

Risoluzione dei problemi relativi a Hyperflex Storage NFS all paths down (APD)

Sommario

[Introduzione](#)

[Come vengono montati gli archivi dati HX su ESXi?](#)

[Tutti i percorsi in basso](#)

[Descrizione del problema](#)

[Flusso di lavoro di risoluzione dei problemi](#)

[Archivia vCenter Server:](#)

[Archivia tutte le macchine virtuali StCtl:](#)

[StCtlVM: StCtlVM di un host ESXi interessato](#)

[Archiviazione host ESXi:](#)

Introduzione

In questo documento viene fornita una rapida comprensione delle operazioni di risoluzione dei problemi che è possibile eseguire per valutare l'origine del problema se viene visualizzato il messaggio di errore "NFS tutti i percorsi inattivi" in vCenter a cui è integrato il cluster Hyperflex.

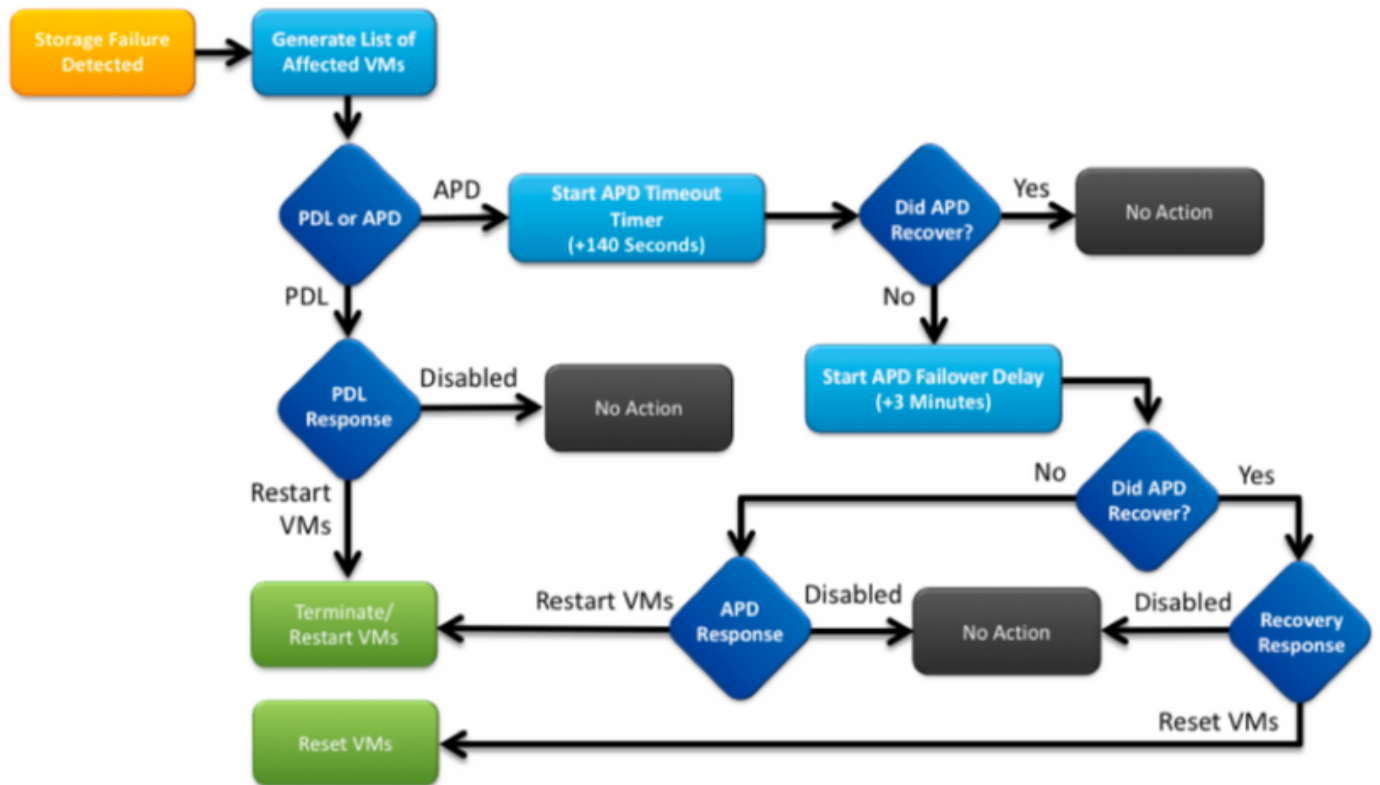
Come vengono montati gli archivi dati HX su ESXi?

- Gli archivi dati Hyperflex vengono installati sugli host ESXi durante l'installazione di NFS. Per poter installare un archivio dati NFS è necessario l'IP del server NFS, che nel nostro caso è l'interfaccia mobile virtuale eth1:0.
- Il cluster Hyperflex sfrutta l'uso di **IP mobili** virtuali sia per la gestione (eth0:mgmtip) che per i **dati di storage (eth1:0)** su cui ogni IP verrà assegnato a una VM del controller di storage (**StCtlVM**). Si noti che potrebbero finire in diverse StCtlVM.
- L'importanza di questo è che l'**IP dei dati di archiviazione** del cluster (**eth1:0**) è **quello utilizzato per montare gli archivi dati** creati nel cluster Hyperflex. È quindi essenziale che sia **assegnato e raggiungibile** da tutti i **nodi** del cluster.
- Notare che in caso di guasto della StCtlVM che attualmente possiede l'IP virtuale eth1:0, deve "migrare" a un'altra StCtlVM disponibile che funziona in modo simile a un FHRP (First Hop Redundancy Protocol).

Tutti i percorsi in basso

- APD significa che l'host non può raggiungere lo storage e non è stato restituito alcun codice SCSI PDL (Permanent Device Lost) dall'array di storage.
- Poiché non sa se la perdita è temporanea o meno, continuerà a tentare di stabilire la comunicazione per più di 140 secondi per impostazione predefinita (timeout) + 3 minuti (ritardo per il failover). L'host ESXi inizia a bloccare qualsiasi traffico di I/O di macchine non virtuali inviato al dispositivo di storage.

- Per ulteriori dettagli sull'APD in ambiente vSphere 6.x, fare riferimento all'articolo della Knowledge Base di VMware [PDL \(Permanent Device Loss\)](#) e [APD \(All-Paths-Down\) in vSphere 5.x e 6.x](#)
- Nell'immagine seguente viene illustrata ogni fase intermittente:



Di seguito è riportato un tipico messaggio di errore in vCenter.

	Status	Name	Defined In
.02.corplex....	Alert	NFS all paths down	SV-VC

Descrizione del problema

Una volta visualizzati gli avvisi APD sugli host, ottenere le seguenti informazioni per comprendere meglio la descrizione del problema:

- Se sono stati interessati uno/più/tutti gli host e se sono stati interessati alcuni host specifici
- Eventuali modifiche apportate in precedenza (configurazione/aggiornamento/ecc.)
- Timestamp del momento in cui il problema è stato rilevato per la prima volta e se il problema è ricorrente

Flusso di lavoro di risoluzione dei problemi

Per risolvere i problemi di APD, è necessario esaminare tre componenti: vCenter, SCVM ed host ESXi.

Questi passaggi rappresentano un flusso di lavoro consigliato per individuare o restringere

l'origine del sintomo Tutti i percorsi inattivi osservato. Si noti che questo ordine non deve essere seguito meticolosamente e può essere adeguato secondo i sintomi particolari osservati sull'ambiente del cliente.

Archivia vCenter Server:

Connettersi a vCenter Server (VCS) e passare a un **host interessato**

1. **Oggetti correlati -> Macchine virtuali** e conferma che StCtIVM è attivo e in esecuzione
2. **Oggetti correlati -> Datastore** e verificare se i datastore NFS mostrano "**inaccessibile**". Se gli archivi dati sembrano **accessibili** e lo stato, è possibile provare a ripristinare l'evento APD nella scheda **Riepilogo** e verificare in seguito se l'avviso viene visualizzato in un secondo momento
3. **Monitor -> Problemi** e **Monitor -> Eventi** deve fornire informazioni su quando l'APD è stato individuato per la prima volta.

Archivia tutte le macchine virtuali StCtI:

Connettersi a **tutte** le **StCtIVM** e verificare i puntatori seguenti, è possibile utilizzare [MobaXterm](#) software.

1. Verificare se tutte le StCtIVM hanno la stessa ora utilizzando **date** o **ntpq -p**. L'asimmetria temporale su StCtIVM può causare problemi con la sincronizzazione del database di zookeeper ed è quindi fondamentale che sia sincronizzato tra tutte le StCtIVM. Il segnale di errore del server ntp indica che l'NTP dello SCVM è sincronizzato.

```
root@SpringpathControllerPZTMTRSH7K:~# date
Tue May 28 12:47:27 PDT 2019
```

```
root@SpringpathControllerPZTMTRSH7K:~# ntpq -p -4
remote refid st t when poll reach delay offset jitter
=====
*abcdefghijkl.GNSS. 1 u 429 1024 377 225.813 -1.436 0.176
```

2. Se l'APD si è verificato durante un **aggiornamento**, è possibile verificare quali **StCtIVM non sono state completamente aggiornate** e, in particolare, identificare l'ultima operazione non riuscita. È possibile che sia stato lui a tenere in mano l'eth1:0 Utilizzare **dpkg -l | grep -i springpath** per identificare le StCtIVM non completamente aggiornate, in quanto disporranno di pacchetti Springpath con versioni miste.

```
root@SpringpathControllerPZTMTRSH7K:~# dpkg -l | grep -i springpath
ii storfs-appliance 4.0.1a-33028 amd64 Springpath Appliance
ii storfs-asup 4.0.1a-33028 amd64 Springpath ASUP and SCH
ii storfs-core 4.0.1a-33028 amd64 Springpath Distributed Filesystem
ii storfs-fw 4.0.1a-33028 amd64 Springpath Appliance
ii storfs-mgmt 4.0.1a-33028 amd64 Springpath Management Software
ii storfs-mgmt-cli 4.0.1a-33028 amd64 Springpath Management Software
ii storfs-mgmt-hypervcli 4.0.1a-33028 amd64 Springpath Management Software
ii storfs-mgmt-ui 4.0.1a-33028 amd64 Springpath Management UI Module
ii storfs-mgmt-vcplugin 4.0.1a-33028 amd64 Springpath Management UI and vCenter Plugin
ii storfs-misc 4.0.1a-33028 amd64 Springpath Configuration
ii storfs-pam 4.0.1a-33028 amd64 Springpath PAM related modules
ii storfs-replication-services 4.0.1a-33028 amd64 Springpath Replication Services
ii storfs-restapi 4.0.1a-33028 amd64 Springpath REST Api's
ii storfs-robo 4.0.1a-33028 amd64 Springpath Appliance
ii storfs-support 4.0.1a-33028 amd64 Springpath Support
```

ii storfs-translations 4.0.1a-33028 amd64 Springpath Translations

3. Verificare che tutti i servizi pertinenti siano *in esecuzione service_status.sh*: Alcuni dei servizi principali sono Springpath File System (*storfs*), SCVM Client (*scvmclient*), System Management Service (*stMgr*) o Cluster IP Monitor (*cip-monitor*).

```
root@SpringpathController5L0GTCR8SA:~# service_status.sh
Springpath File System ... Running
SCVM Client ... Running
System Management Service ... Running
HyperFlex Connect Server ... Running
HyperFlex Platform Agnostic Service ... Running
HyperFlex HyperV Service ... Not Running
HyperFlex Connect WebSocket Server ... Running
Platform Service ... Running
Replication Services ... Running
Data Service ... Running
Cluster IP Monitor ... Running
Replication Cluster IP Monitor ... Running
Single Sign On Manager ... Running
Stats Cache Service ... Running
Stats Aggregator Service ... Running
Stats Listener Service ... Running
Cluster Manager Service ... Running
Self Encrypting Drives Service ... Not Running
Event Listener Service ... Running
HX Device Connector ... Running
Web Server ... Running
Reverse Proxy Server ... Running
Job Scheduler ... Running
DNS and Name Server Service ... Running
Stats Web Server ... Running
```

4. Se uno di questi o altri servizi rilevanti non è attivo, avviarlo *utilizzando start <nomeServizio>* eg: *start storfs* È possibile fare riferimento allo script *service_status.sh* per ottenere i nomi dei servizi. Eseguire un'operazione *head -n25 /bin/service_status.sh* e identificare il nome reale del servizio.

```
root@SpringpathController5L0GTCR8SA:~# head -n25 /bin/service_status.sh
#!/bin/bash
declare -a upstart_services=("Springpath File System:storfs"\
"SCVM Client:scvmclient"\
"System Management Service:stMgr"\
"HyperFlex Connect Server:hxmanager"\
"HyperFlex Platform Agnostic Service:hxSvcMgr"\
"HyperFlex HyperV Service:hxHyperVSvcMgr"\
"HyperFlex Connect WebSocket Server:zkupdates"\
"Platform Service:stNodeMgr"\
"Replication Services:replsvc"\
"Data Service:stDataSvcMgr"\
"Cluster IP Monitor:cip-monitor"\
"Replication Cluster IP Monitor:repl-cip-monitor"\
"Single Sign On Manager:stSSOMgr"\
"Stats Cache Service:carbon-cache"\
"Stats Aggregator Service:carbon-aggregator"\
"Stats Listener Service:statsd"\
"Cluster Manager Service:exhibitor"\
"Self Encrypting Drives Service:sedsvc"\
"Event Listener Service:storfsevents"\
"HX Device Connector:hx_device_connector");
declare -a other_services=("Web Server:tomcat8"\
"Reverse Proxy Server:nginx"\
"Job Scheduler:cron"\
"DNS and Name Server Service:resolvconf");
```

5. Identificare quale *StCtlVM* contiene l'IP (eth1:0) del cluster di archiviazione utilizzando

ifconfig -a Se nessuna StCtlVM contiene tale IP, è possibile che storfs non sia in esecuzione su uno o più nodi.

```
root@help:~# ifconfig
eth0:mgmtip Link encap:Ethernet HWaddr 00:50:56:8b:4c:90
inet addr:10.197.252.83 Bcast:10.197.252.95 Mask:255.255.255.224
UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1
```

6. Verificare se StCtlVM è in contatto con CRMaster e se il servizio zookeeper è attivo e in esecuzione **echo server | nc localhost 2181** e verificare se la modalità è Leader, Follower o Standalone e se le connessioni sono > 0

```
root@help:~# echo srvr | nc localhost 2181
Zookeeper version: 3.4.12-d708c3f034468a4da767791110332281e04cf6af, built on 11/19/2018
21:16 GMT
Latency min/avg/max: 0/0/137
Received: 229740587
Sent: 229758548
Connections: 13
Outstanding: 0
Zxid: 0x140000526c
Mode: leader
Node count: 3577
```

stato service exitor o ps -ef | grep -i espositore

```
root@help:~# service exhibitor status
exhibitor start/running, process 12519
root@help:~# ps -ef | grep -i exhibitor
root 9765 9458 0 13:19 pts/14 00:00:00 grep --color=auto -i exhibitor
root 12519 1 0 May19 ? 00:05:49 exhibitor
```

In caso di errori o di servizi non in esecuzione, è possibile verificare i registri seguenti e provare ad avviare il servizio zookeeper **/var/log/springpath/exhibitor.log** e **/var/log/springpath/stMgr.log** **espositore del servizio avviare il servizio zookeeper**

7. Verifica se il VC è raggiungibile da tutte le StCtlVM informazioni sul cluster **stcli | grep -i "url"** per visualizzare l'URL utilizzato contenente FQDN o IP di VC. Verificare la connettività a VC utilizzando **ping <VC>**

```
root@help:~# stcli cluster info | grep -i "url"
vCenterUrl: https://10.197.252.101
vCenterURL: 10.197.252.101
root@help:~# ping 10.197.252.101
PING 10.197.252.101 (10.197.252.101) 56(84) bytes of data.
64 bytes from 10.197.252.101: icmp_seq=1 ttl=64 time=0.435 ms
```

8. Conferma se il DNS è raggiungibile in caso di cluster che utilizzano il nome di dominio completo **I DNS dei servizi stcli mostrano** come elencare i server configurati DNS su StCtlVM. Verificare la connettività e la risoluzione dei server DNS utilizzando **ping <DNS_IP>** e **host <FQDN> <DNS_IP>**

```
root@help:~# stcli services dns show
1.1.128.140
root@help:~# ping 1.1.128.140
PING 1.1.128.140 (1.1.128.140) 56(84) bytes of data.
64 bytes from 1.1.128.140: icmp_seq=1 ttl=244 time=1.82 ms
```

9. Confermare se tutti gli oggetti StCtlVM hanno la stessa quantità di voci iptable: **iptables -L | wc -l** In caso di mancata corrispondenza, aprire una richiesta TAC.

```
root@SpringpathControllerI51U7U6QZX:~# iptables -L | wc -l
48
```

10. Stato e integrità del cluster **informazioni sul cluster stcli | meno** o **informazioni sul cluster stcli | grep -i "active\|state\|unavailable"** se si cerca di individuare determinati nodi con spazio di archiviazione non disponibile.

```
root@SpringpathControllerI51U7U6QZX:~# stcli cluster info | grep -i
"active\|state\|unavailable"
locale: English (United States)
```

```
state: online
upgradeState: ok
healthState: healthy
state: online
state: 1
activeNodes: 3
state: online
```

riepilogo archiviazione cluster stcli - dettaglio per ottenere i dettagli del cluster di memoria

```
root@SpringpathControllerI51U7U6QZX:~# stcli cluster storage-summary --detail
address: 10.197.252.106
name: HX-Demo
state: online
uptime: 185 days 12 hours 48 minutes 42 seconds
activeNodes: 3 of 3
compressionSavings: 85.45%
deduplicationSavings: 0.0%
freeCapacity: 4.9T
healingInfo:
inProgress: False
resiliencyDetails:
current ensemble size:3
# of caching failures before cluster shuts down:3
minimum cache copies remaining:3
minimum data copies available for some user data:3
minimum metadata copies available for cluster metadata:3
# of unavailable nodes:0
# of nodes failure tolerable for cluster to be available:1
health state reason:storage cluster is healthy.
# of node failures before cluster shuts down:3
# of node failures before cluster goes into readonly:3
# of persistent devices failures tolerable for cluster to be available:2
# of node failures before cluster goes to enospace warn trying to move the existing
data:na
# of persistent devices failures before cluster shuts down:3
# of persistent devices failures before cluster goes into readonly:3
# of caching failures before cluster goes into readonly:na
# of caching devices failures tolerable for cluster to be available:2
resiliencyInfo:
messages:
Storage cluster is healthy.
state: 1
nodeFailuresTolerable: 1
cachingDeviceFailuresTolerable: 2
persistentDeviceFailuresTolerable: 2
zoneResInfoList: None
spaceStatus: normal
totalCapacity: 5.0T
totalSavings: 85.45%
usedCapacity: 85.3G
zkHealth: online
clusterAccessPolicy: lenient
dataReplicationCompliance: compliant
dataReplicationFactor: 3
```

11. Quali datastore sono installati e disponibili

```
root@bsv-hxaf220m5-sc-4-3:~# stcli datastore list
-----
virtDatastore:
  status:
    EntityRef(idtype=None, confignum=None, type=6, id='235ea35f-6c85-9448-bec7-
06f03b5adf16', name='bsv-hxaf220m5-hv-4-3.cisco.com'):
      accessible: True
      mounted: True
    EntityRef(idtype=None, confignum=None, type=6, id='d124203c-3d9a-ba40-a229-
```

```

4dffbe96ae13', name='bsv-hxaf220m5-hv-4-2.cisco.com'):
    accessible: True
    mounted: True
EntityRef(idtype=None, confignum=None, type=6, id='e85f1980-b3c7-a440-9f1e-
20d7a1110ae6', name='bsv-hxaf220m5-hv-4-1.cisco.com'):
    accessible: True
    mounted: True

```

12. Se i comandi stcli richiedono troppo tempo o hanno esito negativo, provare i seguenti comandi sysmtool (da non utilizzare se stcli funziona) **sysmtool —cluster ns —informazioni sul comando sysmtool —ns cluster —cmd health detail sysmtool —ns datastore —elenco di comandi**

StCtIVM: StCtIVM di un host ESXi interessato

Connettersi a **StCtIVM** dell'host **ESXi** interessato

1. Verificare la connettività all'IP del cluster di storage (eth1:0) e ad altri server sulla rete di storage (eth1 su StCtIVM)
Esegui **informazioni cluster stcli | grep -i -B 1 "stctl|hypervisor"** per identificare tutti gli IP di gestione ESXi, StCtIVM eth0 (Mgmt) e StCtIVM eth1 (dati di storage) che partecipano rispettivamente al cluster. Verificare la connettività **ping -I eth1 [-M do -s 8972] <indirizzo IP target>**, verificare i frame jumbo tra ESXi VMK1 e SCVMeth1.
2. Se il problema persiste, è possibile esaminare i seguenti registri **/var/log/springpath/debug-storfs.log**. Verificare se si verificano errori di sistema, errori di segmento o eventi critici **grep -ai "segmentation|critical|panic" debug-storfs.log/var/log/springpath/stmgr.log**. Verificare se la memoria insufficiente presenta il problema **grep -i "Room\|out of mem" /var/log/kern.log**
3. In ultima analisi, è possibile provare a riavviare la **StCtIVM** del nodo che presenta ancora il problema e verificare se il problema persiste.

Archiviazione host ESXi:

Collegarsi a un host **ESXi** interessato tramite **SSH** ed eseguire le seguenti azioni:

1. **elenco nfs storage esxcli oEsxcfg-nas -l** per elencare gli archivi dati NFS attualmente montati e se sono accessibili

```

[root@bsv-hx220m5-hv-4-3:~] esxcli storage nfs list
Volume Name  Host                               Share                               Accessible
Mounted Read-Only  isPE Hardware Acceleration
-----
test          8352040391320713352-8294044827248719091  192.168.4.1:test                   true
true         false false Supported
sradzevi     8352040391320713352-8294044827248719091  192.168.4.1:sradzevi              true
true         false false Supported
[root@bsv-hx220m5-hv-4-3:~] esxcfg-nas -l
test is 192.168.4.1:test from 8352040391320713352-8294044827248719091 mounted available
sradzevi is 192.168.4.1:sradzevi from 8352040391320713352-8294044827248719091 mounted available

```

È inoltre possibile confermare da **/etc/vmware/esx.conf** la coerenza della configurazione ESXi sugli archivi dati montati su NFS, utilizzando il comando **cat /etc/vmware/esx.conf | grep -l nas**

2. Verificare **/var/log/vmkernel.log** e cercare ad esempio stato non riuscito, problemi di montaggio o errori relativi al timestamp identificato nei passaggi precedenti

3. Verifica dello stato di IOVisor/NFS Proxy/SCVMClient Verificare se il servizio è in esecuzione su ESXi utilizzando il comando `/etc/init.d/scvmclient status` [Facoltativo] È possibile verificare se sono presenti connessioni aperte utilizzando l'*elenco delle connessioni di rete ip esxcli | grep -i "proto/client scvm"* Verificare che la versione di SCVMClient VIB corrisponda alla versione del cluster HX, *elenco delle versioni software Esxcli | grep -i primavera*

```
[root@bsv-hx220m5-hv-4-3:~] esxcli software vib list | grep -i spring
scvmclient                3.5.1a-31118                Springpath
VMwareAccepted            2018-12-13
stHypervisorSvc           3.5.1a-31118                Springpath
VMwareAccepted            2018-12-06
vmware-esx-STFSNasPlugin  1.0.1-21                    Springpath
VMwareAccepted            2018-11-16
```

Selezionare `/var/log/scvmclient.log` per verificare se sono presenti errori, ovvero "impossibile ottenere clustermap". Se necessario, è possibile riavviare il servizio SCVMClient tramite il comando `etc/init.d/scvmclient restart`

4. Verificare la connettività di rete con altri host ESXi sulla rete vmk1, in particolare per il cluster di storage IP eth1:0 `esxcfg-vmknic -l` per ottenere informazioni sui dettagli della nic del vmk, ad esempio IP, maschera e MTU `vmkping -l vmk1 [-v -s 8972] -d <indirizzo IP destinazione>` a impostare la connettività [facoltativa con frame jumbo] tra gli host ESXi sulla rete dati del controller
5. La piattaforma hardware esxcli consente di ottenere il numero di serie del server utilizzato sul nome della StCtIVm e di identificare rapidamente su quale host è in esecuzione una specifica StCtIVM.