



RHEL 9

SAN Host

NetApp
September 30, 2022

目錄

RHEL 9	1
適用於RHEL 9.0 ONTAP 的NVMe主機組態 (含功能性)	1

RHEL 9

適用於RHEL 9.0 ONTAP 的NVMe主機組態（含功能性）

支援能力

RHEL 9.0支援NVMe（包括NVMe / FC和NVMe / TCP）、並具備ONTAP 非對稱命名空間存取（ANA）功能、可在不間斷的儲存容錯移轉（SFO）上執行。ANA是與NVM環境中的ALUA等效產品、目前是以內核NVMe多重路徑來實作。本文件詳細說明如何在RHEL 9.0上啟用NVMe、並ONTAP 將ANA用作目標的內核NVMe多重路徑。

功能

- 從RHEL 9.0開始、NVMe/TCP不再是技術預覽功能（不像RHEL 8）、而是完全支援的企業功能本身。
- 從RHEL 9.0開始、預設會針對NVMe命名空間啟用內核NVMe多重路徑、而不需要明確設定（不像RHEL 8）。

限制

- 不像NVMe / FC、NVMe / TCP沒有自動連線功能。這反映出Linux主機的兩大限制：
 - 路徑恢復後不自動重新連線：在路徑關閉後的10分鐘內、NVMe / TCP無法自動重新連線至恢復的路徑。
 - 主機開機期間不自動連線-在主機開機期間、NVMe / TCP也無法自動連線。為了輕鬆克服ONTAP 諸如適用於NVMe / TCP的SFO等故障切換事件、建議您調整「Ctrl-Loss_tmo」定時器、以設定較長的重試時間、例如30分鐘。

組態需求

請參閱 "[NetApp 互通性對照表](#)" 如需支援組態的詳細資訊、請參閱。

啟用內核NVMe多重路徑

步驟

1. 在伺服器上安裝RHEL 9.0。安裝完成後、請確認您執行的是指定的RHEL 9.0核心。請參閱 "[NetApp 互通性對照表](#)" 以取得最新的支援版本清單。
2. 安裝完成後、請確認您執行的是指定的RHEL 9.0核心。請參閱 "[NetApp 互通性對照表](#)" 以取得最新的支援版本清單。

```
# uname -r
5.14.0-70.13.1.el9_0.x86_64
```

3. 安裝「NVMe - CLI」套件。

```
# rpm -qa|grep nvme-cli
nvme-cli-1.16-3.el9.x86_64
```

4. 在主機上、檢查位於「/etc/nape/hostnqn」的主機NQN字串、並驗證其是否符合ONTAP 位於「the」（子系統）上之對應子系統的主機NQN字串。例如、

```
# cat /etc/nvme/hostnqn
nqn.2014-08.org.nvmexpress:uuid:9ed5b327-b9fc-4cf5-97b3-1b5d986345d1
```

```
::> vserver nvme subsystem host show -vserver vs_fcnvme_141
Vserver      Subsystem Host      NQN
-----
vs_fcnvme_14 nvme_141_1 nqn.2014-08.org.nvmexpress:uuid:9ed5b327-b9fc-4cf5-97b3-1b5d986345d1
```



如果主機NQN字串不相符、您應該使用「vserver modify」命令來更新對應ONTAP 的NVMe子系統上的主機NQN字串、使其符合主機上的「/etc/nvm/hostnqn」中的主機NQN字串。

5. 重新啟動主機。

設定NVMe/FC

Broadcom / Emulex

1. 確認您使用的是支援的介面卡。如需最新的支援介面卡清單、請參閱 "[NetApp 互通性對照表](#)"。

```
# cat /sys/class/scsi_host/host*/modelname
LPe32002-M2
LPe32002-M2
```

```
# cat /sys/class/scsi_host/host*/modeldesc
Emulex LightPulse LPe32002-M2 2-Port 32Gb Fibre Channel Adapter
Emulex LightPulse LPe32002-M2 2-Port 32Gb Fibre Channel Adapter
```

2. 請確認您使用的是建議的Broadcom lfit韌體和收件匣驅動程式。如需支援的介面卡驅動程式和韌體版本的最清單、請參閱 "[NetApp 互通性對照表](#)"。

```
# cat /sys/class/scsi_host/host*/fwrev
12.8.351.47, sli-4:2:c
12.8.351.47, sli-4:2:c
```

```
# cat /sys/module/lpfc/version
0:14.0.0.4
```

3. 確認「lffc_enable_FC4_type」設定為「3」。

```
# cat /sys/module/lpfc/parameters/lpfc_enable_fc4_type
3
```

4. 確認啟動器連接埠已啟動並正在執行、而且您可以看到目標LIF。

```
# cat /sys/class/fc_host/host*/port_name
0x100000109b1c1204
0x100000109b1c1205
```

```
# cat /sys/class/fc_host/host*/port_state
Online
Online
```

```

# cat /sys/class/scsi_host/host*/nvme_info

NVME Initiator Enabled
XRI Dist lpfc0 Total 6144 IO 5894 ELS 250
NVME LPORT lpfc0 WWPN x100000109b1c1204 WWNN x200000109b1c1204 DID
x011d00 ONLINE
NVME RPORT WWPN x203800a098dfdd91 WWNN x203700a098dfdd91 DID x010c07
TARGET DISCSRVC ONLINE
NVME RPORT WWPN x203900a098dfdd91 WWNN x203700a098dfdd91 DID x011507
TARGET DISCSRVC ONLINE

NVME Statistics
LS: Xmt 0000000f78 Cmpl 0000000f78 Abort 00000000
LS XMIT: Err 00000000 CMPL: xb 00000000 Err 00000000
Total FCP Cmpl 000000002fe29bba Issue 000000002fe29bc4 OutIO
0000000000000000a
abort 00001bc7 noxri 00000000 nondlp 00000000 qdepth 00000000 wqerr
00000000 err 00000000
FCP CMPL: xb 00001e15 Err 0000d906

NVME Initiator Enabled
XRI Dist lpfc1 Total 6144 IO 5894 ELS 250
NVME LPORT lpfc1 WWPN x100000109b1c1205 WWNN x200000109b1c1205 DID
x011900 ONLINE
NVME RPORT WWPN x203d00a098dfdd91 WWNN x203700a098dfdd91 DID x010007
TARGET DISCSRVC ONLINE
NVME RPORT WWPN x203a00a098dfdd91 WWNN x203700a098dfdd91 DID x012a07
TARGET DISCSRVC ONLINE

NVME Statistics
LS: Xmt 0000000fa8 Cmpl 0000000fa8 Abort 00000000
LS XMIT: Err 00000000 CMPL: xb 00000000 Err 00000000
Total FCP Cmpl 000000002e14f170 Issue 000000002e14f17a OutIO
0000000000000000a
abort 000016bb noxri 00000000 nondlp 00000000 qdepth 00000000 wqerr
00000000 err 00000000
FCP CMPL: xb 00001f50 Err 0000d9f8

```

5. 啟用1MB I/O大小。

需要將"lfc_sg_seg_cnt"參數設為"256"、以讓"lffc"驅動程式發出大小高達1 MB的I/O要求。

```

# cat /etc/modprobe.d/lpfc.conf
options lpfc lpfc_sg_seg_cnt=256

```

- a. 執行「dracut -f」命令、然後重新啟動主機。
- b. 主機開機後、請確認「lfc_sg_seg_cnt」設定為「256」。

```
# cat /sys/module/lpfc/parameters/lpfc_sg_seg_cnt
256
```

Marvell / QLogic

RHEL 9.0核心隨附的原生內建內建qla2xxx驅動程式、提供ONTAP 最新的上游修正程式、是支援此功能不可或缺的要素。確認您執行的是支援的介面卡驅動程式和韌體版本：

```
# cat /sys/class/fc_host/host*/symbolic_name
QLE2742 FW:v9.06.02 DVR:v10.02.00.200-k
QLE2742 FW:v9.06.02 DVR:v10.02.00.200-k
```

驗證是否已設定「ql2xnvmeenable」、以便Marvell介面卡能作為NVMe / FC啟動器運作：

```
# cat /sys/module/qla2xxx/parameters/ql2xnvmeenable
1
```

設定NVMe/TCP

不像NVMe / FC、NVMe / TCP沒有自動連線功能。這反映出Linux NVMe/TCP主機的兩大限制：

- 路徑恢復後不自動重新連線：在路徑關閉後的10分鐘內、NVMe / TCP無法自動重新連線至恢復的路徑。
- 主機開機期間不自動連線-在主機開機期間、NVMe / TCP也無法自動連線。為了輕鬆克服ONTAP SFO等故障切換事件、建議您調整「Ctrl-Loss_tmo」定時器、設定較長的重試時間、例如30分鐘。以下是詳細資料：

步驟

1. 驗證啟動器連接埠是否能夠擷取支援的NVMe / TCP LIF上的探索記錄頁面資料：

```
# nvme discover -t tcp -w 192.168.1.8 -a 192.168.1.51

Discovery Log Number of Records 10, Generation counter 119
=====Discovery Log Entry 0=====
trtype: tcp
adrfam: ipv4
subtype: nvme subsystem
treq: not specified
portid: 0
trsvcid: 4420
subnqn: nqn.1992-
08.com.netapp:sn.56e362e9bb4f11ebbaded039ea165abc:subsystem.nvme_118_tcp
_1
traddr: 192.168.2.56
sectype: none
=====Discovery Log Entry 1=====
trtype: tcp
adrfam: ipv4
subtype: nvme subsystem
treq: not specified
portid: 1
trsvcid: 4420
subnqn: nqn.1992-
08.com.netapp:sn.56e362e9bb4f11ebbaded039ea165abc:subsystem.nvme_118_tcp
_1
traddr: 192.168.1.51
sectype: none
=====Discovery Log Entry 2=====
trtype: tcp
adrfam: ipv4
subtype: nvme subsystem
treq: not specified
portid: 0
trsvcid: 4420
subnqn: nqn.1992-
08.com.netapp:sn.56e362e9bb4f11ebbaded039ea165abc:subsystem.nvme_118_tcp
_2
traddr: 192.168.2.56
sectype: none
...
```

2. 同樣地、請確認其他的NVMe/TCP啟動器目標LIF組合能夠成功擷取探索記錄頁面資料。例如、


```
# nvme discover -t tcp -w 192.168.1.8 -a 192.168.1.51
# nvme discover -t tcp -w 192.168.1.8 -a 192.168.1.52
# nvme discover -t tcp -w 192.168.2.9 -a 192.168.2.56
# nvme discover -t tcp -w 192.168.2.9 -a 192.168.2.57
```

3. 在所有支援的NVMe /TCP啟動器目標LIF上、跨節點執行「NVMe Connect All」命令。請確保在連線期間經過較長的「Ctrl-Loss_tmo」期間（例如30分鐘、可設定至「-l 1800」）、以便在路徑遺失時重試更長時間。例如、

```
# nvme connect-all -t tcp -w 192.168.1.8 -a 192.168.1.51 -l 1800
# nvme connect-all -t tcp -w 192.168.1.8 -a 192.168.1.52 -l 1800
# nvme connect-all -t tcp -w 192.168.2.9 -a 192.168.2.56 -l 1800
# nvme connect-all -t tcp -w 192.168.2.9 -a 192.168.2.57 -l 1800
```

驗證NVMf

1. 檢查下列項目、確認內核NVMe多重路徑確實已啟用：

```
# cat /sys/module/nvme_core/parameters/multipath
Y
```

2. 驗證ONTAP 適當的NVMf設定（例如、設定為「NetApp支援控制器」的模型、以及將「iopolicing」設定為「循環配置資源」的負載平衡）ONTAP、以正確反映在主機上：

```
# cat /sys/class/nvme-subsystem/nvme-subsys*/model
NetApp ONTAP Controller
NetApp ONTAP Controller
```

```
# cat /sys/class/nvme-subsystem/nvme-subsys*/iopolicy
round-robin
round-robin
```

3. 確認ONTAP 支援的名稱空間能正確反映在主機上。例如（A） 、

```
# nvme list
Node          SN                      Model                      Namespace
Usage
-----
-----
/dev/nvme0n1 814vWBNRwf9HAAAAAAB NetApp ONTAP Controller 1
85.90 GB / 85.90 GB

Format          FW Rev
-----
4 KiB + 0 B    FFFFFFFF
```

範例 (b) :

```
# nvme list
Node          SN                      Model                      Namespace
Usage
-----
-----
/dev/nvme0n1 81CZ5BQuUNfGAAAAAAB NetApp ONTAP Controller 1
85.90 GB / 85.90 GB

Format          FW Rev
-----
4 KiB + 0 B    FFFFFFFF
```

4. 確認每個路徑的控制器狀態均為有效、且具有適當的ANA狀態。例如 (A) 、

```
# nvme list-subsys /dev/nvme0n1
nvme-subsys0 - NQN=nqn.1992-
08.com.netapp:sn.5f5f2c4aa73b11e9967e00a098df41bd:subsystem.nvme_141_1
\
+- nvme0 fc traddr=nn-0x203700a098dfdd91:pn-0x203800a098dfdd91
host_traddr=nn-0x200000109b1c1204:pn-0x100000109b1c1204 live
inaccessible
+- nvme1 fc traddr=nn-0x203700a098dfdd91:pn-0x203900a098dfdd91
host_traddr=nn-0x200000109b1c1204:pn-0x100000109b1c1204 live
inaccessible
+- nvme2 fc traddr=nn-0x203700a098dfdd91:pn-0x203a00a098dfdd91
host_traddr=nn-0x200000109b1c1205:pn-0x100000109b1c1205 live optimized
+- nvme3 fc traddr=nn-0x203700a098dfdd91:pn-0x203d00a098dfdd91
host_traddr=nn-0x200000109b1c1205:pn-0x100000109b1c1205 live optimized
```

範例 (b) :

```
# nvme list-subsys /dev/nvme0n1
nvme-subsys0 - NQN=nqn.1992-
08.com.netapp:sn.56e362e9bb4f11ebbade039ea165abc:subsystem.nvme_118_tcp
_1
\
+- nvme0 tcp traddr=192.168.1.51 trsvcid=4420 host_traddr=192.168.1.8
live optimized
+- nvme10 tcp traddr=192.168.2.56 trsvcid=4420 host_traddr=192.168.2.9
live optimized
+- nvme15 tcp traddr=192.168.2.57 trsvcid=4420 host_traddr=192.168.2.9
live non-optimized
+- nvme5 tcp traddr=192.168.1.52 trsvcid=4420 host_traddr=192.168.1.8
live non-optimized
```

5. 驗證NetApp外掛程式是否顯示每ONTAP 個不支援的名稱空間裝置的正確值。例如 (A) 、

```
# nvme netapp ontapdevices -o column
Device          Vserver          Namespace Path
NSID
-----
-----
/dev/nvme0n1    vs_fcnvme_141    /vol/fcnvme_141_vol_1_1_0/fcnvme_141_ns    1

UUID                                Size
-----
72b887b1-5fb6-47b8-be0b-33326e2542e2    85.90GB

# nvme netapp ontapdevices -o json
{
  "ONTAPdevices" : [
    {
      "Device" : "/dev/nvme0n1",
      "Vserver" : "vs_fcnvme_141",
      "Namespace_Path" : "/vol/fcnvme_141_vol_1_1_0/fcnvme_141_ns",
      "NSID" : 1,
      "UUID" : "72b887b1-5fb6-47b8-be0b-33326e2542e2",
      "Size" : "85.90GB",
      "LBA_Data_Size" : 4096,
      "Namespace_Size" : 20971520
    }
  ]
}
```

範例 (b) :

```
# nvme netapp ontapdevices -o column
Device          Vserver          Namespace Path
-----
/dev/nvme0n1    vs_tcp_118
/vol/tcpcnvme_118_1_0_0/tcpcnvme_118_ns

NSID   UUID                               Size
-----
1      4a3e89de-b239-45d8-be0c-b81f6418283c 85.90GB
```

```
# nvme netapp ontapdevices -o json
{
  "ONTAPdevices" : [
    {
      "Device" : "/dev/nvme0n1",
      "Vserver" : "vs_tcp_118",
      "Namespace_Path" : "/vol/tcpcnvme_118_1_0_0/tcpcnvme_118_ns",
      "NSID" : 1,
      "UUID" : "4a3e89de-b239-45d8-be0c-b81f6418283c",
      "Size" : "85.90GB",
      "LBA_Data_Size" : 4096,
      "Namespace_Size" : 20971520
    },
  ]
}
```

疑難排解

在開始任何NVMe / FC故障的疑難排解之前、請務必確保您所執行的組態符合IMT 《支援技術規範的要求。然後繼續執行下列步驟、以偵錯任何主機端問題。

lffc詳細記錄

以下是適用於NVMe/FC的lffcs驅動程式記錄位元遮罩清單、如「驅動程式/scsi/lffc/lffc_logmsg.h」所示：

```
#define LOG_NVME 0x00100000 /* NVME general events. */
#define LOG_NVME_DISC 0x00200000 /* NVME Discovery/Connect events. */
#define LOG_NVME_ABTS 0x00400000 /* NVME ABTS events. */
#define LOG_NVME_IOERR 0x00800000 /* NVME IO Error events. */
```

您可以將「lffc_log_verbose」驅動程式設定（附加至lffc行、位於「/etc/modprobe.d/lffc.conf」）設定為上述任何值、以便從「lffc」驅動程式觀點記錄nvm/FC事件。然後執行「dracut -f」命令重新建立「initramfs」、然後重新啟動主機。重新開機後、請使用上述「log_nvm_disc」位元遮罩、檢查下列項目、確認是否已套用詳細記錄：

```
# cat /etc/modprobe.d/lpfc.conf
options lpfc_enable_fc4_type=3 lpfc_log_verbose=0xf00083
```

```
# cat /sys/module/lpfc/parameters/lpfc_log_verbose
15728771
```

qla2xxx詳細記錄

NVMe / FC沒有類似的特定qla2xxx記錄、如「lffc」中所示。您可以在此處設定一般qla2xxx記錄層級、例如「ql2xextended_error_logging=0x1e400000」。您可以將此值附加至對應的「modprobe qla2xxx.conf」檔案、以完成此作業。然後執行「dracut -f」重新建立「initramfs」、然後重新啟動主機。重新開機後、請確認已套用詳細記錄、如下所示：

```
# cat /etc/modprobe.d/qla2xxx.conf
options qla2xxx ql2xnvmeenable=1 ql2xextended_error_logging=0x1e400000
```

```
# cat /sys/module/qla2xxx/parameters/ql2xextended_error_logging
507510784
```

常見的NVMe-CLI錯誤和因應措施

「NVMe - CLI」顯示錯誤	可能原因	因應措施
「無法寫入/開發/NVMe架構：無效的引數」錯誤發生於「NVMe探索」、「NVMe連線」或「NVMe連線-全部」	如果語法錯誤、通常會顯示此錯誤訊息。	請確定上述NVMe命令的語法正確無誤。

「NVMe - CLI」顯示錯誤	可能原因	因應措施
<p>「無法寫入/開發/NVMe架構：在「NVMe探索」、「NVMe連線」或「NVMe連線全部」期間、沒有這類檔案或目錄</p>	<p>可能會引發多個問題。常見的情況包括：您將錯誤的引數傳遞給上述NVMe命令。</p>	<p>請確定您已針對上述命令傳遞適當的引數（例如適當的WWNN字串、WWPN字串等）。如果引數正確、但仍看到此錯誤、請檢查NVMe啟動器的「/sys/class/scsi_host/host/nvm_info」輸出是否正確、並在遠端連接埠區段下方正確顯示「已啟用」和nvm/FC目標LIF。例如、</p> <pre data-bbox="743 436 1479 1816"> # cat /sys/class/scsi_host/host*/nvme_info NVME Initiator Enabled NVME LPORT lpfc0 WWPN x10000090fae0ec9d WWNN x20000090fae0ec9d DID x012000 ONLINE NVME RPORT WWPN x200b00a098c80f09 WWNN x200a00a098c80f09 DID x010601 TARGET DISCSRVC ONLINE NVME Statistics LS: Xmt 0000000000000006 Cmpl 0000000000000006 FCP: Rd 0000000000000071 Wr 0000000000000005 IO 0000000000000031 Cmpl 00000000000000a6 Outstanding 0000000000000001 NVME Initiator Enabled NVME LPORT lpfc1 WWPN x10000090fae0ec9e WWNN x20000090fae0ec9e DID x012400 ONLINE NVME RPORT WWPN x200900a098c80f09 WWNN x200800a098c80f09 DID x010301 TARGET DISCSRVC ONLINE NVME Statistics LS: Xmt 0000000000000006 Cmpl 0000000000000006 FCP: Rd 0000000000000073 Wr 0000000000000005 IO 0000000000000031 Cmpl 00000000000000a8 Outstanding 0000000000000001 </pre> <p>因應措施：如果「NVMe資訊」輸出中未顯示上述目標LIF、請檢查「/var/log/Messages」和「dmesg」輸出是否有任何可疑的NVMe / FC故障、並據此報告或修正。</p>

「NVMe - CLI」顯示錯誤	可能原因	因應措施
在「NVMe探索」、「NVMe連線」或「NVMe連線」期間、「無探索記錄項目可擷取」	如果NetApp陣列上的對應子系統未新增「/etc/nexe/hostnqn」字串、或是在個別子系統中新增不正確的「hostnqn」字串、通常會出現此錯誤訊息。	請確定NetApp陣列上的對應子系統中已新增確切的「/etc/nvm/hostnqn」字串（請透過「vserver NVMe子系統主機show」驗證）。
「無法寫入/開發/NVMe架構：作業已在進行中」、「NVMe探索」、「NVMe連線」或「NVMe連線-全部」	如果已建立控制器關聯或指定的作業、或正在建立、就會出現此錯誤訊息。這可能是上述安裝之自動連線指令碼的一部分。	無。若為「NVMe探索」、請稍後嘗試執行此命令。對於「NVMe Connect」和「Connect All」、請執行「NVMe清單」、以驗證命名空間裝置是否已建立並顯示在主機上。

何時聯絡技術支援

如果您仍遇到問題、請收集下列檔案和命令輸出、然後將其傳送以供進一步分類：

```

cat /sys/class/scsi_host/host*/nvme_info
/var/log/messages
dmesg
nvme discover output as in:
nvme discover --transport=fc --traddr=nn-0x200a00a098c80f09:pn
-0x200b00a098c80f09 --host-traddr=nn-0x20000090fae0ec9d:pn
-0x10000090fae0ec9d
nvme list
nvme list-subsys /dev/nvmeXnY

```

版權資訊

Copyright©2022 NetApp、Inc.版權所有。美國印製本文件中版權所涵蓋的任何部分、不得以任何形式或任何方式（包括影印、錄製、在未事先取得版權擁有者書面許可的情況下、在電子擷取系統中進行錄音或儲存。

衍生自受版權保護之NetApp資料的軟體必須遵守下列授權與免責聲明：

本軟體係由NetApp「依現狀」提供、不含任何明示或暗示的保證、包括但不限於適售性及特定用途適用性的暗示保證、特此聲明。在任何情況下、NetApp均不對任何直接、間接、偶發、特殊、示範、或衍生性損害（包括但不限於採購替代商品或服務；使用損失、資料或利潤損失；或業務中斷）、無論是在合約、嚴格責任或侵權行為（包括疏忽或其他）中、無論是因使用本軟體而產生的任何責任理論（包括疏忽或其他）、即使已被告知可能造成此類損害。

NetApp保留隨時變更本文所述之任何產品的權利、恕不另行通知。除非NetApp以書面明確同意、否則NetApp不承擔因使用本文所述產品而產生的任何責任或責任。使用或購買本產品並不代表NetApp擁有任何專利權利、商標權利或任何其他智慧財產權。

本手冊所述產品可能受到一或多個美國國家/地區的保護專利、國外專利或申請中。

限制權利圖例：政府使用、複製或揭露受DFARS 252.277-7103（1988年10月）和FAR 52-227-19（1987年6月）技術資料與電腦軟體權利條款（c）（1）（ii）分段所述限制。

商標資訊

NetApp、NetApp標誌及所列的標章 <http://www.netapp.com/TM> 為NetApp、Inc.的商標。其他公司和產品名稱可能為其各自所有者的商標。