96,10.63.150.97:/maprnfs3/tg4
10.63.150.85,10.63.150.86:/datapipeline_dataset/tg4_dest
XCP 1.6-dev; (c) 2019 NetApp, Inc.; Licensed to Karthikeyan
Nagalingam [NetApp Inc] until Wed Feb  5 11:07:27 2020
xcp: WARNING: No index name has been specified, creating one with
name: autaname_copy_2019-12-06_21.14.38.652652
xcp: mount '10.63.150.96,10.63.150.97:/maprnfs3/tg4': WARNING: This
NFS server only supports 1-second timestamp granularity. This may
cause sync to fail because changes will often be undetectable.
  130 scanned, 128 giants, 3.59 GiB in (723 MiB/s), 3.60 GiB out (724
MiB/s), 5s
  130 scanned, 128 giants, 8.01 GiB in (889 MiB/s), 8.02 GiB out (890
MiB/s), 11s
  130 scanned, 128 giants, 12.6 GiB in (933 MiB/s), 12.6 GiB out (934
MiB/s), 16s
  130 scanned, 128 giants, 16.7 GiB in (830 MiB/s), 16.7 GiB out (831
MiB/s), 21s
  130 scanned, 128 giants, 21.1 GiB in (907 MiB/s), 21.1 GiB out (908
MiB/s), 26s
```

```

130 scanned, 128 giants, 25.5 GiB in (893 MiB/s), 25.5 GiB out (894
MiB/s), 31s
130 scanned, 128 giants, 29.6 GiB in (842 MiB/s), 29.6 GiB out (843
MiB/s), 36s
...
[root@workr-140 linux]# ./xcp_yatin copy --parallel 20
10.63.150.92:/maprnfs3/tg4_2
10.63.150.85,10.63.150.86:/datapipeline_dataset/tg4_2_dest
XCP 1.6-dev; (c) 2019 NetApp, Inc.; Licensed to Karthikeyan
Nagalingam [NetApp Inc] until Wed Feb  5 11:07:27 2020
xcp: WARNING: No index name has been specified, creating one with
name: autoname_copy_2019-12-06_21.14.24.637773
xcp: mount '10.63.150.92:/maprnfs3/tg4_2': WARNING: This NFS server
only supports 1-second timestamp granularity. This may cause sync to
fail because changes will often be undetectable.
130 scanned, 128 giants, 4.39 GiB in (896 MiB/s), 4.39 GiB out (897
MiB/s), 5s
130 scanned, 128 giants, 9.94 GiB in (1.10 GiB/s), 9.96 GiB out
(1.10 GiB/s), 10s
130 scanned, 128 giants, 15.4 GiB in (1.09 GiB/s), 15.4 GiB out
(1.09 GiB/s), 15s
130 scanned, 128 giants, 20.1 GiB in (953 MiB/s), 20.1 GiB out (954
MiB/s), 20s
130 scanned, 128 giants, 24.6 GiB in (928 MiB/s), 24.7 GiB out (929
MiB/s), 25s
130 scanned, 128 giants, 29.0 GiB in (877 MiB/s), 29.0 GiB out (878
MiB/s), 31s
130 scanned, 128 giants, 33.2 GiB in (852 MiB/s), 33.2 GiB out (853
MiB/s), 36s
130 scanned, 128 giants, 37.8 GiB in (941 MiB/s), 37.8 GiB out (942
MiB/s), 41s
130 scanned, 128 giants, 42.0 GiB in (860 MiB/s), 42.0 GiB out (861
MiB/s), 46s
130 scanned, 128 giants, 46.1 GiB in (852 MiB/s), 46.2 GiB out (853
MiB/s), 51s
130 scanned, 128 giants, 50.1 GiB in (816 MiB/s), 50.2 GiB out (817
MiB/s), 56s
130 scanned, 128 giants, 54.1 GiB in (819 MiB/s), 54.2 GiB out (820
MiB/s), 1m1s
130 scanned, 128 giants, 58.5 GiB in (897 MiB/s), 58.6 GiB out (898
MiB/s), 1m6s
130 scanned, 128 giants, 62.9 GiB in (900 MiB/s), 63.0 GiB out (901
MiB/s), 1m11s
130 scanned, 128 giants, 67.2 GiB in (876 MiB/s), 67.2 GiB out (877
MiB/s), 1m16s

```

f. Check the load distribution on the storage controller.

```
Hadoop-AFF8080::*> statistics show-periodic -interval 2 -iterations 0
-summary true -object nic_common -counter rx_bytes|tx_bytes -node
Hadoop-AFF8080-01 -instance e3b
Hadoop-AFF8080: nic_common.e3b: 12/6/2019 15:55:04
rx_bytes tx_bytes
-----
879MB    4.67MB
856MB    4.46MB
973MB    5.66MB
986MB    5.88MB
945MB    5.30MB
920MB    4.92MB
894MB    4.76MB
902MB    4.79MB
886MB    4.68MB
892MB    4.78MB
908MB    4.96MB
905MB    4.85MB
899MB    4.83MB

Hadoop-AFF8080::*> statistics show-periodic -interval 2 -iterations 0
-summary true -object nic_common -counter rx_bytes|tx_bytes -node
Hadoop-AFF8080-01 -instance e9b
Hadoop-AFF8080: nic_common.e9b: 12/6/2019 15:55:07
rx_bytes tx_bytes
-----
950MB    4.93MB
991MB    5.84MB
959MB    5.63MB
914MB    5.06MB
903MB    4.81MB
899MB    4.73MB
892MB    4.71MB
890MB    4.72MB
905MB    4.86MB
902MB    4.90MB
```

[Next: Where to find additional information.](#)

Where to find additional information

[Previous: MapR-FS to ONTAP NFS.](#)

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

- NetApp In-Place Analytics Module Best Practices
<https://www.netapp.com/us/media/tr-4382.pdf>
- NetApp FlexGroup Volume Best Practices and Implementation Guide
<https://www.netapp.com/us/media/tr-4571.pdf>
- NetApp Product Documentation
<https://www.netapp.com/us/documentation/index.aspx>

Version history

| Version | Date | Document version history |
|-------------|---------------|--|
| Version 3.0 | January 2022 | Directly move data from HDFS and MapR-FS to NFS by using NetApp XCP. |
| Version 2.0 | January 2020 | XCP included as the default data mover. Added MapR-FS to NFS and GPFS to NFS data transfer. |
| Version 1.0 | November 2018 | Initial release. |

Copyright information

Copyright © 2023 NetApp, Inc. All Rights Reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system—without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP “AS IS” AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

LIMITED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (b)(3) of the Rights in Technical Data -Noncommercial Items at DFARS 252.227-7013 (FEB 2014) and FAR 52.227-19 (DEC 2007).

Data contained herein pertains to a commercial product and/or commercial service (as defined in FAR 2.101) and is proprietary to NetApp, Inc. All NetApp technical data and computer software provided under this Agreement is commercial in nature and developed solely at private expense. The U.S. Government has a non-exclusive, non-transferrable, nonsublicensable, worldwide, limited irrevocable license to use the Data only in connection with and in support of the U.S. Government contract under which the Data was delivered. Except as provided herein, the Data may not be used, disclosed, reproduced, modified, performed, or displayed without the prior written approval of NetApp, Inc. United States Government license rights for the Department of Defense are limited to those rights identified in DFARS clause 252.227-7015(b) (FEB 2014).

Trademark information

NETAPP, the NETAPP logo, and the marks listed at <http://www.netapp.com/TM> are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.