NetApp In-Place Analytics Module 3.1

Installation and Setup Guide

July 2019 | 215-14067_A0
doccomments@netapp.com

NetApp®

Contents

Overview of NetApp In-Place Analytics Module 4
Support for Ranger authorization and Kerberos authentication 4
Installing NetApp In-Place Analytics Module 6
System requirements for NetApp In-Place Analytics Module 6
Configuring ONTAP for NetApp In-Place Analytics Module 7
Configuring the Hadoop cluster for NetApp In-Place Analytics Module 8
Specifying NFS as the scheme for the Hadoop file system 8
Installing NetApp In-Place Analytics Module on a Hortonworks cluster 9
Updating the NetApp In-Place Analytics Module configurations for Hadoop components 10
Creating the JSON configuration file 11
Parameters for the JSON configuration file 12
Verifying the installation of NetApp In-Place Analytics Module 17
Uninstalling the NetApp In-Place Analytics Module 19
Copyright 20
Trademark 21
How to send comments about documentation and receive update notifications 22
Index 23

Overview of NetApp In-Place Analytics Module

The NetApp In-Place Analytics Module enables analytics software such as Apache Hadoop and Apache Spark to access data by using the NetApp ONTAP® data management software, the Network File System (NFS) protocol, and a simple configuration file change.

The NetApp In-Place Analytics Module works with Apache Hadoop and Apache Spark. It uses a simple configuration file change that enables data analytics on ONTAP storage. By using ONTAP software, the NetApp In-Place Analytics Module decouples analytics (computation) from storage, thereby utilizing the benefits of NAS to share data.

You can deploy NetApp In-Place Analytics Module by running the Hadoop Distributed File System (HDFS) as the primary file system and using the NetApp In-Place Analytics Module to analyze data on the NFS storage systems in the same Hadoop cluster.

Related information

NetApp Technical Report 4382: NetApp In-Place Analytics Module Best Practices

Support for Ranger authorization and Kerberos authentication

NetApp In-Place Analytics Module supports Ranger for secure authorization and Kerberos for authentication, encryption and data integrity.

Ranger authorization

NetApp In-Place Analytics Module consists of a Ranger client that communicates with the Ranger admin server to download configured policies and has a Ranger policy enforcement engine to enforce policies when data is accessed using NFS.

When configuring Ranger for NetApp In-Place Analytics Module, you must specify an HDFS repository to create policies for NFS. NetApp In-Place Analytics Module uses this HDFS repository to connect to the Ranger server to get the policies. Only one repository is supported with NetApp In-Place Analytics Module. You can create multiple policies under the same HDFS repository. When creating a Ranger policy for NFS, you must specify the policy name, resource path (path to the NFS server), user or groups for the policy, and permissions (read, write, and execute).

Kerberos authentication

NetApp In-Place Analytics Module 3.1 supports Kerberos for authentication and enables data integrity and encryption of the data that is exchanged between the Hadoop nodes and ONTAP storage. In a Hadoop environment with Kerberos enabled, the user is already authenticated. NetApp In-Place Analytics Module uses Kerberos to enable data access over NFS for authenticated users by maintaining data integrity and confidentiality depending on the Kerberos flavor that is configured in the NetApp In-Place Analytics Module configuration file. The following three flavors of Kerberos v5 are supported:

- **`krb5p`**: Authenticated users access data that is integrity-protected and is also encrypted.
- **`krb5i`**: Authenticated users access data that is only integrity-protected.
- **`krb5`**: Allows authenticated users to access data that is neither integrity-protected nor encrypted.

You must ensure that the following requirements are met for using Kerberos with NetApp In-Place Analytics Module:

- All user keytab files must be installed on all the Hadoop nodes that use NetApp In-Place Analytics Module.

  For using Kerberos service in Hadoop, each user must be added to the Kerberos Key Distribution Center (KDC) and a keytab file for the user principal name (for example, username@realm.com) must be created. The keytab file contains the machine credentials for the user.

- The ONTAP NFS server must be configured with the KDC of the Hadoop cluster.

  The Hadoop cluster and ONTAP storage system must be in the same Kerberos realm and also all of the data LIFs must be configured in the same realm. All data LIFs in the NFS server must be configured with identical service principal name (SPN) for the NFS service in the SVM.

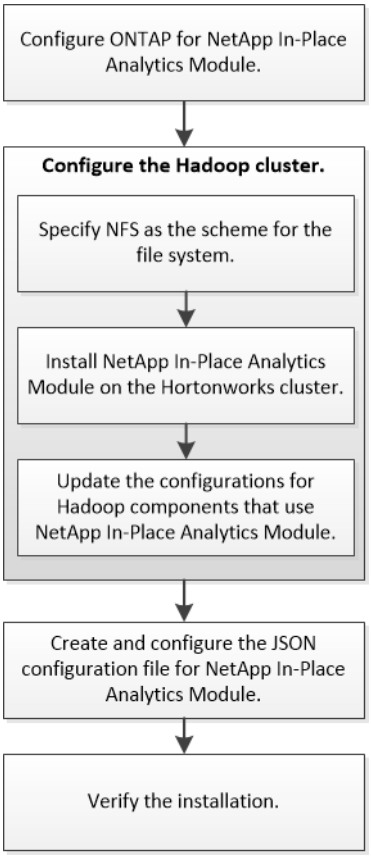
- All user to UID mapping and groups to GID mapping must be synchronized across all Hadoop nodes and ONTAP.

Related information

NetApp Technical Report 4744: Secure Hadoop using Apache Ranger with NetApp In-Place Analytics Module Deployment Guide

Installing NetApp In-Place Analytics Module

Installing NetApp In-Place Analytics Module involves configuring ONTAP, configuring the Hadoop cluster, downloading and installing the NetApp In-Place Analytics Module plug-in, creating the JSON configuration file, and verifying the installation.



Steps

1. System requirements for NetApp In-Place Analytics Module on page 6
2. Configuring ONTAP for NetApp In-Place Analytics Module on page 7
3. Configuring the Hadoop cluster for NetApp In-Place Analytics Module on page 8
4. Creating the JSON configuration file on page 11
5. Verifying the installation of NetApp In-Place Analytics Module on page 17

System requirements for NetApp In-Place Analytics Module

You must ensure that the ONTAP storage system and Hadoop cluster meet the required configuration to install NetApp In-Place Analytics Module.

ONTAP requirements

Your ONTAP cluster must be running one of the following versions:

- Clustered Data ONTAP 8.3 or later

- ONTAP 9 or later

Hadoop requirements

The following versions of Hadoop and its distributions are required:

- Hadoop 2.4 or later
- Hortonworks 2.1 or later

Configuring ONTAP for NetApp In-Place Analytics Module

You must create a storage virtual machine (SVM), a volume, data LIFs, and configure NFS access on the ONTAP storage system so that the NetApp In-Place Analytics Module can enable Hadoop clusters to access data using NFS.

Steps

1. Create an SVM with NFS access.
2. Create a volume in the SVM.
3. Create at least one data logical interface (LIF) that is accessible from each of the NodeManagers in the Hadoop cluster.

   The best practice is to create a private network or separate VLAN with NodeManagers and storage systems.

   Note: Using more data LIFs improves NFS performance.

4. Enable NFS mount requests from nonreserved ports.

   a. Allow NFS mount requests from nonreserved ports for the SVM:

   ```
   vserver nfs modify -vserver svm_name -mount-rootonly disabled
   ```

   b. Allow all NFS requests from nonreserved ports for the SVM:

   ```
   vserver nfs modify -vserver svm_name -nfs-rootonly disabled
   ```

   Note: You can skip this step if you are running ONTAP 9.5 or later.

5. At the advanced privilege level, set the maximum transfer sizes for all TCP connections using NFS.

If you are running...	Run this command...
Clustered Data ONTAP 8.3 or later	`vserver nfs modify -vserver svm_name -v3-tcp-max-read-size 1048576` `vserver nfs modify -vserver svm_name -v3-tcp-max-write-size 65536`
ONTAP 9 or later	`vserver nfs modify -vserver svm_name -tcp-max-xfer-size 1048576`

6. If you want to use Kerberos for authentication, configure the SVM with the Key Distribution Center (KDC) of the Hadoop cluster.

   a. Enable Kerberos on the data LIFs of the SVM:

   ```
   vserver nfs kerberos interface enable -vserver svm_name -lif data_lif -spn service_principal_name
   ```

All data LIFs that use NetApp In-Place Analytics Module must use the same service principal name (SPN).

Example

```
cluster1::> vserver nfs kerberos interface enable -vserver vs1 -
lif data1 -spn nfs/hadooplif.netapp.com@NETAPP.COM
```

b. Verify the Kerberos configuration for the SVM:

`vserver nfs kerberos interface show -vserver svm_name`

Example

```
cluster1::> vserver nfs kerberos interface show -vserver vs1

               Logical
Vserver        Interface     Address         Kerberos SPN
-------------- ------------- --------------- -------- ----------------------
vs1            data1         10.231.43.115   enabled  nfs/
hadooplif.netapp.com@NETAPP.COM
vs1            data2         10.231.43.116   enabled  nfs/
hadooplif.netapp.com@NETAPP.COM
2 entries were displayed.
```

Related concepts

Support for Ranger authorization and Kerberos authentication on page 4

Related information

ONTAP 9 Documentation Center

Configuring the Hadoop cluster for NetApp In-Place Analytics Module

Configuring the Hadoop cluster involves specifying NFS as the scheme for the Hadoop file system, installing the NetApp In-Place Analytics Module, and updating the configurations for Hadoop components, such as YARN and MapReduce.

Specifying NFS as the scheme for the Hadoop file system

You must specify NFS as the scheme for the Hadoop file system by updating the custom `core-site.xml` file in the Hadoop cluster so that NFS file system can be used in parallel with the default Hadoop Distributed File System (HDFS).

Steps

1. Log in to the Hortonworks cluster through the Ambari UI:

   `http://ambari_server_name:8080/`

2. In the **HDFS** tab, click **Configs** > **Advanced**.
3. In the **Custom core-site** section, click **Add Property** and add the following information in the text box.

```
fs.AbstractFileSystem.nfs.impl=org.apache.hadoop.netapp.fs.nfs.NFSv3Ab
stractFilesystem
fs.nfs.impl=org.apache.hadoop.netapp.fs.nfs.NFSv3FileSystem
fs.nfs.prefetch=false
fs.nfs.configuration=/etc/NetAppNFSConnector/conf/nfs-mapping.json
```

The `fs.nfs.configuration` parameter points to the path where the `nfs-mapping.json` file is saved.

After you finish

Verify the installation by running a service check and resolve errors, if any.

Click **NetApp In-Place Analytics Module > Service Actions > Run Service Check**.

Installing NetApp In-Place Analytics Module on a Hortonworks cluster

You can download and install the NetApp In-Place Analytics Module on a Hortonworks cluster by placing the installation JAR files in the Hadoop client directory and the library path of all of the Hadoop components in all the nodes of the Hortonworks cluster.

Before you begin

The Hortonworks cluster must be already set up.

Steps

1. Download the NetApp In-Place Analytics Module software from the NetApp Support Site to a temporary directory in the Hortonworks cluster.

   The .zip file contains the following JAR files:

   - `hadoop-nfs-connector-`*`version_number`*`.jar`
   - `hadoop-nfs-2.7.1.jar`
   - `ranger-plugins-audit-1.0.0.jar`

   NetApp Downloads: Software

2. Rename the `hadoop-nfs-2.7.1.jar` file to `hadoop-nfs-`*`hdp_version`*`.jar`, where *`hdp_version`* is the HDP version of the Hortonworks cluster.

3. On each node of the Hortonworks cluster, copy all of the JAR files to the `/usr/hdp/current/hadoop-client/` directory.

   Important: You must ensure that the permission on the `hadoop-nfs-connector-`*`version_number`*`.jar`, `hadoop-nfs-`*`hdp_version`*`.jar`, and `ranger-plugins-audit-1.0.0.jar` is set to **`read`** and **`execute`** for all the users on the system.

4. On each node of the Hortonworks cluster, copy all of the JAR files to the library path of the Hadoop components that must use NetApp In-Place Analytics module:

   a. Copy all of the JAR files to the `usr/hdp/`*`hdp_version`*`/hadoop/lib/` directory.

   b. For YARN, copy all of the JAR files to the `usr/hdp/`*`hdp_version`*`/hadoop-yarn/lib/` directory.

   c. For Hive, copy all of the JAR files to the `/usr/hdp/`*`hdp_version`*`/hive/lib/` directory.

   d. For Spark, copy all of the JAR files to the `/usr/hdp/`*`hdp_version`*`/spark/lib/` directory.

e. For HBase, copy all of the JAR files to the `/usr/hdp/hdp_version/hbase/lib/` directory.

Example

```
[root@node1 ~]# cp /root/temp/hadoop-nfs-connector-3.1.jar /usr/hdp/
2.6.0.3-8/hadoop/lib/hadoop-nfs-connector-3.1.jar
[root@node1 ~]# cp /root/temp/hadoop-nfs-2.6.0.3-8.jar /usr/hdp/
2.6.0.3-8/hadoop/lib/hadoop-nfs-2.6.0.3-8.jar
[root@node1 ~]# cp /root/temp/ranger-plugins-audit-1.0.0.jar /usr/hdp/
2.6.0.3-8/hadoop/lib/ranger-plugins-audit-1.0.0.jar
[root@node1 ~]# cp /root/temp/hadoop-nfs-connector-3.1.jar /usr/hdp/
2.6.0.3-8/hadoop/hadoop-nfs-connector-3.1.jar
[root@node1 ~]# cp /root/temp/hadoop-nfs-2.6.0.3-8.jar /usr/hdp/
2.6.0.3-8/hadoop/hadoop-nfs-2.6.0.3-8.jar
[root@node1 ~]# cp /root/temp/ranger-plugins-audit-1.0.0.jar /usr/hdp/
2.6.0.3-8/hadoop/ranger-plugins-audit-1.0.0.jar
[root@node1 ~]# cp /root/temp/hadoop-nfs-connector-3.1.jar /usr/hdp/
2.6.0.3-8/hadoop-yarn/lib/hadoop-nfs-connector-3.1.jar
[root@node1 ~]# cp /root/temp/hadoop-nfs-2.6.0.3-8.jar /usr/hdp/
2.6.0.3-8/hadoop-yarn/lib/hadoop-nfs-2.6.0.3-8.jar
[root@node1 ~]# cp /root/temp/ranger-plugins-audit-1.0.0.jar /usr/hdp/
2.6.0.3-8/hadoop-yarn/lib/ranger-plugins-audit-1.0.0.jar
[root@node1 ~]# cp /root/temp/hadoop-nfs-connector-3.1.jar /usr/hdp/
2.6.0.3-8/hive/lib/hadoop-nfs-connector-3.1.jar
[root@node1 ~]# cp /root/temp/hadoop-nfs-2.6.0.3-8.jar /usr/hdp/
2.6.0.3-8/hive/lib/hadoop-nfs-2.6.0.3-8.jar
[root@node1 ~]# cp /root/temp/ranger-plugins-audit-1.0.0.jar /usr/hdp/
2.6.0.3-8/hive/lib/ranger-plugins-audit-1.0.0.jar
```

If the library path is not updated for any Hadoop component that uses NetApp In-Place Analytics Module, a `java.lang.ClassNotFoundException` error is generated.

5. If you want to use Kerberos authentication, copy the `keytab` file of each user to all nodes of the Hadoop cluster.

   The default location for the `keytab` directory is the `/etc/security/keytab` directory. You can modify the path of the `keytab` directory in the `nfs-mapping.json` file, if required.

Updating the NetApp In-Place Analytics Module configurations for Hadoop components

You can update the configurations for NetApp In-Place Analytics Module for various Hadoop components, such as MapReduce and YARN.

Steps

1. From the Ambari UI, update the classpath of MapReduce for using the NetApp In-Place Analytics Module for Hadoop jobs.

   a. Click **MapReduce2** > **Configs** > **Advanced** > **Advanced mapred-site**.

   b. In the `mapreduce.application.classpath` field, add the `/usr/hdp/current/hadoop-client/*`.

      Note: The separator is a colon (:) for the `mapreduce.application.classpath` field.

2. Update the classpath of YARN for using the NetApp In-Place Analytics Module for Hadoop jobs.

   a. Click **YARN** > **Configs** > **Advanced** > **Advanced yarn-site**.

   b. In the `yarn.application.classpath` field, add the `/usr/hdp/current/hadoop-client/*`.

Note: The separator is a comma (,) for the `yarn.application.classpath` field.

After you finish

1. Click **Services > Restart All Required**.
   It is required to restart all of the affected components, such as HDFS, YARN, and MapReduce, before you run any Hadoop jobs.
2. Click **NetApp In-Place Analytics Module > Service Actions > Run Service Check**.
   You must verify the installation by running a service check and resolve errors, if any.

Creating the JSON configuration file

After configuring ONTAP and Hadoop clusters for the NetApp In-Place Analytics Module, you must create a JSON configuration file and then distribute the file to all the nodes on your Hadoop cluster. The file should be accessible to all the Hadoop users in the system.

Steps

1. Create the `nfs-mapping.json` file.

   You can create the file in any folder based on your environment. For example, you can create a folder `/etc/NetAppNFSConnector/conf` in all the Hadoop members.

2. Set up read permissions for all the Hadoop users in the system:

   `chmod 755 nfs-mapping.json`

3. Set up the parameters for the JSON file.

```
{
        "spaces": [
        {
            "endpoints": [
                {
                  "path":"",
                  "exportPath":"",
                  "hosts": [""],
                   "nfsKerberosSpn":""
                   }
            ],
            "name": "",
            "options": {
                "nfsAuthScheme": "",
                "nfsGssServiceScheme" : "",
                "nfsDefaultKerberosSpn" : ""
                "nfsExportPath": "",
                "nfsMountPort": ,
                "nfsPort": ,
                "nfsReadSizeBits": ,
                "nfsReadDirPlusCountBytes": ,
                "nfsReceiveTCPBufferSizeBytes": ,
                "nfsSendTCPBufferSizeBytes": ,
                "nfsReadMinPoolThreads": ,
                "nfsReadMaxPoolThreads": ,
                "nfsWriteMinPoolThreads":  ,
                "nfsWriteMaxPoolThreads":,
                "nfsRpcbindPort": ,
                "nfsSplitSizeBits": ,
                "nfsWriteSizeBits": ,
                "nfsRpcTimeout_ms": ,
                "nfsIsRangerEnabled":,
                "nfsRangerConfig": {
                   "nfsRangerServiceName":"",
                   "nfsRangerAdminUrl":"",
                   "nfsRangerAuthDefaultAllow": ,
                   "nfsRangerAuditLogPath" : "",
                   "nfsRangerAuditEnabled" :
                }
            },
            "uri": ""
```

```
        }

        ]
}
```

Parameters for the JSON configuration file on page 12

Example JSON configuration file

The following example shows a sample JSON file without Ranger details.

```
{
        "spaces": [
        {
            "endpoints": [
                {
                  "path":"/",
                  "exportPath":"/vol1",
                  "hosts": ["nfs://10.231.43.115:2049/"],
                   "nfsKerberosSpn":"nfs/hadooplif.netapp.com@NETAPP.COM"
                   }
            ],
            "name": "demo",
            "options": {
                "nfsAuthScheme": "RPCSEC_GSS",
                "nfsGssServiceScheme" : "RPCSEC_GSS_KRB5",
                "nfsDefaultKerberosSpn" : "nfs/svm_nfs_lif1.example.com@EXAMPLE.COM",
                "nfsExportPath": "/",
                "nfsMountPort": -1,
                "nfsPort": 2049,
                "nfsReadSizeBits": 16,
                "nfsReadDirPlusCountBytes": 7680,
                "nfsReceiveTCPBufferSizeBytes": 1048576,
                "nfsSendTCPBufferSizeBytes": 1048576,
                "nfsReadMinPoolThreads":64,
                "nfsReadMaxPoolThreads":128,
                "nfsWriteMinPoolThreads":64,
                "nfsWriteMaxPoolThreads":128,
                "nfsRpcbindPort": 111,
                "nfsSplitSizeBits": 16,
                "nfsWriteSizeBits": 16,
                "nfsRpcTimeout_ms": 120000,
                "nfsIsRangerEnabled":false,
                "nfsRangerConfig": {
                   "nfsRangerServiceName":"hdpmaster_nfsconnector_hadoop",
                   "nfsRangerAdminUrl":"http://10.141.46.222:6080/",
                   "nfsRangerAuthDefaultAllow":false,
                   "nfsRangerAuditLogPath" : "/root/RANGER_AUDIT_NFS",
                   "nfsRangerAuditEnabled" :false
                }
            },
            "uri": "nfs://10.231.43.115:2049/"
        }

        ]
}
```

Parameters for the JSON configuration file

You must be aware of the global, NFS, Ranger, and Kerberos parameters for `nfs-mapping.json` file.

Storage system details

Parameter	Description	Default value	Example value
`name`	Name for the SVM to which you are connecting from the Hadoop servers	NA	hadoop_cluster

Parameter	Description	Default value	Example value
`uri`	Primary data LIF of the SVM that is mentioned in the `name` field **Note:** You can access the NFS file system only with the data LIF that is specified in the `uri` parameter. Accessing the NFS file system with any other data LIF in the SVM is not supported.	NA	`nfs://192.168.120.1:2049/` or `nfs://hadoop_cluster:2049/`

The `name` and `uri` are mandatory parameters.

Endpoint configuration

Parameter	Description	Default value	Example value
`path`	Alias name given to the local mount path in a Hadoop node for accessing the NFS file system that is specified in the `exportPath` parameter		`/nfs_volume_name`
`exportPath`	Junction path of the NFS volume that is exported from the SVM This path name should either be the same as the volume name or the same as the subfolders of the `nfsExportPath` value. You must ensure that the exported volume or folder exists before using it in the NetApp In-Place Analytics Module configuration. This parameter is mandatory.	`/`	`/nfs_volume_name`
`hosts`	List of data LIFs (IP addresses or host names) that are configured to access the NFS volume junction path specified in the `exportPath` field that is tagged under an endpoint. This parameter is mandatory.	NA	`["nfs://192.168.1.1:2049/","nfs://192.168.1.2:2049/"]`

You can configure multiple endpoints. However, you must ensure that all endpoints are configured with unique values.

For example, if the `nfs-mapping.json` endpoint is configured with the following values:

- `path`: **`/local_path`**
- `exportPath`: **`/volume1`**
- `hosts`: 192.168.1.1:2049

The contents of the **`/volume1`** NFS volume are listed when you run the following command:

```
hadoop fs -ls nfs://192.168.1.1:2049/:2049/local_path
```

Note: If the `path` parameter is specified in an endpoint configuration, the `path` parameter value is used to access the files and folders from the NFS export that is specified in the `exportPath`

parameter. If the `path` parameter is not specified in the endpoint configuration, the files and folders are accessed from the NFS export that is specified in the `nfsExportPath` parameter.

NFS configuration parameters

Parameter	Description	Default value	Example value
`nfsAuthScheme`	Authentication type of the NFS request Valid values are **`AUTH_SYS`**, **`AUTH_NONE`**, and **`RPSEC_GSS`**. Use **`RPSEC_GSS`** for Kerberos authentication. See other parameters for *Kerberos configuration*. This is a mandatory parameter.	**`AUTH_SYS`**	**`AUTH_NONE`**
`nfsExportPath`	Absolute junction path for a specific volume in the SVM This export path is used if the `path` parameter is not specified in the endpoint configuration. **Note:** It can be considered as the default export path or junction path for an SVM. It overrides the `exportPath` parameter that is configured for all endpoints under a namespace if the ***`path`*** specified in the Hadoop command **`nfs://`*`ip`*`:2049/`*`path`*** does not match any value specified for the `path` parameter in all of the available endpoints in the corresponding namespace.	/	`/hadoop`
`nfsReadSizeBits`	Size (bits) of the NFS read request It is the read size per request and can be configured up to a maximum value that the NFS server supports.	16	20
`nfsWriteSizeBits`	The size (bits) of the NFS write requests. It is the write size per request and can be configured up to a maximum value that NFS server supports.	16	20

Parameter	Description	Default value	Example value
`nfsReadDirPlusCountBytes`	dircount parameter as per specification for NFS protocol under operation READIRPLUS request Configuring a higher value improves performance for directory listing on NFS volumes. The values depend on the maximum read I/O size that is supported by the NFS server. It is recommended to configure this parameter in such a way that 8 times the `nfsReadDirPlusCountBytes` parameter value is within the limit of the maximum read I/O size.	7680	7680
`nfsReceiveTCPBufferSizeBytes`	Networking module (Java Netty library package) receive buffer size It is not the actual socket buffer, but the buffer maintained by the Netty library specific to the TCP socket for receiving packets.	1048576 (1 MB)	1048576
`nfsSendTCPBufferSizeBytes`	Networking module (Java Netty package) send buffer size It is not the actual socket buffer, but a buffer maintained by the Netty library specific to the TCP socket for sending packets.	1048576 (1 MB)	1048576
`nfsSplitSizeBits`	The size (bits) of the input split used. This is used to determine the number of blocks to read/prefetch. Number of read blocks per split = Minimum of file length and “nfsSplitSizeBits” divided by “readBlockSizeBits”.	20	28
`nfsPort`	Port used for NFS connections. This parameter is used if the PortMapper service is not running.	2049	2049
`nfsMountPort`	Port used for mount in case the PortMapper is not running	-1	12345
`nfsRpcbindPort`	Port used for the PortMapper connections	111	111
`nfsUserCacheTimeout`	Timeout (in minutes) after which the user and group information cache is refreshed. This advanced parameter is optional.	30	15
`nfsReadMinPoolThreads`	Minimum read threads for I/O operations on TCP channels	32	32
`nfsReadMaxPoolThreads`	Maximum read threads for I/O operations on TCP channels	128	128

Parameter	Description	Default value	Example value
`nfsWriteMinPoolThreads`	Minimum write threads for I/O operations on TCP channels	32	32
`nfsWriteMaxPoolThreads`	Maximum write threads for I/O operations on TCP channels	128	128
`nfsRpcTimeout_ms`	Timeout value in milliseconds of each NFS operation to the NetApp storage system	120000	120000

Ranger configuration

Parameter	Description	Default value	Example value
`nfsIsRangerEnabled`	Enable Ranger authorization support for NFS	**`false`**	**`true`**
`nfsRangerServiceName`	The HDFS service name to be used per NFS namespace that is configured in the Ranger admin server under service manager for resource-based policies for the HDFS component. Only one service name for HDFS is supported. You can create multiple policies under one service name.	NA	`hdpmaster_nfsconnector_hadoop`
`nfsRangerAdminUrl`	Ranger admin URL to import the policies to NFS client instance	NA	`http://10.141.46.9:6080/`
`nfsRangerAuditLogPath`	Path where the audit logs for this module are stored You must create the folder in all the Hadoop members of the cluster.	NA	`/usr/nfs/ranger/logs`
`nfsRangerAuditEnabled`	Enable audit logging for file system API operation calls	**`false`**	**`true`**
`nfsRangerAuthDefaultAllow`	Determines whether all users who do not have Ranger policies configured have permissions to access the file system	**`false`**	**`true`**

Kerberos configuration

The `nfsAuthScheme` parameter must be set to **`RPSEC_GSS`** for Kerberos authentication.

Parameter	Description	Default value	Example value
`nfsKerberosSpn`	Kerberos service principal name (SPN) It is the host name of the NFS server in the Kerberos realm. This is a mandatory parameter if Kerberos is configured.	NA	`nfs/hadooplif.netapp.com@NETAPP.COM`

Parameter	Description	Default value	Example value
`nfsDefaultKerberosSpn`	SPN for NFS service that is used when no matching endpoint is found in the configuration This is a mandatory parameter if Kerberos is configured. **Note:** This parameter can be considered as the default SPN for the default export path (`nfsExportPath`) in the configuration.		`"nfsDefaultKerberosSpn" : "nfs/svm_nfs_lif1.example.com@EXAMPLE.COM",`
`nfsGssServiceScheme`	Kerberos flavor for authentication Valid values are the following: • **`RPSEC_GSS_KRB5`**: For **`krb5`** that allows authenticated users to access data that is neither integrity-protected nor encrypted. • **`RPSEC_GSS_KRB5i`**: For **`krb5i`** that allows authenticated users access data that is only integrity-protected. • **`RPSEC_GSS_KRB5p`**: For **`krb5p`** that allows authenticated users to access data that is both integrity-protected and encrypted. This is a mandatory parameter if Kerberos is configured.	NA	**`RPSEC_GSS_KRB5`**

Verifying the installation of NetApp In-Place Analytics Module

You should run any Hadoop job to check the installation of the NetApp In-Place Analytics Module.

Steps

1. Verify the installation of the NetApp In-Place Analytics Module: `hadoop fs`

Example

```
[root@node1 ~]#hadoop fs -ls nfs://10.63.150.213:2049/
Found 1 items
drwxrwxrwx - root root 4096 2018-05-11 04:31 nfs://
10.63.150.213:2049/.snapshot
[root@node1 ~]#
[root@node1 ~]# hadoop jar /usr/hdp/2.5.3.0-37/hadoop-mapreduce/
hadoop-mapreduce-examples.jar teragen 10000 nfs://
10.63.150.213:2049/tg
[root@node1 ~]# hadoop jar /usr/hdp/2.5.3.0-37/hadoop-mapreduce/
hadoop-mapreduce-examples.jar terasort nfs://10.63.150.213:2049/tg
nfs://10.63.150.213:2049/ts
[root@node1 ~]# hadoop jar /usr/hdp/2.5.3.0-37/hadoop-mapreduce/
hadoop-mapreduce-examples.jar teravalidate nfs://
10.63.150.213:2049/ts nfs://10.63.150.213:2049/tv
[root@node1 ~]# hadoop fs -ls nfs://10.63.150.213:2049/
Found 4 items
```

```
drwxrwxrwx - root root 4096 2018-05-11 04:31 nfs://
10.63.150.213:2049/.snapshot
drwxrwxrwx - root root 4096 2018-05-11 06:28 nfs://
10.63.150.213:2049/tg
drwxrwxrwx - root root 4096 2018-05-11 06:29 nfs://
10.63.150.213:2049/ts
drwxrwxrwx - root root 4096 2018-05-11 06:30 nfs://
10.63.150.213:2049/tv
[root@node1 ~]#
```

2. Copy a local file by using NFS.

Example

```
[root@hdp1 nc_volumev3]# touch hello_world.txt

[root@hdp1 nc_volumev3]# echo "Hello World" >> hello_world.txt

[root@hdp1 nc_volumev3]# cat hello_world.txt
Hello World

[root@hdp1 nc_volumev3]# hadoop fs -copyFromLocal hello_world.txt
nfs://10.63.150.213:2049//nc_volumev3/test/

[root@hdp1 nc_volumev3]# hadoop fs -ls nfs://10.63.150.213:2049/
nc_volumev3/test/
Found 6 items
-rw-r--r-- 1 hdfs hadoop 1145 2018-02-22 12:15 nfs://
10.63.150.213:2049/nc_volumev3/test/README
drwxrwxrwx - root root 4096 2018-02-23 06:28 nfs://10.63.150.213:2049/
nc_volumev3/test/abc
-rw-r--r-- 1 root root 12 2018-02-24 17:43 nfs://10.63.150.213:2049/
nc_volumev3/test/hello_world.txt
drwxr-xr-x - root root 4096 2018-02-23 05:20 nfs://10.63.150.213:2049/
nc_volumev3/test/subfolderinexportPath
drwxrwxrwx - hdfs hadoop 4096 2018-02-22 09:32 nfs://
10.63.150.213:2049/nc_volumev3/test/tg_absolutepath_exportPath_path
drwxrwxrwx - root root 4096 2018-02-23 06:15 nfs://10.63.150.213:2049/
nc_volumev3/test/tg_in_exportPath

[root@hdp1 nc_volumev3]# hadoop fs -cat nfs://10.63.150.213:2049/
nc_volumev3/test/hello_world.txt
Hello World
[root@hdp1 nc_volumev3]#
```

After you finish

For validating NetApp In-Place Analytics Module with various Hadoop components, see *NetApp Technical Report 4382: NetApp In-Place Analytics Module Best Practices*.

Uninstalling the NetApp In-Place Analytics Module

You can uninstall NetApp In-Place Analytics Module if you no longer need it. If you want to install a new version of the NetApp In-Place Analytics Module, you must uninstall the current version and then install the new version of the NetApp In-Place Analytics Module.

Steps

1. On each node of the Hortonworks cluster, delete all of the JAR files from the `/usr/hdp/current/hadoop-client/` directory.
2. On each node of the Hortonworks cluster, delete all of the JAR files from the library path of the Hadoop components that use NetApp In-Place Analytics module.

Example

For removing NetApp In-Place Analytics Module from the YARN classpath, delete all of the JAR files from the `usr/hdp/`*`hdp_version`*`/hadoop-yarn/lib/` directory.

Related tasks

Installing NetApp In-Place Analytics Module on a Hortonworks cluster on page 9

Copyright

Trademark

How to send comments about documentation and receive update notifications

You can help us to improve the quality of our documentation by sending us your feedback. You can receive automatic notification when production-level (GA/FCS) documentation is initially released or important changes are made to existing production-level documents.

If you have suggestions for improving this document, send us your comments by email.

doccomments@netapp.com

To help us direct your comments to the correct division, include in the subject line the product name, version, and operating system.

If you want to be notified automatically when production-level documentation is released or important changes are made to existing production-level documents, follow Twitter account @NetAppDoc.

You can also contact us in the following ways:

- NetApp, Inc., 1395 Crossman Ave., Sunnyvale, CA 94089 U.S.
- Telephone: +1 (408) 822-6000
- Fax: +1 (408) 822-4501
- Support telephone: +1 (888) 463-8277

Index

C

comments
  how to send feedback about documentation *22*

D

documentation
  how to receive automatic notification of changes to *22*
  how to send feedback about *22*

F

feedback
  how to send comments about documentation *22*

I

information
  how to send feedback about improving documentation *22*

N

NetApp In-Place Analytics Module
  configuring ONTAP *7*
  creating the JSON configuration file *11*
  installing *6*
  installing on Hortonworks cluster *9*
  Kerberos *4*
  overview *4*
  parameters for NFS mapping JSON file *12*
  Ranger *4*
  specifying NFS as the scheme for the Hadoop file system *8*
  system requirements for installing *6*
  uninstalling *19*
  updating classpath of Hadoop components *10*
  verifying the installation *17*

S

suggestions
  how to send feedback about documentation *22*

T

Twitter
  how to receive automatic notification of documentation changes *22*