

# FlexPod常見效能問題

## 目錄

[簡介](#)

[FlexPod概念概述](#)

[效能注意事項](#)

[環境](#)

[測量](#)

[基線](#)

[FlexPod中的效能問題](#)

[常見問題](#)

[幀和資料包丟失](#)

[MTU不相符](#)

[Nexus 5000和UCS平台上的MTU顯示](#)

[端到端配置](#)

[測試端到端巨型幀](#)

[緩衝區相關問題](#)

[驅動程式問題](#)

[介面卡資訊](#)

[邏輯封包流](#)

[輸入/輸出模組](#)

[設計注意事項](#)

[埠速度選擇和埠通道注意事項](#)

[儲存特定問題](#)

[儲存位置](#)

[最佳路徑選取](#)

[虛擬機器和虛擬機器監控程式流量共用](#)

[疑難排解提示](#)

[縮小問題範圍](#)

[思科](#)

[計數器限制](#)

[控制平面注意事項](#)

[捕獲流量](#)

[NetApp](#)

[VMware](#)

[已知問題和增強功能](#)

[TAC案例](#)

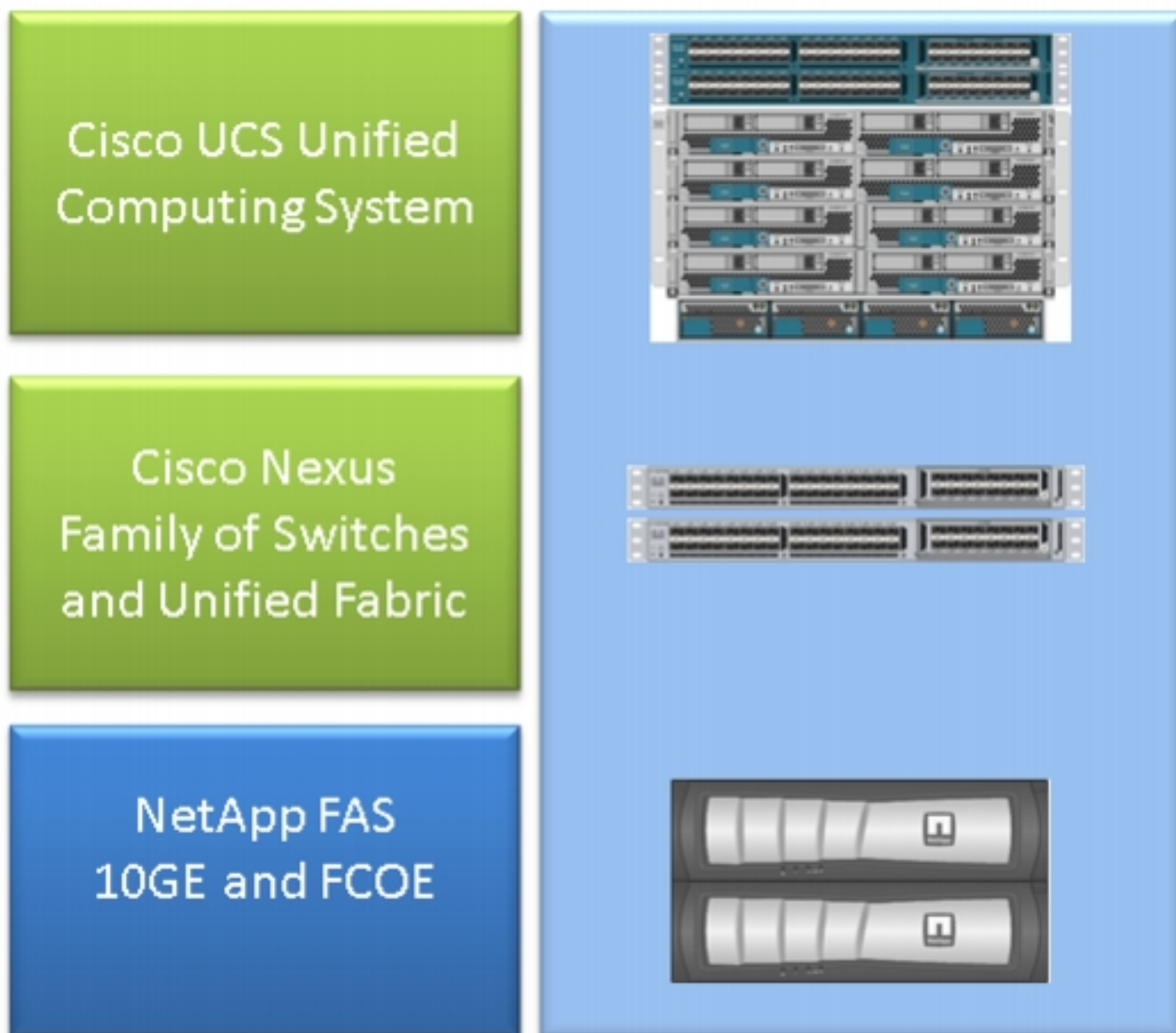
[意見回饋](#)

# 簡介

本文檔介紹FlexPod環境中的常見效能問題，提供故障排除方法並提供緩解步驟。它旨在為希望對FlexPod環境中的效能進行故障排除的客戶提供起點。本文是根據資料中心解決方案技術支援中心(TAC)團隊最近幾個月發現的問題而撰寫的。

## FlexPod概念概述

FlexPod包括通過Nexus交換機連線到NetApp儲存和IP網路的統一計算系統(UCS)電腦。



最常見的FlexPod包含通過交換矩陣互聯(FI)連線到Nexus 5500交換機再連線到NetApp檔案管理器的Cisco UCS B系列機箱。另一個稱為FlexPod Express的解決方案使用連線到Nexus 3000交換機的UCS C系列機箱。本文討論最常見的FlexPod。

## 效能注意事項

在FlexPod中常見的、具有多個責任方的複雜環境中，您需要考慮多個方面，以便排除故障。第2層和IP網路中的典型效能問題源於：

- 資料包或幀丟失 — 資料位丟失會對應用程式效能產生不利影響。
- 緩衝 — 如果資料包或幀在隊列/緩衝區中花費太多時間，應用程式可能會看到某些效能影響，特別是在儲存網路的情況下。延遲、重新排序和規範器問題屬於此類別。
- MTU不相符問題和分段 — 達到更高的效能時經常遇到的問題。與分段和MTU不一致相關的問題屬於此類別。

## 環境

瞭解用來衡量效能的環境非常重要。應提出有關儲存型別和協定以及受影響伺服器的作業系統(OS)和位置的問題，以適當縮小問題範圍。概述連通性的拓撲圖是最基本的最小值。

## 測量

你需要知道什麼是測量的，它是如何測量的。某些應用程式以及大多數儲存和虛擬機器監控程式供應商提供某種型別的測量結果，以指示系統的效能/運行狀況。這些測量值是一個很好的起點，因為它們不能替代大多數故障排除方法。

例如，虛擬機器監控程式中的網路檔案系統(NFS)儲存延遲測量可能表示效能下降，但就其本身而言，它並不影響網路。在NFS的情況下，從主機到NFS儲存IP網路的簡單ping可能表明是否該責怪網路。

## 基線

這一點強調得不夠，尤其是當您建立TAC案例時。為了表示效能不理想，需要表示測量的引數。這包括預期值和測試值。理想情況下，您應該顯示以前數據和用來實現該資料的測試方法。

例如；當測試時(從單個啟動器到單個邏輯單元號(LUN)僅寫入時)達到10毫秒的延遲，可能並不表示滿載系統的延遲應該是多少。

## FlexPod中的效能問題

由於本文檔旨在作為大多數FlexPod環境的參考，因此它僅概述了負責資料中心解決方案的TAC團隊所發現的最常見問題。

## 常見問題

本節將討論儲存和IP/第2層網路的常見問題。

### 幀和資料包丟失

幀和資料包丟失是影響效能的最常見因素。查詢問題指示的常見位置之一是在介面級別。從Nexus 5000或UCS Nexus作業系統(NX-OS)CLI輸入 `show interface | sec "is up" | egrep`

^(Eth|fc)|discard|drop|CRC命令。對於已啟動的介面，它會列出名稱並丟棄計數器和丟棄。同樣，當您輸入show interface counters error命令時，會顯示一個很好的概述，該命令會顯示所有介面的錯誤統計資訊。

## 乙太網世界

知道非0處的計數器可能不指示問題是很重要的。在某些情況下，這些計數器可能在初始設定或以前的操作更改中引發。應監控計數器的增加。

您還可以從ASIC級別收集計數器，這可能更具指示性。具體而言，對於介面上的循環備援檢查(CRC)錯誤，要輸入的TAC常用命令是show hardware internal carmel crc。Carmel是負責埠級轉發的ASIC的名稱。

每個埠都可以從6100系列FI或Nexus 5600交換機獲得類似的輸出。對於FI 6100 (即gatos ASIC)，請輸入以下命令：

```
show hardware internal gatos port ethernet X/Y | grep
"OVERSIZE|TOOLONG|DISCARD|UNDERSIZE|FRAGMENT|T_CRC|ERR|JABBER|PAUSE"
```

對於Nexus 5600，從bigsur ASIC輸入以下命令：

```
show hardware internal bigsur port eth x/y | egrep
"OVERSIZE|TOOLONG|DISCARD|UNDERSIZE|FRAGMENT|T_CRC|ERR|JABBER|PAUSE"
```

carmel ASIC的命令顯示CRC資料包的接收位置和轉發到的位置，更重要的是，它們是否已經堆積。

由於Nexus 5000和UCS NX-OS操作都是直通式操作，因此具有不正確幀校驗序列(FCS)的交換模式幀僅在轉發之前進行堆疊。找出損壞的幀來自何處非常重要。

```
bdsol-6248-06-A(nxos)# show hardware internal carmel crc
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+
| Port   | MM rx CRC | MM Rx Stomp| FI rx CRC | FI Rx Stomp | FI tx CRC | FI tx Stomp| MM tx CRC |
+-----+-----+-----+-----+-----+-----+-----+-----+
(....)
| Eth 1/17 |    --- |    --- |    --- |    908100 |    --- |    --- |    --- |
| Eth 1/18 |    --- |    --- |    --- |    298658 |    --- |    --- |    --- |
(....)
| Eth 1/34 |    --- |    --- |    --- |    --- |    --- | 1206758 | 1206758 |
```

此範例顯示來自Eth 1/17和Eth 1/18的堆疊資料包，它們是Nexus 5000的上行鏈路。可以假設這些幀後來被傳送到乙太網1/34，例如Eth 1/17 + Eth 1/18 rx Stomp = Eth 1/34 tx Stomp。

Nexus 5000的類似外觀顯示：

```
bdsol-n5548-05# show hardware internal carmel crc
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+
| Port   | MM rx CRC | MM Rx Stomp| FI rx CRC | FI Rx Stomp| FI tx CRC | FI tx Stomp| MM tx CRC |
+-----+-----+-----+-----+-----+-----+-----+-----+
(....)
| Eth 1/14 |    13 |    --- |    --- |    --- |    13 |    --- |    --- |
```

```
-- |
(.....)
| Eth 1/19 |      7578 |      --- |      --- |      7463 |      --- |      --- |
--- |
```

此輸出顯示兩個鏈路上接收的CRC，並在轉發前標籤為儲存。有關詳細資訊，請參閱[Nexus 5000故障排除指南](#)。

## 光纖通道世界

查詢丟棄（丟棄、錯誤、CRC、B2B信用耗盡）的一種簡單方法是通過**show interface counters fc**命令。

此命令在Nexus 5000和交換矩陣互聯上可用，可很好地指示光纖通道世界中的情況。

例如：

```
bdsol-n5548-05# show interface counters fc | i fc|disc|error|B2B|rate|put
fc2/16
1 minute input rate 72648 bits/sec, 9081 bytes/sec, 6 frames/sec
1 minute output rate 74624 bits/sec, 9328 bytes/sec, 5 frames/sec
96879643 frames input, 155712103332 bytes
0 discards, 0 errors, 0 CRC
113265534 frames output, 201553309480 bytes
0 discards, 0 errors
0 input OLS, 1 LRR, 0 NOS, 0 loop inits
1 output OLS, 2 LRR, 0 NOS, 0 loop inits
0 transmit B2B credit transitions from zero
0 receive B2B credit transitions from zero
16 receive B2B credit remaining
32 transmit B2B credit remaining
0 low priority transmit B2B credit remaining
(...)
```

此介面不忙，並且輸出顯示沒有丟棄或發生錯誤。

此外，B2B信貸從0的轉變也得到了強調；由於Cisco錯誤ID [CSCue80063](#)和[CSCut08353](#)，這些計數器無法信任。它們在Cisco MDS上工作正常，但在Nexus5k平台的UCS上工作不正常。此外，您還可以驗證思科錯誤ID [CSCsz95889](#)。

與乙太網中光纖通道(FC)的卡梅爾類似，可以使用fc-mac設施。例如，對於埠fc2/1，輸入**show hardware internal fc-mac 2 port 1 statistics**命令。顯示的計數器為十六進位制格式。

```
bdsol-6248-06-A(nxos)# show interface fc1/32 | i disc
    15 discards, 0 errors
    0 discards, 0 errors
bdsol-6248-06-A(nxos)# show hardware internal fc-mac 1 port 32 statistics
ADDRESS          STAT                                     COUNT
-----
0x0000003d FCP_CNTR_MAC_RX_BAD_WORDS_FROM_DECODER 0x70
0x00000042 FCP_CNTR_MAC_CREDIT_IG_XG_MUX_SEND_RRDY_REQ 0x1e4f1026
0x00000043 FCP_CNTR_MAC_CREDIT_EG_DEC_RRDY 0x66cafd1
0x00000061 FCP_CNTR_MAC_DATA_RX_CLASS3_FRAMES 0x1e4f1026
0x00000069 FCP_CNTR_MAC_DATA_RX_CLASS3_WORDS 0xe80946c708
0x000d834c FCP_CNTR_PIF_RX_DROP 0xf
0x00000065 FCP_CNTR_MAC_DATA_TX_CLASS3_FRAMES 0x66cafd1
0x0000006d FCP_CNTR_MAC_DATA_TX_CLASS3_WORDS 0x2b0fae9588
```

```
0xffffffff FCP_CNTR_OLS_IN 0x1
0xffffffff FCP_CNTR_LRR_IN 0x1
0xffffffff FCP_CNTR_OLS_OUT 0x1
```

輸出顯示輸入時丟棄15個。它可以與FCP\_CNTR\_PIF\_RX\_DROP匹配，後者計數為0xf ( 15為小數 )。此資訊可再次與FWM ( 轉發管理器 ) 資訊關聯。

```
bdsol-6248-06-A(nxos)# show platform fwm info pif fc 1/32 verbose | i drop|discard|asic
fc1/32 pd: slot 0 logical port num 31 slot_asic_num 3 global_asic_num 3 fwm_inst 7
fc 0
fc1/32 pd: tx stats: bytes 191196731188 frames 107908990 discard 0 drop 0
fc1/32 pd: rx stats: bytes 998251154572 frames 509332733 discard 0 drop 15
fc1/32 pd fcoe: tx stats: bytes 191196731188 frames 107908990 discard 0 drop 0
fc1/32 pd fcoe: rx stats: bytes 998251154572 frames 509332733 discard 0 drop 15
```

但是，這會向管理員告知丟棄量，以及相應的ASIC編號。需要查詢有關丟棄ASIC原因的get資訊。

```
bdsol-6248-06-A(nxos)# show platform fwm info asic-errors 3
Printing non zero Carmel error registers:
DROP_SHOULD_HAVE_INT_MULTICAST: res0 = 25 res1 = 0 [36]
DROP_INGRESS_ACL: res0 = 15 res1 = 0 [46]
```

在這種情況下，流量被入口訪問控制清單(ACL)丟棄，通常在FC世界 — 分割槽中。

## MTU不相符

在FlexPod環境中，必須根據需要為應用程式和協定容納端到端最大過渡單元(MTU)設定。對於大多數環境，這是乙太網光纖通道(FCoE)和巨型幀。

此外，如果發生分段，預期效能會降低。對於網路檔案系統(NFS)和網際網路小型電腦系統介面(iSCSI)等協定，測試和證明端到端IP最大傳輸單元(MTU)和TCP最大分段大小(MSS)非常重要。

無論您對巨型幀還是FCoE進行故障排除，必須記住這兩個問題都需要在整個環境中進行一致的配置和服務類別(CoS)標籤，以便正常運行。

對於UCS和Nexus，用於驗證每個介面、每個QoS組MTU設定的命令是**show queuing interface | i queuing|qos-group|MTU**。

## Nexus 5000和UCS平台上的MTU顯示

UCS和Nexus的已知方面是介面上的MTU顯示。此輸出演示了一個配置為將巨型幀和FCoE排隊的介面：

```
bdsol-6248-06-A(nxos)# show queuing interface e1/1 | i MTU
q-size: 360640, HW MTU: 9126 (9126 configured)
q-size: 79360, HW MTU: 2158 (2158 configured)
```

同時，**show interface**命令顯示1500位元組：

```
bdsol-6248-06-A(nxos)# show int e1/1 | i MTU
MTU 1500 bytes, BW 10000000 Kbit, DLY 10 usec
```

如果與卡梅爾ASIC資訊相比，ASIC顯示給定埠的MTU能力。

```
show hardware internal carmel port ethernet 1/1 | egrep -i MTU
      mtu                : 9260
```

上述平台上預期會出現這種MTU不匹配的情況，可能會誤導新生。

## 端到端配置

端到端一致配置是保證正確效能的唯一方法。思科端以及VMware ESXi的巨型幀配置和步驟在[UCS with VMware ESXi End-to-End Jumbo MTU配置示例中說明](#)。

[UCS FCoE上行鏈路配置](#)示例顯示了UCS和Nexus 5000配置。有關基本Nexus 5000配置的大綱，請參閱參考文檔中的附錄A。

[為Cisco UCS刀片設定FCoE連線](#)，重點為FCoE的UCS配置。[帶FCoE NPV連線的UCS的Nexus 5000 NPV FCoE配置示例](#)重點介紹Nexus配置。

## 測試端到端巨型幀

大多數現代作業系統都能夠通過簡單的網際網路控制消息協定(ICMP)測試來測試適當的巨型幀配置。

### 計算

9000位元組 — 不含選項的IP標頭 ( 20位元組 ) — ICMP標頭 ( 8位元組 ) = 8972位元組資料

常用作業系統中的命令

### Linux

```
ping a.b.c.d -M do -s 8972
```

### Microsoft Windows

```
ping -f -l 8972 a.b.c.d
```

### ESXi

```
vmkping -d -s 8972 a.b.c.d
```

## 緩衝區相關問題

緩衝和其他延遲相關問題是FlexPod環境中常見的效能下降原因之一。並非所有報告為延遲的問題都源於實際緩衝問題，許多測量值可能表示端到端延遲。例如，在NFS的情況下，可能需要報告的時間段才能成功讀取/寫入到儲存，而不是實際網路延遲。

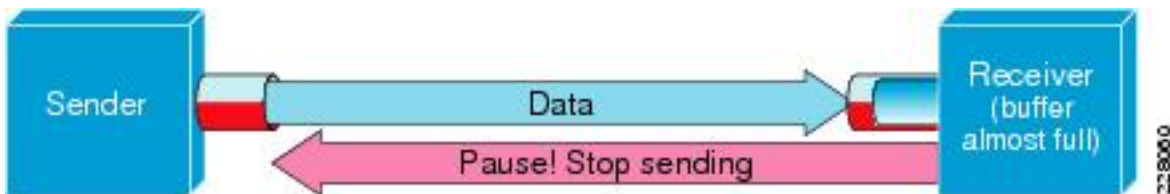
擁塞是最常見的緩衝原因。在第2層世界中，擁塞會導致幀緩衝甚至尾部丟棄。為了避免在擁塞期間丟棄，引入了IEEE 802.3x暫停訊框和優先順序流量控制(PFC)。兩者都依賴於要求端點在擁塞持續

期間保持短時間的傳輸。這可能是由於網路擁塞（用大量資料淹沒收到的資料）或需要經過優先順序的幀（如FCoE）造成的。

### 流量控制 — 802.3x

若要驗證哪些介面啟用了流量控制，請輸入 `show interface flowcontrol` 命令。必須遵循儲存供應商關於啟用流量控制的建議。

此處顯示說明802.3x流量控制如何運作的圖示。

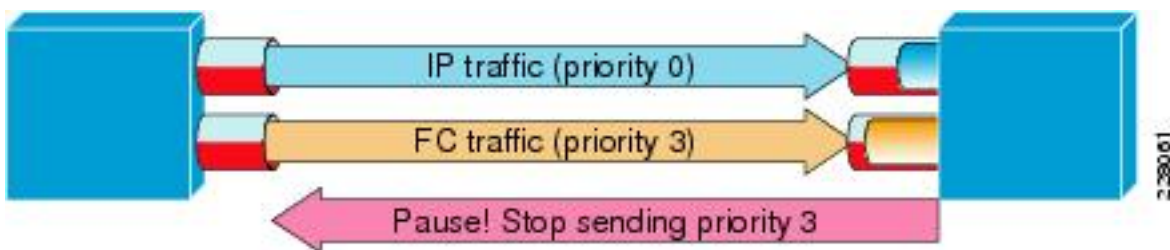


### PFC - 802.1Qbb

並非所有設定都需要PFC，但建議大多數情況下使用PFC。為了驗證哪些介面啟用了PFC，請輸入 `show interface priority-flow-control | i On` 命令可在UCS的NX-OS和Nexus 5000上運行。

FI和Nexus 5000之間的介面應在該清單中可見。如果不是，則需要驗證QoS配置。為了利用PFC，QoS需要端到端保持一致。要檢查PFC為什麼沒有在特定介面上啟動，請輸入 `show system internal dcbx log interface ethernet x/y` 命令以獲取資料中心橋接功能交換協定(DCBX)日誌。

此處顯示暫停訊框如何與PFC配合使用的圖示。



`show interface priority-flow-control` 命令允許管理員觀察優先順序暫停幀的每QoS類行為。

以下是範例：

```
bdsol-6120-05-A(nxos)# show queuing interface ethernet 1/1 | i prio
Per-priority-pause status : Rx (Inactive), Tx (Inactive)
Per-priority-pause status : Rx (Inactive), Tx (Active)
```

此輸出顯示，在第二類中，裝置僅傳輸(TX)PPP幀。

在這種情況下，Ethernet 1/1是面向IOM的埠，雖然整個埠不會啟用PFC，但它可能會處理FEX埠的PPP幀。

```
bdsol-6120-05-A(nxos)# show interface e1/1 priority-flow-control
=====
Port Mode Oper(VL bmap) RxPPP TxPPP
```



```
=====  
Ethernet1/1 Auto Off 4885 3709920
```

在這種情況下，會涉及FEX介面。

```
bdsol-6120-05-A(nxos)# show interface priority-flow-control | egrep .*\/.*\  
Ethernet1/1/1 Auto Off 0 0  
Ethernet1/1/2 Auto Off 0 0  
Ethernet1/1/3 Auto Off 0 0  
Ethernet1/1/4 Auto Off 0 0  
Ethernet1/1/5 Auto On (8) 8202210 15038419  
Ethernet1/1/6 Auto On (8) 0 1073455  
Ethernet1/1/7 Auto Off 0 0  
Ethernet1/1/8 Auto On (8) 0 3956077  
Ethernet1/1/9 Auto Off 0 0
```

也可通過show fex Detail檢查相關的FEX埠，其中X是機箱編號。

```
bdsol-6120-05-A(nxos)# show fex 1 detail | section "Fex Port"  
Fex Port State Fabric Port  
Eth1/1/1 Down Eth1/1  
Eth1/1/2 Down Eth1/2  
Eth1/1/3 Down None  
Eth1/1/4 Down None  
Eth1/1/5 Up Eth1/1  
Eth1/1/6 Up Eth1/2  
Eth1/1/7 Down None  
Eth1/1/8 Up Eth1/2  
Eth1/1/9 Up Eth1/2
```

如需暫停機制的詳細資訊，請參閱以下檔案。

- [乙太網路光纖通道操作](#)
- [統一交換矩陣白皮書 — 乙太網光纖通道\(FCoE\)](#)

## 排隊丟棄

Nexus 5000和UCS NX-OS均會跟蹤因每個QOS組排隊而產生的入口丟棄。例如：

```
bdsol-6120-05-A(nxos)# show queuing interface  
Ethernet1/1 queuing information:  
TX Queuing  
  qos-group sched-type oper-bandwidth  
    0         WRR         50  
    1         WRR         50  
RX Queuing  
  qos-group 0  
  q-size: 243200, HW MTU: 9280 (9216 configured)  
  drop-type: drop, xon: 0, xoff: 243200  
Statistics:  
  Pkts received over the port           : 31051574  
  Ucast pkts sent to the cross-bar      : 30272680  
  Mcast pkts sent to the cross-bar      : 778894  
  Ucast pkts received from the cross-bar : 27988565  
  Pkts sent to the port                 : 34600961  
  Pkts discarded on ingress             : 0  
  Per-priority-pause status            : Rx (Inactive), Tx (Active)
```

入口丟棄應僅在配置為允許丟棄的隊列中發生。

可能因為以下原因而發生輸入佇列捨棄：

- 交換器連線埠分析器(SPAN)/監控作業階段已在某些介面上啟用(請參閱思科錯誤ID [CSCur25521](#))
- 來自另一個介面的背壓，啟用時通常出現暫停幀
- 發往CPU的流量

## 驅動程式問題

Cisco為UCS提供兩種作業系統驅動程式：enic和fnic。Enic負責乙太網連線，fnic負責光纖通道和FCoE連線。非常重要，引擎和fnic驅動程式與UCS互操作性矩陣中指定的完全一致。錯誤驅動程式引起的問題包括丟包、延遲增加、啟動過程延長或完全缺乏連線。

## 介面卡資訊

思科提供的介面卡可以很好地測量通過流量和丟棄流量。此示例說明如何連線到機箱X、伺服器Y和介面卡Z。

```
bdsol-6248-06-A# connect adapter X/Y/Z
adapter X/Y/Z # connect
No entry for terminal type "dumb";
using dumb terminal settings.
```

管理員可以從此處登入到效能監視中心(MCP)設施。

```
adapter 1/2/1 (top):1# attach-mcp
No entry for terminal type "dumb";
using dumb terminal settings
```

MCP工具允許您監控每個邏輯介面(LIF)的流量使用情況。

```
adapter 1/2/1 (mcp):1# vnic
(...)
```

```
-----
          v n i c                l i f                v i f
id  name      type      bb:dd.f state  lif state  uif  ucsm  idx  vlan  state
-----
 13 vnic_1     enet     06:00.0 UP      2  UP     =>0  834   20  3709  UP
 14 vnic_2     fc       07:00.0 UP      3  UP     =>0  836   17   970  UP
```

機箱1、伺服器1和介面卡1有兩個與虛擬介面(虛擬乙太網或虛擬光纖通道)834和836關聯的虛擬網路介面卡(VNIC)。這兩個介面卡具有數字2和3。可以檢查LIF 2和3的統計資訊，如下所示：

```
adapter 1/2/1 (mcp):3# lifstats 2
      DELTA          TOTAL DESCRIPTION
      4              4 Tx unicast frames without error
53999          53999 Tx multicast frames without error
69489          69489 Tx broadcast frames without error
      500              500 Tx unicast bytes without error
8361780        8361780 Tx multicast bytes without error
22309578       22309578 Tx broadcast bytes without error
      2                2 Rx unicast frames without error
2791371        2791371 Rx multicast frames without error
4595548        4595548 Rx broadcast frames without error
```

```

188 188 Rx unicast bytes without error
260068999 260068999 Rx multicast bytes without error
514082967 514082967 Rx broadcast bytes without error
3668331 3668331 Rx frames len == 64
2485417 2485417 Rx frames 64 < len <= 127
655185 655185 Rx frames 128 <= len <= 255
434424 434424 Rx frames 256 <= len <= 511
143564 143564 Rx frames 512 <= len <= 1023
94.599bps Tx rate
2.631kbps Rx rate

```

必須注意的是，UCS管理員將獲得total和delta ( 兩個後續的提升統計值執行之間 ) 列，以及每個LIF的當前流量負載和有關可能已發生的任何錯誤的資訊。

上一個示例顯示介面在非常小的負載下沒有任何錯誤。此示例顯示了不同的伺服器。

```
adapter 4/4/1 (mcp):2# lifstats 2
```

```

DELTA TOTAL DESCRIPTION
127927993 127927993 Tx unicast frames without error
273955 273955 Tx multicast frames without error
122540 122540 Tx broadcast frames without error
50648286058 50648286058 Tx unicast bytes without error
40207322 40207322 Tx multicast bytes without error
13984837 13984837 Tx broadcast bytes without error

28008032 28008032 Tx TSO frames
262357491 262357491 Rx unicast frames without error
55256866 55256866 Rx multicast frames without error
51088959 51088959 Rx broadcast frames without error
286578757623 286578757623 Rx unicast bytes without error
4998435976 4998435976 Rx multicast bytes without error
7657961343 7657961343 Rx broadcast bytes without error

96 96 Rx rq drop pkts (no bufs or rq disabled)

136256 136256 Rx rq drop bytes (no bufs or rq disabled)
5245223 5245223 Rx frames len == 64
136998234 136998234 Rx frames 64 < len <= 127
9787080 9787080 Rx frames 128 <= len <= 255
14176908 14176908 Rx frames 256 <= len <= 511
11318174 11318174 Rx frames 512 <= len <= 1023
61181991 61181991 Rx frames 1024 <= len <= 1518
129995706 129995706 Rx frames len > 1518

136.241kbps Tx rate

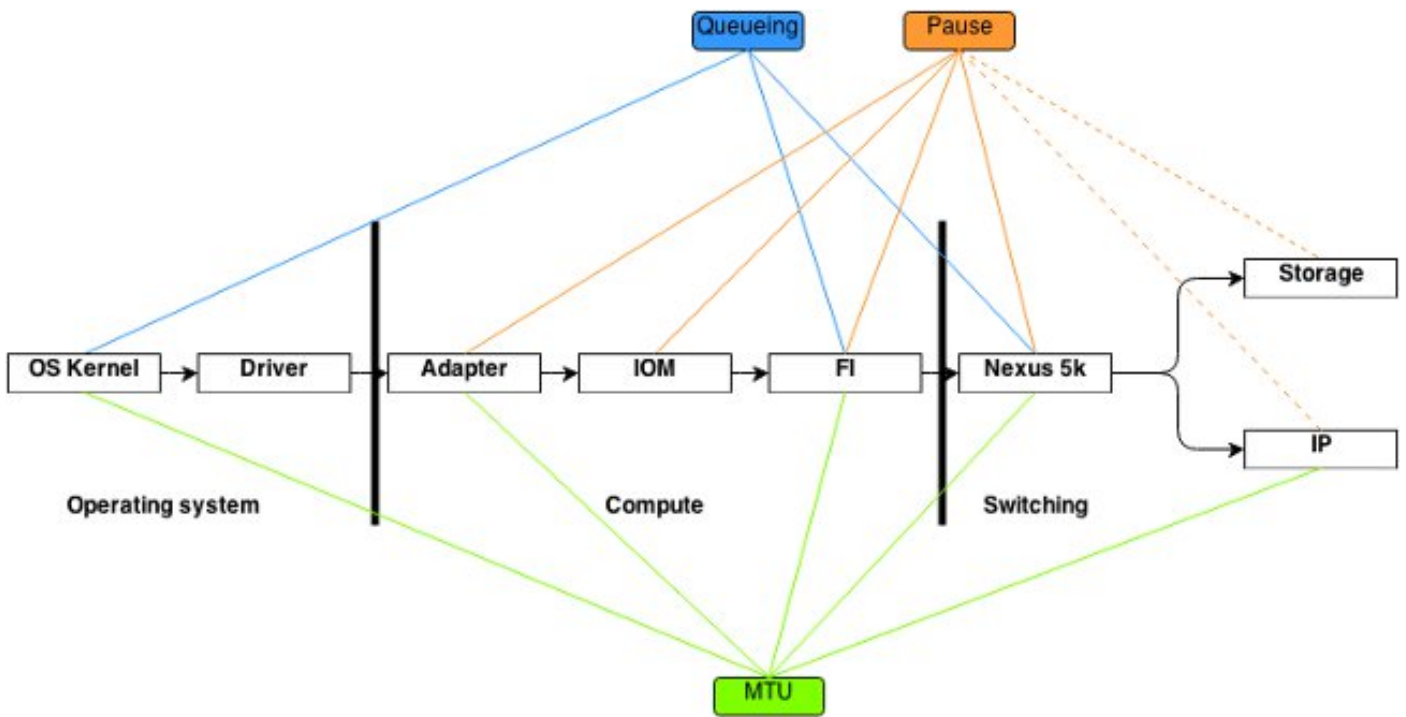
784.185kbps Rx rate

```

兩個有趣的資訊位顯示，介面卡丟棄96個幀，原因是缺少緩衝或緩衝已禁用，並且正在處理TCP段解除安裝(TSO)段。

## 邏輯封包流

此處顯示的圖表概述了FlexPod環境中的邏輯資料包流。



此圖表示幀在通過FlexPod環境的過程中通過的元件的細分。它不反映任何塊的複雜性，只是一種記住特定功能應配置和驗證位置的方式。

## 輸入/輸出模組

如邏輯資料包流圖所示，輸入/輸出模組(IOM)是通過UCS的所有通訊中間的一個元件。要連線到機箱X中的IOM，請輸入**connect iom x**命令。

以下是其他幾個有用的命令：

- 拓撲資訊 — **show platform software [woodside|redwood] sts**命令從IOM的角度顯示拓撲資訊。



```

# show platform software woodside loss

```

Port	SMD			Port Extra Drop	S8 Loss Counters	Cos_u														
	Tx Pkts	Rx Pkts	Errors			Counters	COS													
							0	1	2	3	4	5	6	7	8	9				
0-M12	0	82	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0-H123	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

由於底層基礎架構的工作方式，計數器只針對在執行兩個命令之間遇到任何丟失的介面顯示。在以下示例中，您會看到NI2介面接收了82個暫停幀，並且28個暫停幀已傳輸到介面HI23，您知道該介面已連線到刀片3。

## 設計注意事項

FlexPod允許靈活配置和設定儲存和資料網路。靈活性也帶來了額外的挑戰。請務必遵循最佳實踐文檔和思科驗證設計(CVD):

- CVD - FlexPod [部署指南](#)
- NetApp儲存最佳實踐 (不特定於Flexpod) — [Cisco Unified Computing System\(UCS\)儲存連線選項和NetApp儲存最佳實踐](#)

## 埠速度選擇和埠通道注意事項

TAC工程師發現的一個常見問題是由於選擇了最佳實踐文檔中提到的1 Gbit乙太網路而不是10 Gbit乙太網路，導致連結過度使用。舉一個尖銳的例子，在十個1 Gbit鏈路上單流效能不會優於一個10 Gbit鏈路。在埠通道中，單個流可以經過單個鏈路。

若要瞭解Nexus和/或FI的NX-OS使用的是哪種負載均衡方法，請輸入**show port-channel load-balance**命令。管理員還可以瞭解埠通道中的哪個介面將被選擇作為資料包或幀的傳出介面。以下為兩台主機之間VLAN49上幀的簡單示例：

```

show port-channel load-balance forwarding-path interface port-channel 928 vlan 49
src-mac 70ca.9bce.ee24 dst-mac 8478.ac55.2fc2
Missing params will be substituted by 0's.
Load-balance Algorithm on switch: source-dest-ip
crc8_hash: 2      Outgoing port id: Ethernet1/27
Param(s) used to calculate load-balance:
  dst-mac: 8478.ac55.2fc2
  src-mac: 70ca.9bce.ee24

```

## 儲存特定問題

前面討論的問題對於資料和儲存網路來說都很常見。為了完整起見，還提到了特定於儲存區域網路(SAN)的效能問題。儲存協定的構建具有恢復能力，並且多路徑功能仍然得到增強。隨著非對稱邏輯單元分配(ALUA)和多路徑IO(MPIO)等技術的出現，為管理員提供了更大的靈活性和更多選項。

## 儲存位置

另一個考慮因素是存放位置。FlexPod設計要求在Nexus交換機上連線儲存。直接連線的儲存不符合CVD。如果遵循最佳實踐，則支援帶有直接連線儲存的設計。同時，這些設計並非嚴格的FlexPod。

## 最佳路徑選取

這在技術上不是思科的問題，因為這些選項大多數對思科裝置都是透明的。選擇並堅持最佳路徑是一個常見問題。現代設備特定模組(DSM)可以有多個路徑，需要根據一定的標準選擇最佳路徑，以提供恢復能力和負載平衡。此螢幕截圖顯示了適用於Microsoft Windows的NetApp DSM和負載平衡選項的四個可用路徑。

The screenshot shows the NetApp DSM configuration interface. The top part displays a table of paths for Disk0, and the bottom part shows the 'Data ONTAP(R) DSM Properties' dialog box with the 'MPIO' tab selected.

Disk ID	Path ID	Operational State	Admin State	Initiator Name	Initiator Address
Disk0	01000101	Active/Optimized	Enabled	com.ciscosystem...	20:00:00:25:b5:00:a...
Disk0	02000002	Active/Non-Optimized	Enabled	com.ciscosystem...	20:00:00:25:b5:00:b...
Disk0	01000001	Active/Optimized	Enabled	com.ciscosystem...	20:00:00:25:b5:00:a...
Disk0	02000102	Active/Non-Optimized	Enabled	com.ciscosystem...	20:00:00:25:b5:00:b...

**Data ONTAP(R) DSM Properties**

Default Load Balance Property

- Auto Assign
- Failover Only
- Round Robin
- Round Robin with Subset
- Least Weighted Paths
- Least Queue Depth

應根據與儲存供應商的討論來選擇建議的設定。這些設定可能會影響效能問題。TAC可能會要求您執行的典型測試是僅通過交換矩陣A或交換矩陣B進行的讀/寫測試。這通常允許您將效能問題縮小到本文檔的「常見問題」部分中討論的情況。

## 虛擬機器和虛擬機器監控程式流量共用

此點特定於計算元件，無論供應商是誰。從計算角度來看，為虛擬機器管理程式設定儲存網路的簡單方法是建立兩個主機匯流排介面卡(HBA)（每個光纖一個），並通過這兩個介面運行啟動LUN流量和虛擬機器(VM)儲存流量。始終建議拆分引導LUN流量和VM儲存流量。這樣可以實現更好的效能，並且還允許在兩種流量之間進行邏輯拆分。有關示例，請參見「已知問題」部分。

## 疑難排解提示

### 縮小問題範圍

如同快速故障排除一樣，縮小問題範圍並提出正確的問題非常重要。

- 哪些裝置/應用/VM受 ( /未受 ) 影響？
- 哪個儲存控制器受 ( /未受 ) 影響？
- 哪些路徑受(/not)影響？
- 問題 ( /不是 ) 出現的頻率？

## 思科

### 計數器限制

本文檔中討論了ASIC隊列計數器。計數器也會提供某個時間點的檢視，因此監控計數器的增加非常重要。某些計數器無法被設計清除。例如，前面提到的carmel ASIC。

為了給出一個明確的示例，介面上存在CRC或丟棄可能並不理想，但可以預期它們的值非零。計數器可能會在某個時間點上升，可能是在轉換或初始設定期間。因此，必須注意計數器的增加，以及最後一次清除計數器的時間。

### 控制平面注意事項

雖然檢查計數器很有用，但必須知道某些資料平面問題可能無法找到控制平面計數器和工具的簡單反射。舉個例子，Ethanalyzer是一個非常有用的工具，可在UCS和Nexus 5000上使用。但是，它只能捕獲控制平面流量。流量擷取是TAC經常要求的，尤其是當不清楚錯誤的位置時。

### 捕獲流量

在終端主機上進行可靠的流量擷取，可以很快發現效能問題並將其縮小。Nexus 5000和UCS都提供流量SPAN。具體來說，UCS的SPAN控制特定HBA和交換矩陣端選項非常有用。要在監控UCS上的會話時瞭解有關流量捕獲功能的詳細資訊，請參閱以下參考資料：

- [適用於實體和虛擬配接器的UCS流量分析 \( 視訊 \)](#)
- [Cisco UCS Manager GUI配置指南 — 監控流量](#)

## NetApp

NetApp提供了一整套實用工具，以便對其儲存控制器進行故障排除，其中包括：

- perfstat — 一個非常有用的實用程式，通常用於NetApp支援人員
- systat — 提供有關檔案伺服器的繁忙程度以及檔案伺服器正在執行的任務的資訊 — [NetApp支援庫](#)

最常用的命令包括：

- ```
sysstat -x 2
```
- ```
sysstat -M 2
```



以下是在sysstat -x 2輸出中查詢的一些內容，這可能表示NetApp陣列或磁碟超載：

- 持續CPU列，包括:或F
- 持續HDD使用率列超過20%

本文描述如何配置NetApp:[NetApp乙太網儲存最佳實踐](#)。

- VLAN標籤
- VLAN中繼
- 巨量MTU
- IP雜湊
- 禁用FlowControl

## VMware

ESXi提供安全殼層(SSH)訪問，您可以通過它進行故障排除。提供給管理員的最有用的工具包括esxtop和perfmon。

- esxtop — 與Linux/BSD top非常相似，它允許使用者監控與即時效能相關的引數  
[使用esxtop確定ESX/ESXi的儲存效能問題](#)
- perfmon — 允許使用者對Microsoft Windows虛擬機器(VM)進行故障排除  
[收集Windows Perfmon日誌資料以診斷虛擬機器效能問題](#)
- 收集ESXi上的診斷捆綁包 — [使用vSphere客戶端\(653\)收集VMware ESX/ESXi的診斷資訊](#)
- Cisco B系列伺服器的VMware vSwitch負載平衡要求 — [使用UCS 6100系列交換矩陣互聯的Cisco UCS B200 M1/M2刀鋒伺服器不支援基於IP雜湊的路由](#)

## 已知問題和增強功能

- 思科錯誤ID [CSCuj86736](#) — 使用被動雙軸電纜時，CRC錯誤可能會增加。這是在Nexus 5000未最佳化DFE時造成的。輸入show hardware internal carmel eye命令，以驗證「Eye height」引數是否高於100 mv。這一點在版本5.2(1)N1(7)和7.0(4)N1(1)中已修復。
- 思科錯誤ID [CSCuo76425](#) — 與上一個錯誤類似，也存在於UCS交換矩陣互聯上。版本2.2(3a)已修復此問題。
- 思科錯誤ID [CSCuo76425](#) — 與[CSCuj86736](#)相同，但UCS交換矩陣互聯除外。
- 思科錯誤ID [CSCup40056](#) — 計時問題，原因是使用虛擬光纖通道介面卡[統一計算系統虛擬機器即時遷移失敗中所述的VM流量共用啟動流量](#)。
- 檢測並避免漏電緩慢 — FC和FCoE經常受到漏電緩慢的影響。NX-OS版本7.0(0)N1(1)引入了檢測和避免此漏洞的方法。如需詳細瞭解[Cisco Nexus 5500系列NX-OS介面配置指南](#)和[緩排裝置檢測和擁塞迴避中的功能](#)。
- 思科漏洞ID [CSCuj81245](#) — 基於PALO的卡 ( VIC1240和其他 ) 存在限制，導致FC中止。
- 思科錯誤ID [CSCuh61202](#) — 升級至版本2.1(3)後，UCS韌體FC中止，且可見多個其他問題。
- 思科錯誤ID [CSCtw91018](#) — 在基於PALO的單個介面卡上混合使用VNIC的MTU設定可能導致某些流量類別發生耗竭。
- 思科錯誤ID [CSCuq40256](#) — 將導致從交換矩陣互聯向下到伺服器介面卡的鏈路上禁用PFC。這將導致各種問題，首先表現為光纖通道中止，然後在儲存端報告無序幀。可能會報告儲存斷開和其他效能問題。

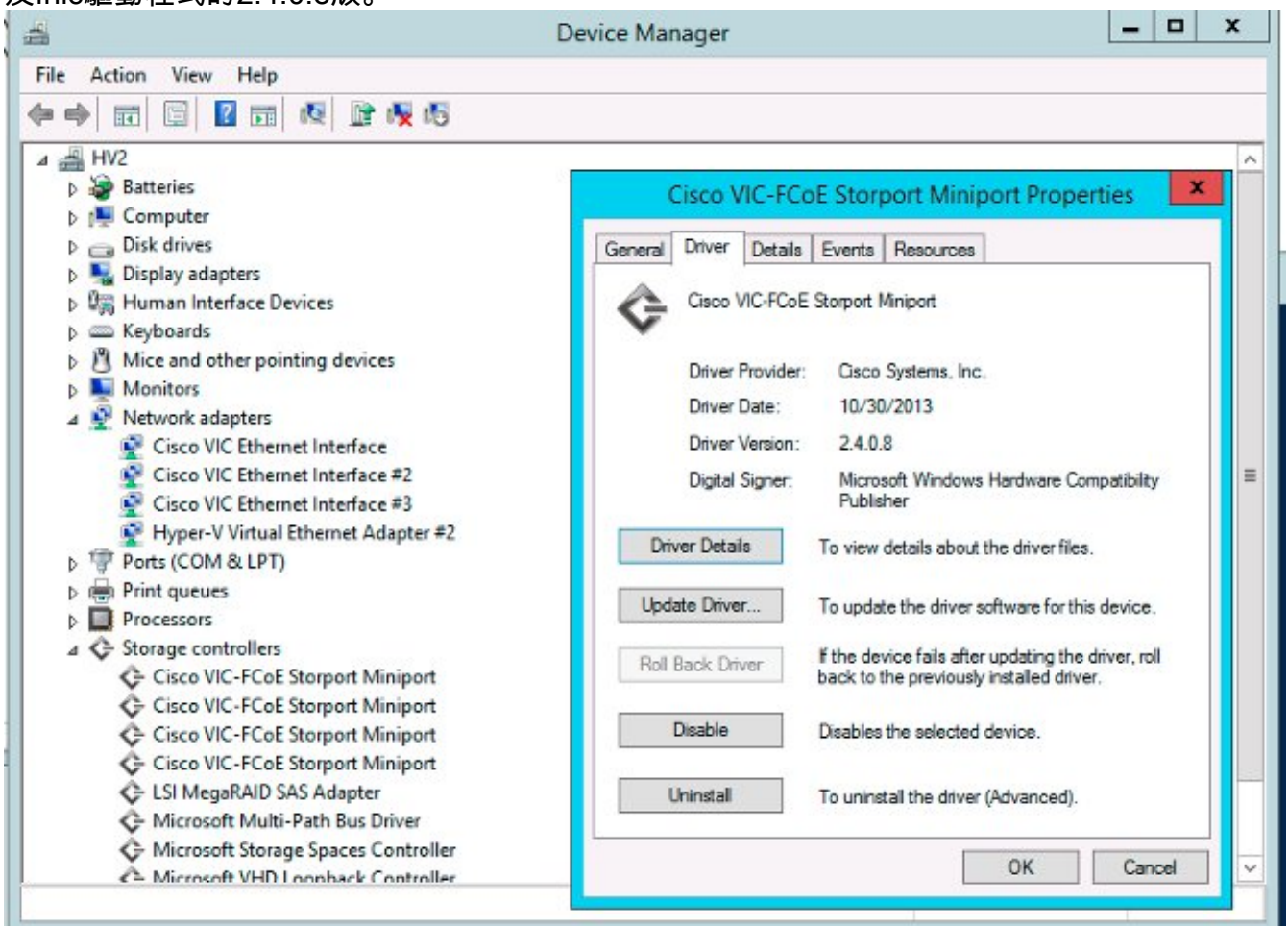
## TAC案例

在許多情況下，TAC工程師會要求您收集一些基本資訊，然後才能開始調查。

- 拓撲圖 — 包括埠號和線速，這是絕對必要的。
- [UCSM技術支援 — 收集技術支援檔案 \( B和C系列 \) 的可視指南。](#)
- UCS機箱技術支援，適用於遇到問題的機箱 — 請參閱前面的連結。
- Nexus 5000技術支援和UCS與NetApp之間的任何其他網路裝置 — [show tech-support details命令的Redirecting輸出。](#)
- 兩個FI上show queuing interface命令的輸出。  

```
connect nxos A|B
show queuing interface | no-more
show interface priority-flow-control | no-more
show interface flowcontrol | no-more.
```
- ESXi上的主機驅動程式版本執行 — 輸入以下命令：vmkload\_mod -s引擎vmkload\_mod -s fnic
- Linux -  

```
dmesg | egrep -i 'enic|fnic'
```
- Windows — 檢查「裝置管理器」中的驅動程式版本。來自Window 2012 R2的示例顯示了三個Cisco VIC乙太網介面和四個VIC FCoE微型埠介面（還負責光纖通道，而不僅僅是FCoE）以及fnic驅動程式的2.4.0.8版。



## 意見回饋

使用「意見回饋」按鈕提供有關本文或您的體驗的回饋。我們將在情況發生時和收到回饋後不斷更新本檔案。