

疑難排解Hyperflex儲存NFS所有路徑中斷(APD)問題

目錄

[簡介](#)

[HX資料儲存如何安裝在ESXI上？](#)

[所有關閉路徑](#)

[問題描述](#)

[故障排除工作流](#)

[簽入vCenter Server:](#)

[簽入所有StCtIVM:](#)

[StCtIVM:受影響的ESXi主機的StCtIVM](#)

[簽入ESXi主機：](#)

簡介

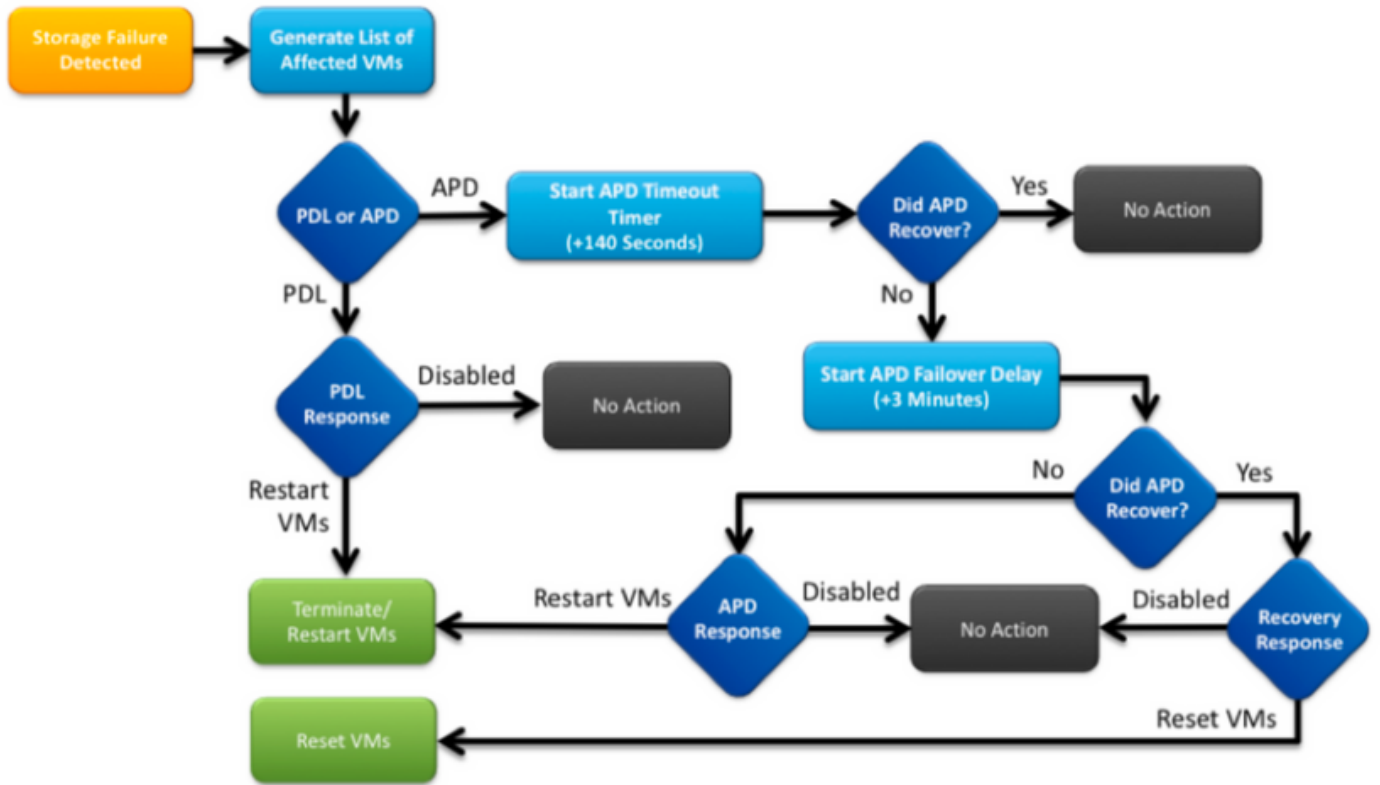
如果您在與整合Hyperflex群集的vCenter中看到「NFS all paths down」錯誤消息，則本文檔為您提供了可執行的快速瞭解和故障排除步驟，以便評估問題的根源。

HX資料儲存如何安裝在ESXI上？

- Hyperflex資料儲存區在NFS裝載時安裝在ESXI主機上，為了裝載NFS資料儲存區，我們需要使用NFS伺服器IP，在本例中是eth1:0虛擬浮動介面。
- Hyperflex群集利用虛擬浮動IP進行管理(eth0:mgmtip)和儲存資料(eth1:0)，每個IP都將分配給一個特定的儲存控制器VM(StCtIVM)。請注意，它們可能最終位於不同的StCtIVM中。
- 其重要性在於群集儲存資料IP(eth1:0)用於裝載在Hyperflex群集中建立的datastore。因此，必須為其分配並從集群的所有節點訪問。
- 請注意，如果當前擁有eth1:0虛擬IP的StCtIVM出現故障，它應將「遷移」到另一個可用的StCtIVM，其運行方式與FHRP（第一跳冗餘協定）類似。

所有關閉路徑

- APD表示主機無法訪問儲存，並且儲存陣列中沒有返回永久裝置丟失(PDL)SCSI代碼。
- 由於不知道丟失是否是暫時的，預設情況下，它會繼續嘗試建立超過140的通訊（超時）+ 3分鐘（故障切換延遲），因此ESXi主機開始使傳送到儲存裝置的所有非虛擬機器I/O流量出現故障。
- 有關vSphere 6.x環境中的APD的詳細資訊，請參閱Vmware知識庫文章[vSphere 5.x和6.x中的Permanent Device Loss\(PDL\)and All-Paths-Down\(APD\)](#)
- 下圖說明每個間歇階段：



vCenter中的典型錯誤消息如下。

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

問題描述

在主機上看到APD警報後，獲取以下資訊以更好地瞭解問題說明：

- 如果一部/多部/所有主機受到影響，以及如果某些特定主機受到影響
- 如果之前執行過任何更改（配置/升級等）
- 首次發現問題時的時間戳以及問題是否重複出現

故障排除 workflow

為了對APD進行故障排除，我們需要瞭解3個元件 — vCenter、SCVM和ESXi主機。

這些步驟是建議的工作流程，旨在查明或縮小出現的所有路徑關閉症狀的來源。請注意，此訂單不必嚴格遵守，您可以根據在客戶環境中觀察到的特定症狀來充分遵循此訂單。

簽入vCenter Server:

連線到vCenter Server(VCS)並導航到受影響的主機

1. **Related Objects -> Virtual Machines**並確認StCtIVM已啟動且正在運行
2. **Related Objects -> Datastore**並確認NFS資料儲存區是否顯示「**inaccessible**」。如果資料儲存似乎可以訪問並處於狀態，您可以在**Summary**頁籤上嘗試在APD事件中「重置為綠色」，稍後驗證警報是否再次彈出
3. **Monitor -> Issues**和**Monitor -> Events**應提供有關首次發現APD的時間資訊。

簽入所有StCtIVM:

連線到所有StCtIVM並驗證下面的指標，您可以使用 [MobaXterm](#) 軟體的路由器之間的 CLI 設定站點對站點 (LAN 對 LAN) IPsec 網際網路金鑰交換版本 1 (IKEv1) 通道。

1. 使用 **date**或**ntpq -p**驗證所有StCtIVM是否具有相同的時間。StCtIVM上的時間偏差可能導致縮放管理員資料庫同步問題，因此必須在所有StCtIVM之間保持同步。ntp server的strick sign infort表示您的SCVM的NTP已同步。

```
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. 如果在升級期間發生APD，則可以考慮驗證哪些StCtIVMs尚未完全升級，特別是識別最後發生故障的StCtIVM。可能是之前持有eth1:0的那個路由器 使用**dpkg -l | grep -i springpath**以標識未完全升級的StCtIVM，因為它們將具有混合版本的springpath包。

```
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. 驗證所有相關服務是否都**正在運行****service_status.sh**: 部分主要服務包括Springpath檔案系統 (storfs)、SCVM客戶端(scvmclient)、系統管理服務(stMgr)或**集群IP監控(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. 如果其中任何一項或其他相關服務未啟動，請使用 **start <serviceName>** 例如：**start storfs** 啟動。可以參考 `service_status.sh` 指令碼獲取服務名稱。執行 `head -n25 /bin/service_status.sh` 並標識服務的真實名稱。

```

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. 使用 `ifconfig -a` 識別哪個 StCtIVM 包含 storage cluster IP (eth1:0) 如果沒有 StCtIVM 包含該 IP，則可能儲存未在一個或多個節點上運行。

```

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. 驗證 StCtIVM 是否與 CRMMaster 接觸，以及 `zookeeper service` 是 up 和 `echo server | nc localhost 2181` 並檢查模式是否為「引線」、「跟隨器」或「獨立」，以及連線是否大於 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

```

服務展位狀態或 `ps -ef | grep -i 參展商`

```
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
```

如果有任何錯誤或服務未運行，您可以驗證以下日誌並嘗試啟動zookeeper服務

/var/log/springpath/exhibitor.log /var/log/springpath/stMgr.log服務展位器開始啟動zookeeper服務

7. 驗證是否可從所有StCtIVM訪問VC stcli群集資訊 | grep -i "url"顯示使用的URL，其中包含VC的FQDN或IP。使用**ping <VC>**檢驗與VC的連線

```
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. 確認在使用FQDN的群集情況下是否可訪問DNS **stcli services dns show**列出StCtIVM上配置了DNS的伺服器。使用**ping <DNS_IP>**和**host <FQDN> <DNS_IP>**，測試與DNS伺服器的連線性和解析

```
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. 確認所有StCtIVM是否具有相同的iptables條目：**iptables -L | wc -l**。如果兩者不相符，請開啟一個TAC案例。

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

10. 當前群集狀態和運行狀況 **stcli**群集資訊 | 更少或**stcli**群集資訊 | **grep -i "active|state|unavailable"**(如果嘗試查詢儲存不可用時顯示的特定節點)。

```
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

stcli cluster storage-summary --
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. 已掛載和可用的數據儲存數據

```

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. 如果stcli命令花費時間太長或失敗，您可以嘗試以下sysmtool命令（如果stcli工作則不要使用）***sysmtool —ns群集 — cmd資訊 sysmtool —ns群集 — cmd healthdetail sysmtool —ns datastore —cmd list***

StCtIVM:受影響的ESXi主機的StCtIVM

連線到受影響ESXi主機的StCtIVM

1. 驗證與儲存群集IP(eth1:0)和儲存網路上的其他服務器(StCtIVMs上的eth1)的連接。
 運行***stcli群集資訊 | grep -i -B 1 "stctl|hypervisor"***用於標識分別參與群集的所有ESXi管理IP、StCtIVM eth0(Mgmt)和StCtIVM eth1（儲存資料）測試***ESXI VMK1和SCVM eth1之間的連線 ping -l eth1 [-M do -s 8972] <目標IP地址>, Jumbo幀測試。***
2. 如果問題仍然存在，您可以檢視以下日誌 ***/var/log/springpath/debug-storfs.log*** 檢查是否有任何宕機、搜尋故障或嚴重事件***grep -ai "segmentation|critical|panic" debug-storfs.log/var/log/springpath/stmgr.log******grep -i "om|out of mem" /var/log/kern.log***

3. 最後，您可以嘗試重新啟動仍遇到問題的節點的StCtIVM，並驗證問題是否仍然存在。

簽入ESXi主機：

通過SSH連線到受影響的ESXi主機，並執行下列操作：

1. `esxcli storage nfs list`以列NFS

```
[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
```

您也可以使用命令 `cat /etc/vmware/esx.conf` 從 `/etc/vmware/esx.conf` 進行確認，以驗證NFS裝載的資料儲存上ESXi配置的一致性 | `grep -l nas`

2. 驗證 `/var/log/vmkernel.log` 並查詢有關先前步驟中識別的時間戳的示例失敗狀態、裝入問題或錯誤

3. 驗證IOvisor/NFS Proxy/SCVMClient的狀態 使用 `/etc/init.d/scvmclient status` 命令檢查ESXi上是否正在運行服務 [可選] 您可以使用 `esxcli` 網路ip連線清單驗證是否有任何開啟的連線 | `grep -i "protol|scvmclient"` 確認SCVMClient VIB是否與HX群集的版本相同，`Esxcli` 軟件vib清單 | `grep -i 彈簧`

```
[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		

檢查 `/var/log/scvmclient.log` 以檢視是否存在任何錯誤，即「無法獲取群集對映」如有必要，可以通過 `etc/init.d/scvmclient restart` 重新啟動SCVMClient服務

4. `vmk1ESXiIP eth1:0esxcfg-vmknics -l vmk nicIPMTUvmkping -l vmk1 [-v -s 8972] -d <IP> tESXi[]`

5. `esxcli hardware platform` 獲取在StCtIVm名稱上使用的伺服器SN，並幫助您快速確定特定的StCtIVM正在哪個主機上運行。