

# 排除Hyperflex存储NFS所有路径关闭(APD)问题

## 目录

### [简介](#)

### [HX Datastore如何安装在ESXI上？](#)

### [所有向下路径](#)

### [问题说明](#)

### [故障排除工作流程](#)

### [检入vCenter Server:](#)

### [签入所有StCtIVM:](#)

### [StCtIVM:受影响ESXi主机的StCtIVM](#)

### [签入ESXi主机：](#)

## 简介

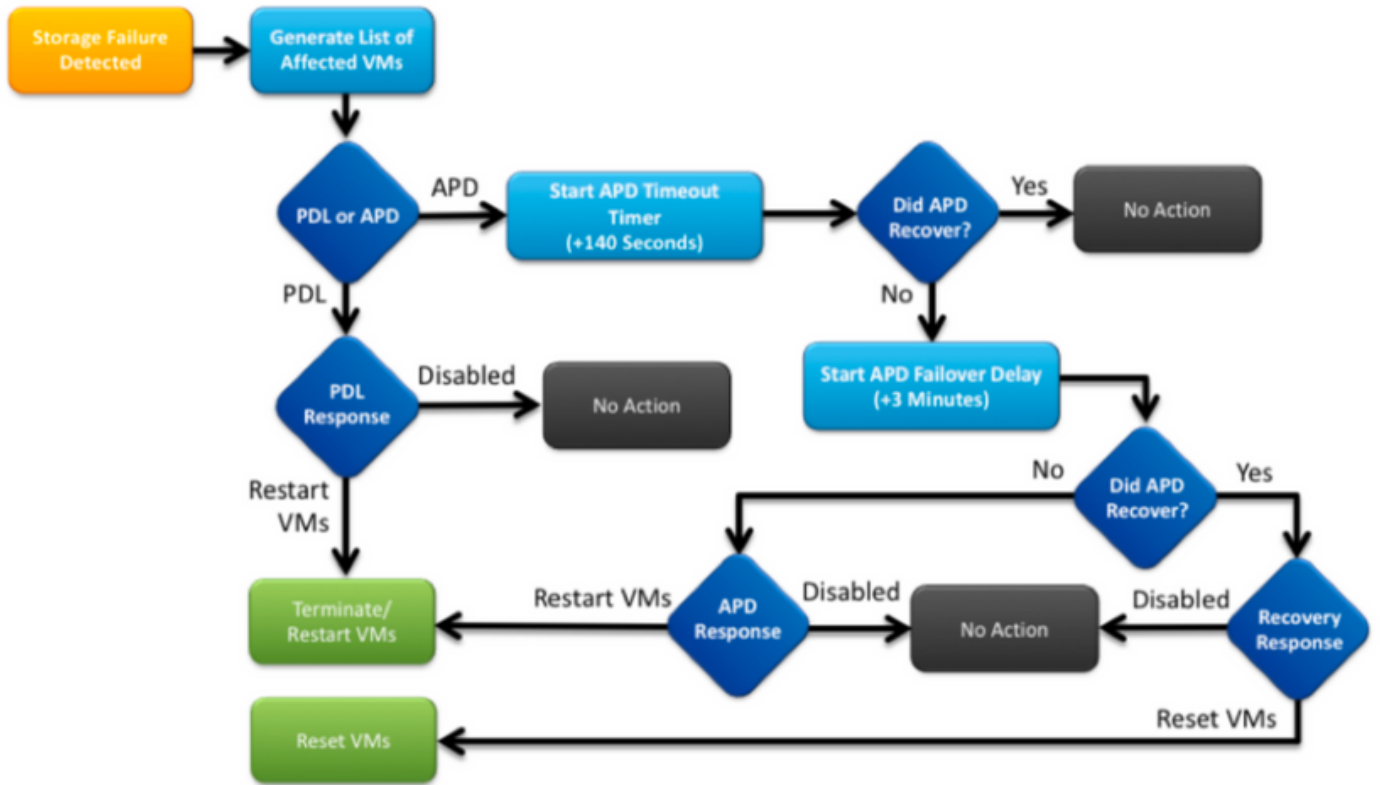
如果您在与Hyperflex集群集成的vCenter中看到“NFS all paths down”错误消息，本文档将为您提供快速了解和故障排除步骤，以便评估问题的根源。

## HX Datastore如何安装在ESXI上？

- Hyperflex Datastore在ESXI主机上装载为NFS装载，为了装载NFS数据存储，我们需要NFS服务器IP，在本例中为eth1:0虚拟浮动接口。
- Hyperflex集群利用虚拟**浮动IP**来管理(eth0:mgmtip)和存储数据(eth1:0)，每个IP将分配到一个特定存储控制器VM(StCtIVM)。请注意，它们可能会以不同的StCtIVM结束。
- 这一点的重要性是**集群存储数据IP(eth1:0)是用于装载Hyperflex集群中创建的数据存储的IP(eth1:0)**。因此，必须从集群的所有节点**分配和可访问该节点**。
- 请注意，如果当前拥有eth1:0虚拟IP的StCtIVM发生故障，它应“迁移”到另一个可用的StCtIVM，其工作方式与FHRP（第一跳冗余协议）类似。

## 所有向下路径

- APD表示主机无法到达存储，并且没有从存储阵列返回永久设备丢失(PDL)SCSI代码。
- 由于它不知道丢失是否是临时的，它将在默认情况下（超时）+ 3分钟（故障转移延迟）继续尝试建立通信140秒以上，ESXi主机将开始使发送到存储设备的任何非虚拟机I/O流量失败。
- 有关vSphere 6.x环境中APD的更多详细信息，请参阅Vmware知识库文章[vSphere 5.x和6.x中的永久设备丢失\(PDL\)和全路径关闭\(APD\)](#)
- 下图说明每个间歇性阶段：



vCenter中的典型错误消息如下所示。

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

## 问题说明

在主机上看到APD警报后，请获取以下信息以更好地了解问题描述：

- 如果一台/多台/所有主机受到影响，以及某些特定主机受到影响
- 如果之前执行过任何更改（配置/升级等）
- 第一次观察问题时以及问题是否反复出现时的时间戳

## 故障排除工作流程

要排除APD故障，我们需要了解3个组件 — vCenter、SCVM和ESXi主机。

这些步骤是建议的工作流程，用于查明或缩小观察到的“所有向下路径”症状的来源。请注意，此订单不必细致遵守，根据客户环境中观察到的特定症状，您可能已满足此要求。

## 检入vCenter Server:

连接到vCenter Server(VCS)并导航到受影响的主机

1. **相关对象** —> **虚拟机**并确认StCtIVM已启动并正在运行
2. **相关对象** —> **Datastore** , 并确认NFS Datastore是否显示“不可访问”。 如果Datastore似乎可访问, 且状态可以尝试在Summary ( 摘要 ) 选项卡上“Reset to Green ( 重置为绿色 ) ”APD事件, 稍后验证警报是否稍后会弹回
3. **Monitor** -> **Issues** 和Monitor -> **Events** 应提供首次发现APD的时间信息。

## 签入所有StCtIVM:

连接到所有StCtIVM并验证以下指针, 您可以使用 [MobaXterm](#) 。

1. 使用日期或`ntpq -p`验证所有StCtIVM是否**具有相同的时间**。StCtIVM上的时间偏差可能导致与zookeeper数据库同步有关的问题, 因此, 在所有StCtIVM之间同步至关重要。ntp服务器的astrick符号表示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APD在升级期间发生, 您可以考虑验证哪些StCtIVM尚未完全升级, 尤其是确定最后一次失败的APD。可能是之前握有eth1:0的 使用`dpkg -l | grep -i springpath`以识别未完全升级的StCtIVM, 因为它们将具有混合版本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. 验证是否所有相关服务都运行`service_status.sh`: 一些主要服务是Springpath File System(storfs)、SCVM Client(`scvmclient`)、 System Management Service(`stMgr`)或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. 如果其中任何一项或其他相关服务未启动，请使用 **start <serviceName>** 例如：**start storfs** 您可以参考 `service_status.sh` 脚本获取服务名称。输入头 — `n25 /bin/service_status.sh` 并确定服务实名。

```

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** 确定哪个 StCtIVM 包含存储集群 IP (eth1:0) 如果 StCtIVM 中未包含 IP，则可能存储未在一个或多个节点上运行。

```

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. 验证 StCtIVM 是否与 CRMMaster 联系，以及 **zookeeper service** 是否已启动并正在运行 **echo srvr | nc localhost 2181** 并检查模式是“引线”(Leader)、“从动”(Follower)还是“独立”(Standolane)，以及连接是否 > 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

```

## 服务展示者状态或ps -ef | grep -i参展商

```
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
```

如果出现任何错误或服务未运行，您可以验证以下日志并尝试启动zookeeper服务

**/var/log/springpath/exhibitor.log /var/log/springpath/stMgr.log**服务参展商开始启动zookeeper服务

7. 验证是否可从所有StCtIVM访问VC stcli cluster info | grep -i "url"以显示所用的URL，其中包含VC的FQDN或IP。使用ping <VC>检验与VC的连接

```
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. 在使用FQDN的群集情况下，确认DNS是否可访问 **stcli services dns show**以列出StCtIVM上配置的DNS服务器。使用ping <DNS\_IP>和host <FQDN> <DNS\_IP> 测试DNS服务器的连接和解析

```
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. 确认所有StCtIVM是否具有相同数量的iptables条目：**iptables -L | wc -l**。如果它们不匹配，请提交TAC案例。

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

10. 当前群集状态和运行状况 **stcli cluster info |小或stcli集群信息 | grep -i "active|state|unavailable"**(如果尝试查找存储不可用时显示哪些特定节点)。

```
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. 哪些数据存储已装载和可用

```

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命令过长或失败，您可以尝试以下sysmtool命令（如果stcli工作，请勿使用）

```

sysmtool --ns cluster --cmd info sysmtool --ns cluster -cmd healthdetail sysmtool -- ns
datastore --cmd list

```

# StCtIVM:受影响ESXi主机的StCtIVM

连接到受影响ESXi主机的StCtIVM。

1. 验证与存储集群IP(eth1:0)和存储网络(StCtIVM上的eth1)上的其他服务器的连接  
运行 **stcli集群信息 | grep -i -B 1 "stctl\|hypervisor"**，以标识分别参与群集的所有ESXi管理IP、StCtIVM eth0（管理）和StCtIVM eth1（存储数据）测试ESXi VMK1和SCVM eth1之间的连接 **ping -l eth1 [-M do -s 8972] <target IP address>**、**巨型帧测试**。
2. 如果问题仍未明确，您可能需要查看以下日志 **/var/log/springpath/debug-storfs.log** 检查是否存在任何紧急情况、seg故障或**严重事件****grep -ai "segmentation\|critical\|panic" debug-**

*storfs.log/var/log/springpath/stmgr.log*grep -i "oom\|out of mem" /var/log/kern.log

3. 最终，您可能尝试重新启动仍然遇到此问题的节点的StCtIVM，并验证问题是否仍然存在。

## 签入ESXi主机：

通过SSH连接到受影响的ESXi主机并执行以下操作：

### 1. *esxcli nfs Esxcfg-nas -l*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
```

您也可以从/etc/vmware/esx.conf确认，以使用命令cat /etc/vmware/esx.conf验证NFS挂载数据存储上ESXi配置的一致性 | *grep -l nas*

2. 验证/var/log/vmkernel.log 并查找在前面步骤中确定的时间戳周围出现故障、挂载问题或错误的示例

3. 验证IOvisor/NFS代理/SCVMClient的状态 使用命令/etc/init.d/scvmclient status检查ESXi上是否正在运行服务 [可选]您可以使用esxcli网络ip连接列表验证是否有打开的连接 | *grep -i "proto|scvmclient"*确认SCVMClient VIB是否与HX集群(Esxcli软件vib列表)的版本相同 | *grep -i春*

```
[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		

检查/var/log/scvmclient.log 以查看是否存在任何错误，即“无法获取集群映射”如有必要，可以通过etc/init.d/scvmclientrestart重新启动SCVMClient服务。

4. *vmk1ESXiIP eth1:0esxcfg-vmknics -lvmk NICIPMTUvmkping -l vmk1 [-v -s 8972] -d <IP>* ESXi

5. *esxcli*硬件平台可获取服务器SN，该SN用于StCtIVm的名称，并可帮助您快速确定特定StCtIVM在哪台主机上运行。