# Häufige Leistungsprobleme bei FlexPod

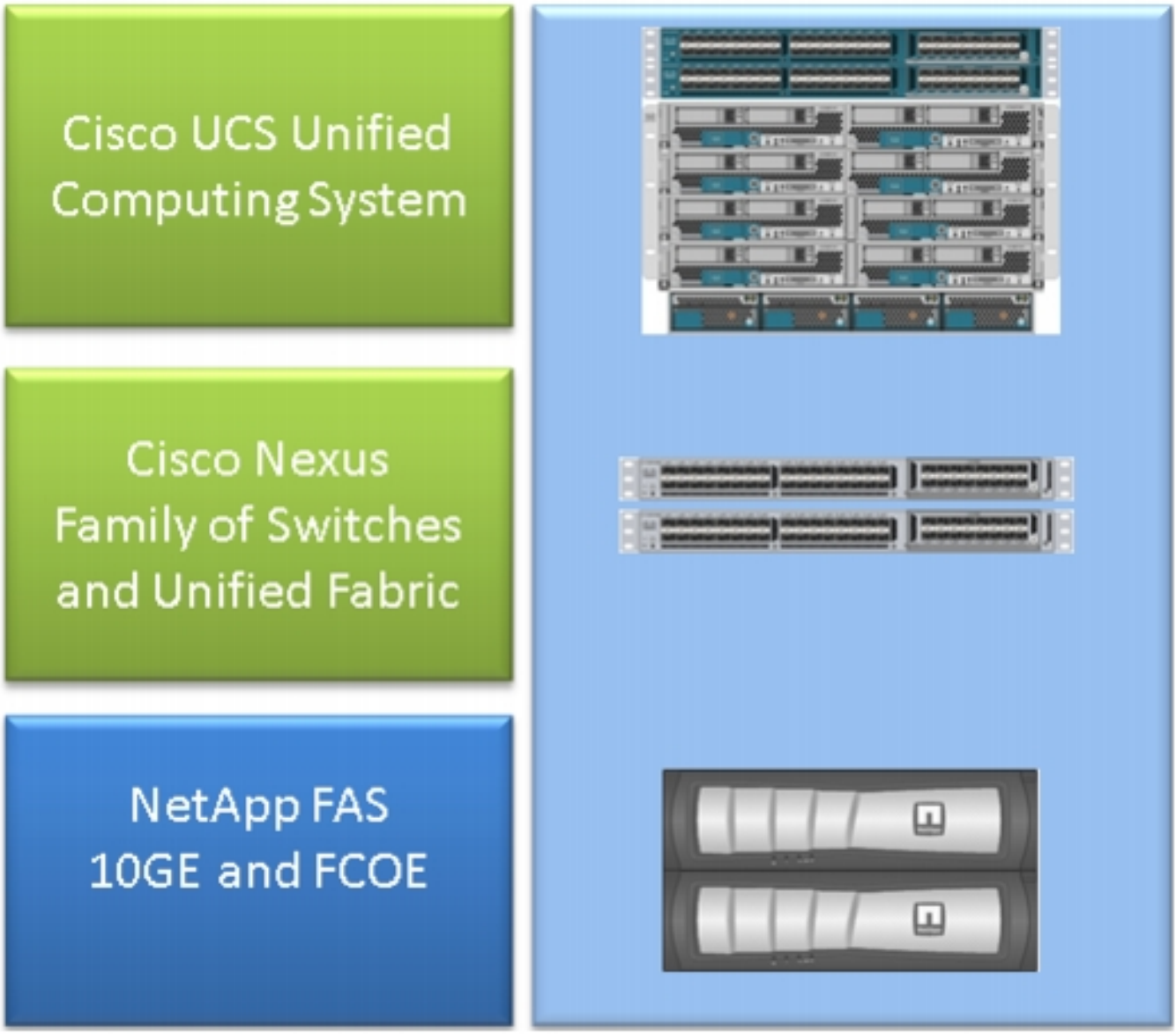
## Inhalt

# Einführung

In diesem Dokument werden häufige Leistungsprobleme in FlexPod-Umgebungen beschrieben, eine Methode zur Problembehebung bereitgestellt und Schritte zur Problemvermeidung beschrieben. Sie ist als Ausgangspunkt für Kunden gedacht, die eine Fehlerbehebung für die Leistung in einer FlexPod-Umgebung durchführen möchten. Dieses Dokument wurde aufgrund von Problemen erstellt, die das Team des Data Center Solutions Technical Assistance Center (TAC) in den letzten Monaten festgestellt hatte.

# Konzeptionelle Übersicht über FlexPod

Ein FlexPod besteht aus einem Unified Computing System (UCS)-Computer, der über einen Nexus-Switch mit Storage- und IP-Netzwerken von NetApp verbunden ist.




Das häufigste FlexPod besteht aus einem Chassis der Cisco UCS B-Serie, das über Fabric Interconnects (FIs) mit Nexus 5500-Switches mit NetApp-Filern verbunden ist. Eine andere Lösung, FlexPod Express genannt, verwendet ein Chassis der UCS C-Serie, das mit Nexus 3000-Switches verbunden ist. In diesem Dokument werden die gängigsten FlexPod-Lösungen

vorgestellt.

# Überlegungen zur Leistung

In komplexen Umgebungen mit mehreren verantwortlichen Parteien, wie in der Regel in einem FlexPod zu sehen ist, müssen Sie mehrere Aspekte berücksichtigen, um das Problem zu beheben. Typische Leistungsprobleme in Layer-2- und IP-Netzwerken ergeben sich aus:

- Paket- oder Frame-Verlust - Datenverluste wirken sich negativ auf die Anwendungsleistung aus.
- Pufferung: Wenn ein Paket oder Frame zu viel Zeit in eine Warteschlange bzw. einen Puffer investiert, kann dies von Anwendungen, insbesondere im Fall von Speichernetzwerken, zu Auswirkungen auf die Leistung führen. Probleme mit Latenz, Neuordnung und Normalisierungsproblemen fallen unter diese Kategorie.
- MTU-Probleme und -Fragmentierung - ein häufiges Problem, wenn Sie eine höhere Leistung erreichen. Probleme im Zusammenhang mit Fragmentierung und MTU-Inkonsistenz fallen in diese Kategorie.

## Umgebung

Es ist wichtig, die Umgebung zu kennen, in der die Leistung gemessen wird. Fragen zum Speichertyp und -protokoll sowie zum Betriebssystem und Speicherort des betroffenen Servers sollten zur ordnungsgemäßen Eingrenzung des Problems gestellt werden. Ein Topologiediagramm, das die Konnektivität beschreibt, ist das absolute Minimum.

## Messung

Man muss wissen, was gemessen wird und wie es gemessen wird. Bestimmte Anwendungen sowie die meisten Storage- und Hypervisor-Anbieter liefern Messwerte, die auf die Leistung/Integrität des Systems hinweisen. Diese Messungen sind ein guter Ausgangspunkt, da sie die meisten Fehlerbehebungsmethoden nicht ersetzen.

Beispiel: Eine NFS-Speicherlatenzmessung (Network File System) im Hypervisor könnte darauf hindeuten, dass die Leistung abnimmt. Allein jedoch impliziert sie keine Auswirkungen auf das Netzwerk. Bei einem NFS kann ein einfacher Ping vom Host zum NFS-Speicher-IP-Netzwerk angeben, ob das Netzwerk schuld ist.

## Ausgangswert

Dieser Punkt kann nicht genug betont werden, besonders wenn Sie ein TAC-Ticket eröffnen. Um anzuzeigen, dass die Leistung nicht zufriedenstellend ist, muss der gemessene Parameter angegeben werden. Dazu gehört der erwartete **und** getestete Wert. Im Idealfall sollten Sie vorherige Daten **und** die Testmethodik anzeigen, die zum Erreichen dieser Daten verwendet wurde.

Als Beispiel: Die bei Tests erzielte Latenz von 10 ms, bei der nur Schreibzugriff von einem Initiator auf eine LUN-Nummer (Logical Unit Number) möglich ist, ist kein Hinweis darauf, wie lange die

Latenz für ein voll ausgelastetes System eigentlich sein sollte.

# Leistungsprobleme in einem FlexPod

Da dieses Dokument als Referenz für die meisten FlexPod-Umgebungen dient, werden nur die häufigsten Probleme beschrieben, die das für Rechenzentrumslösungen zuständige TAC-Team sieht.

## Häufige Probleme

In diesem Abschnitt werden Probleme beschrieben, die häufig mit Storage- und IP/Layer-2-Netzwerken verbunden sind.

### Frame- und Paketverlust

Frame- und Paketverluste sind der häufigste Faktor, der sich auf die Leistung auswirkt. Eine der gängigsten Stellen, um nach Anzeichen für ein Problem zu suchen, ist die Schnittstellenebene. Geben Sie in der CLI des Nexus 5000 oder des UCS Nexus Operating System (NX-OS) die **Show-Schnittstelle ein. | Sek. "ist aktiv"** Befehl **| egrep ^(Eth|fc)|discard|drop|CRC**. Bei Schnittstellen, die aktiv sind, werden der Name und die Zähler und Verwerfungen aufgelistet. Ebenso wird eine große Übersicht angezeigt, wenn Sie den Befehl **show interface counters error** (Fehlerstatistiken für alle Schnittstellen anzeigen) eingeben.

#### Ethernet-Welt

Es ist wichtig zu wissen, dass Zähler bei Nicht-0 möglicherweise kein Problem anzeigen. In bestimmten Szenarien wurden diese Zähler möglicherweise bei der Ersteinrichtung oder bei vorherigen betrieblichen Änderungen ausgelöst. Eine Erhöhung der Zähler sollte überwacht werden.

Sie können auch Zähler der ASIC-Ebene sammeln, was eher indikativ sein kann. Insbesondere für den CRC-Fehler (Cyclical Redundancy Check) an Schnittstellen ist ein TAC-Favoritenbefehl die **Anzeige von hardwareinternem Carmel crc**. Carmel ist der Name des ASIC, der für die Weiterleitung auf Port-Ebene verantwortlich ist.

Eine ähnliche Ausgabe kann von FIs der Serie 6100 oder Nexus 5600-Switches pro Port übernommen werden. Geben Sie für den FI 6100, den gatos ASIC, den folgenden Befehl ein:

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

Geben Sie für den Nexus 5600 von der bigsur-ASIC den folgenden Befehl ein:

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

Der Befehl für Carmel-ASIC zeigt an, wo CRC-Pakete empfangen und an welche weitergeleitet wurden, und vor allem, ob sie gestapelt wurden.

Da der Nexus 5000- und UCS NX-OS-Betrieb per Cut-Through ausgeführt wird, werden Frames im Switching-Modus mit falscher Frame Check Sequence (FCS) erst vor der Weiterleitung gestapelt. Es ist wichtig herauszufinden, woher die beschädigten Frames kommen.

```
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 |
```

Dieses Beispiel zeigt stomplizierte Pakete, die von Eth 1/17 und Eth 1/18 stammen, was ein Uplink zum Nexus 5000 ist. Man kann davon ausgehen, dass diese Frames später nach Eth 1/34 gesendet wurden, wie Eth 1/17 + Eth 1/18 rx Stomp = Eth 1/34 tx Stomp.

Ein ähnlicher Blick auf den Nexus 5000 zeigt Folgendes:

```
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 |        --- |        --- |
--- |
```

Diese Ausgabe zeigt CRCs an, die auf zwei Verbindungen empfangen und vor der Weiterleitung als Stempel markiert wurden. Weitere Informationen finden Sie im Nexus 5000-Leitfaden zur Fehlerbehebung.

#### Fibre Channel-Welt

Eine einfache Methode zur Suche nach Drops (Diskrepanzen, Fehler, CRCs, B2B-Kreditschöpfung) ist der Befehl **show interface counter fc**.

Dieser Befehl, der auf dem Nexus 5000 und Fabric Interconnect verfügbar ist, gibt einen guten Hinweis darauf, was in der Fibre Channel-Welt geschieht.

Beispiel:

```
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
(...)
```

Diese Schnittstelle ist nicht belegt, und die Ausgabe zeigt an, dass keine Rückwürfe oder Fehler aufgetreten sind.

Außerdem wurden B2B-Kreditübergänge von 0 hervorgehoben. Aufgrund der Cisco Bug-IDs CSCue80063 und CSCut08353 können diese Zähler nicht vertrauenswürdig sein. Sie funktionieren auf Cisco MDS, aber nicht auf dem UCS der Nexus 500-Plattformen. Sie können auch die Cisco Bug-ID CSCsz95889 überprüfen.

Ähnlich wie Carmel in Ethernet World für Fibre Channel (FC) kann die FC-MAC-Einrichtung verwendet werden. Geben Sie beispielsweise für Port fc2/1 den Befehl **show hardware internal fc-mac 2 port 1 statistics** ein. Die dargestellten Zähler haben das hexadezimale Format.

```
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
```

In der Ausgabe werden 15 Rückwürfe angezeigt. Dies kann mit FCP_CNTR_PIF_RX_DROP abgeglichen werden, der auf 0xf (15 im Dezimalformat) gezählt wurde. Diese Informationen können wiederum mit Informationen von FWM (Forwarding Manager) korreliert werden.

```
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
```

Dadurch erhält der Administrator jedoch die Anzahl der Tropfen und die entsprechende ASIC-Nummer. Der Abruf von Informationen über den Grund für den Ausfall des ASIC muss abgefragt werden.

```
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]
```

In diesem Fall wurde der Datenverkehr von der ACL (Ingress Access Control List) verworfen, in der Regel in der FC-Welt - Zoning.

## MTU-Abweichung

In FlexPod-Umgebungen ist es wichtig, die MTU-Einstellung (Maximum Transition Unit) für alle Anwendungen und Protokolle einzubinden, für die sie erforderlich ist. In den meisten Umgebungen sind dies Fibre Channel over Ethernet (FCoE) und Jumbo Frames.

Darüber hinaus ist bei einer Fragmentierung eine verminderte Leistung zu erwarten. Bei Protokollen wie Network File System (NFS) und Internet Small Computer System Interface (iSCSI) ist es wichtig, die End-to-End-MTU (Maximum Transmission Unit) und die TCP Maximum Segment Size (MSS) zu testen und zu testen.

Unabhängig davon, ob Sie die Fehlerbehebung für Jumbo Frames oder FCoE durchführen, sollten Sie bedenken, dass beide Geräte eine konsistente Konfiguration und CoS-Kennzeichnung (Class of Service) in der gesamten Umgebung benötigen, um ordnungsgemäß funktionieren zu können.

Im Fall von UCS und Nexus ist ein Befehl zur Validierung der MTU-Einstellung pro Schnittstelle und QoS-Gruppe für die **Warteschlangenschnittstelle** hilfreich**. | i queuing|qos-group|MTU**.

## MTU-Anzeige auf Nexus 5000- und UCS-Plattformen

Ein bekannter Aspekt von UCS und Nexus ist die Anzeige von MTUs auf der Schnittstelle. Diese Ausgabe veranschaulicht eine Schnittstelle, die für die Warteschlange von Jumbo Frames und FCoE konfiguriert ist:

```
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)
```

Gleichzeitig zeigt der Befehl **show interface** 1500 Byte an:

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

Im Vergleich zu Carmel-ASIC-Informationen zeigt der ASIC die MTU-Funktion eines bestimmten Ports an.

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

Diese MTU-Diskrepanz in der Anzeige wird für die oben genannten Plattformen erwartet und könnte Neophyten irreführen.

## End-to-End-Konfiguration

Eine konsistente End-to-End-Konfiguration ist die einzige Möglichkeit, eine ordnungsgemäße Leistung zu gewährleisten. Die Konfiguration von Jumbo Frames und die Schritte für Cisco sowie

VMware ESXi werden im UCS mit VMware ESXi-End-to-End-Jumbo-MTU-Konfigurationsbeispiel beschrieben.

UCS FCoE-Uplink-Konfigurationsbeispiel zeigt eine UCS- und Nexus 5000-Konfiguration. Im Anhang A des Referenzdokuments finden Sie einen Überblick über eine grundlegende Nexus 5000-Konfiguration.

Richten Sie FCoE Connectivity für ein Cisco UCS Blade ein. Der Schwerpunkt liegt auf der UCS-Konfiguration für FCoE. Nexus 5000 NPIV FCoE mit FCoE NPV Angeschlossenem UCS - Konfigurationsbeispiel konzentriert sich auf die Nexus-Konfiguration.

## Testen von End-to-End-Jumbo Frames

Die meisten modernen Betriebssysteme bieten die Möglichkeit, eine ordnungsgemäße Konfiguration von Jumbo Frames mithilfe eines einfachen ICMP-Tests (Internet Control Message Protocol) zu testen.

#### Berechnung

9000 Byte - IP-Header ohne Optionen (20 Byte) - ICMP-Header (8 Byte) = 8972 Byte Daten

#### Befehle in gängigen Betriebssystemen

### 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
```

## Buffer-bezogene Probleme

Pufferung und andere latenzbedingte Probleme gehören zu den häufigen Ursachen für Leistungseinbußen in der FlexPod-Umgebung. Nicht alle als Latenz gemeldeten Probleme sind auf tatsächliche Pufferprobleme zurückzuführen. Zahlreiche Messwerte weisen möglicherweise auf eine End-to-End-Latenz hin. Im Fall von NFS ist beispielsweise der angegebene Zeitraum möglicherweise für das erfolgreiche Lesen/Schreiben in den Speicher und nicht für die tatsächliche Netzwerklatenz erforderlich.

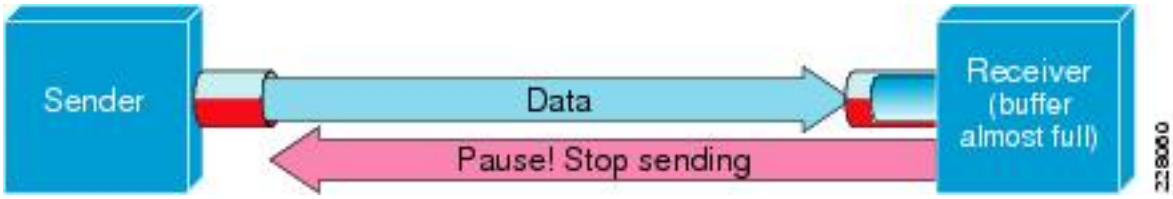
Überlastung ist die häufigste Ursache für Pufferung. In der Layer-2-Welt kann eine Überlastung zu Pufferung und sogar zum Verlust von Frames führen. Um Verwerfungen während Überlastungszeiten zu vermeiden, wurden IEEE 802.3x-Pause-Frames und Priority Flow Control (PFC) eingeführt. Beide setzen voraus, dass der Endpunkt für einen kurzen Zeitraum übertragen

wird, während die Überlastung anhält. Dies kann durch Netzwerküberlastungen (Überlastung der empfangenen Datenmenge) oder durch das Passieren eines priorisierten Frames verursacht werden, wie bei FCoE.

#### Flusskontrolle - 802.3x

Um zu überprüfen, für welche Schnittstellen die Flusskontrolle aktiviert ist, geben Sie den Befehl **show interface flow control ein**. Es ist wichtig, der Empfehlung des Speicheranbieters hinsichtlich der aktivierten Flusskontrolle zu folgen.

Hier wird eine Abbildung gezeigt, die die Funktionsweise der 802.3x-Flusssteuerung veranschaulicht.

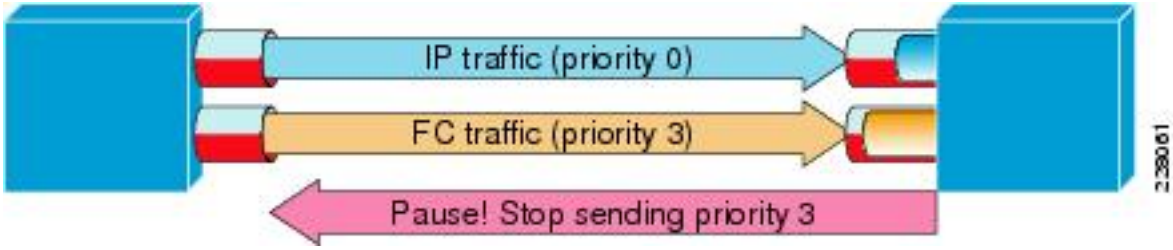



#### PFC - 802.1Qbb

PFC ist nicht für alle Konfigurationen erforderlich, wird jedoch für die meisten empfohlen. Um zu überprüfen, für welche Schnittstellen PFC aktiviert ist, wird die **Show interface priority-flow-control | i Der** Befehl **On** kann auf dem NX-OS und dem Nexus 5000 des UCS ausgeführt werden.

Die Schnittstellen zwischen FIs und dem Nexus 5000 sollten in dieser Liste sichtbar sein. Andernfalls muss die QoS-Konfiguration überprüft werden. QoS muss durchgängig konsistent sein, um die Vorteile von PFC nutzen zu können. Um zu überprüfen, warum die PFC nicht auf einer bestimmten Schnittstelle verfügbar ist, geben Sie den Befehl **show system internal dcbx log interface ethernet x/y** ein, um das DCBX-Protokoll (Data Center Bridging Capabilities Exchange Protocol) abzurufen.

Eine Abbildung, die zeigt, wie Pause-Frames mit PFC funktionieren.




Mit dem Befehl **show interface priority-flow-control** kann der Administrator das Verhalten der Prioritätspausen-Frames pro QoS-Klasse beobachten.

Hier ein Beispiel:

```
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)
```

Diese Ausgabe zeigt, dass das Gerät in der zweiten Klasse lediglich einen PPP-Frame übertrug

(TX).

In diesem Fall ist Ethernet 1/1 mit IOM verbunden, und obwohl PFC für den gesamten Port nicht aktiviert ist, können PPP-Frames für FEX-Ports verarbeitet werden.

```
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
```

In diesem Fall sind FEX-Schnittstellen beteiligt.

```
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
```

Die beteiligten FEX-Ports können auch über **show fex X detail** überprüft werden, wobei X die Chassis-Nummer ist.

```
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
```

Weitere Informationen zu Pausenmechanismen finden Sie in diesen Dokumenten.

- Fibre Channel over Ethernet-Operationen
- Whitepaper zu Unified Fabric: Fibre Channel over Ethernet (FCoE)

#### Warteschlangenrückwürfe

Sowohl der Nexus 5000 als auch das UCS NX-OS protokollieren eingehende Rückwürfe aufgrund von Warteschlangen auf QoS-Gruppenbasis. Beispiel:

```
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)
```

Die Rückwurffunktion *sollte* nur in Warteschlangen erfolgen, die so konfiguriert sind, dass sie Verwerfen zulassen.

Rückwürfe von Eingangswarteschlangen können aus folgenden Gründen auftreten:

- Switched Port Analyzer (SPAN)/Überwachungssitzung aktiviert auf einigen Schnittstellen (siehe Cisco Bug ID CSCur25521)
- Gegendruck von einer anderen Schnittstelle, Pause-Frames werden in der Regel angezeigt, wenn sie aktiviert sind
- An CPU gesendeter Datenverkehr

### Treiberproblem

Cisco bietet zwei Betriebssystem-Treiber für UCS, ENIC und FC. Enic ist für die Ethernet-Konnektivität verantwortlich und fnic ist für die Fibre Channel- und FCoE-Konnektivität verantwortlich. Es ist **sehr wichtig**, dass die Engine- und Netzwerktreiber genau den Angaben in der UCS-Interoperabilitätsmatrix entsprechen. Die Probleme, die durch falsche Treiber verursacht werden, reichen von Paketverlust und erhöhter Latenz bis hin zu einem längeren Boot-Prozess oder einem vollständigen Mangel an Konnektivität.

### Adapterinformationen

Ein von Cisco bereitgestellter Adapter kann eine gute Messung des weitergeleiteten und verworfenen Datenverkehrs liefern. Dieses Beispiel zeigt, wie Sie eine Verbindung zu Chassis X, Server Y und Adapter Z herstellen.

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

Von hier aus kann sich der Administrator beim Monitoring Center for Performance (MCP) anmelden.

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

Mit der MCP-Funktion können Sie die Nutzung des Datenverkehrs pro Logical Interface (LIF) überwachen.

```
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
```

Chassis 1, Server 1 und Adapter 1 verfügen über zwei Virtual Network Interface Cards (VNICs), die virtuellen Schnittstellen (Virtual Ethernet oder Virtual Fibre Channel) 834 und 836 zugeordnet sind. Diese haben die Zahlen 2 und 3. Die Statistiken für LIF 2 und 3 können wie folgt überprüft werden:

```
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
```

Es ist zu beachten, dass der Administrator des UCS über die Spalten "Total" (Gesamt) und "Delta" (zwischen zwei nachfolgenden Ausführen von lifstats) sowie über die aktuelle Datenverkehrslast pro LIF verfügt und Informationen über eventuell auftretende Fehler erhält.

Im vorherigen Beispiel werden Schnittstellen ohne Fehler mit sehr geringer Auslastung angezeigt. Dieses Beispiel zeigt einen anderen Server.

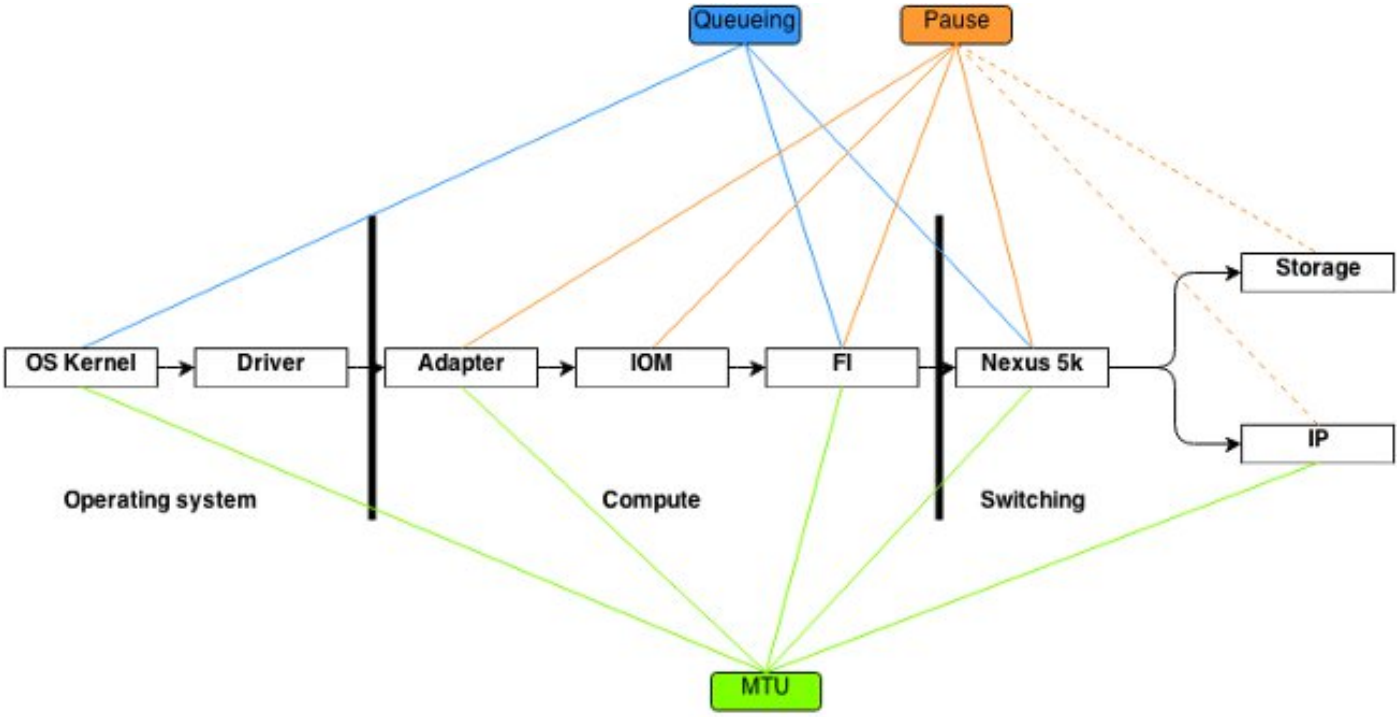
```
adapter 4/4/1 (mcp):2# lifstats 2
              DELTA                TOTAL DESCRIPTION
          127927993            127927993 Tx unicast frames without error
             273955               273955 Tx multicast frames without error
             122540               122540 Tx broadcast frames without 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
```

Zwei interessante Informationsbits zeigen, dass 96 Frames vom Adapter aufgrund eines Mangels an Puffer oder Pufferung deaktiviert wurden und dass zusätzlich TCP Segment Offloading (TSO)-Segmente verarbeitet werden.

## Logischer Paketfluss

Das hier abgebildete Diagramm stellt den logischen Paketfluss in einer FlexPod-Umgebung dar.




Dieses Diagramm soll die Komponenten eines Frames aufschlüsseln, der über die FlexPod-Umgebung durchlaufen wird. Sie spiegelt nicht die Komplexität der einzelnen Blöcke wider und dient lediglich dazu, sich zu merken, wo bestimmte Funktionen konfiguriert und verifiziert werden sollten.

## Eingabe-/Ausgabemodul

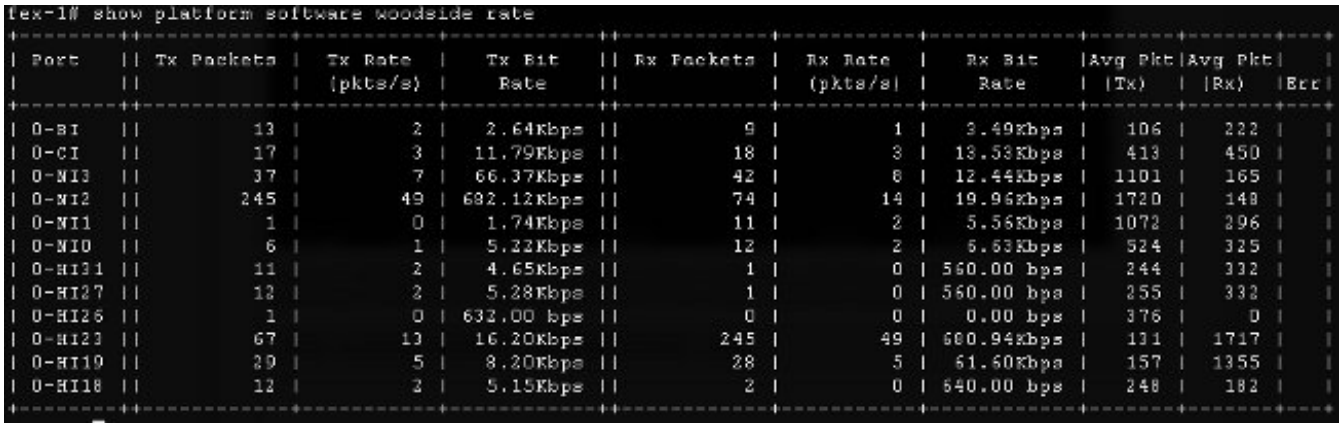
Wie im Diagramm für den logischen Paketfluss gezeigt, ist das Eingabe-/Ausgabemodul (IOM) eine Komponente in der Mitte der gesamten Kommunikation, die das UCS durchläuft. Um eine Verbindung zum IOM im Chassis X herzustellen, geben Sie den Befehl **connect iom x ein**.

Hier sind einige weitere nützliche Befehle:

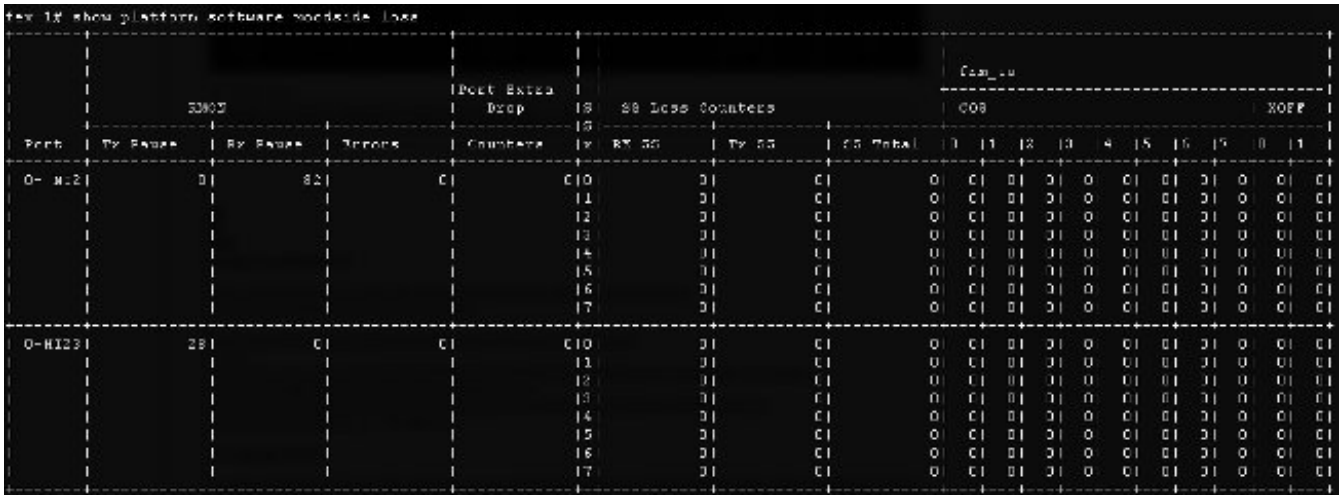
- Topologieinformationen - der Befehl **show plattform software [woodside|redwood] sts** zeigt topologische Informationen aus der Sicht des IOM.


Es zeigt Netzwerkschnittstellen (Network Interfaces, NIs), die zu FIs führen, in diesem Fall gibt es acht, von denen vier hochgefahren sind. Darüber hinaus werden Host-Schnittstellen (HIs) angezeigt, die innerhalb des Chassis zu bestimmten Blades führen.

- Datenverkehrsrate - Der Befehl **show plattform software [woodside|redwood] rate** wird verwendet, um die Datenverkehrsrate zu überprüfen, die die HI-Schnittstellen passiert, sobald die Topologie und die HI-Schnittstelle zur Blade-Zuordnung bekannt sind.




- Datenverkehrsverlust: Geben Sie den Befehl **show platform software [woodside|redwood] loss** ein. Die Ausführung dieses Befehls führt zu einer Null der Verlustindikatoren. Sie können Frames pausieren und auf Schnittstellenbasis ablegen.



Aufgrund der Funktionsweise der zugrunde liegenden Infrastruktur werden die Zähler nur für Schnittstellen angezeigt, bei denen zwischen der Ausführung der beiden Befehle Verluste aufgetreten sind.  In diesem Beispiel sehen Sie, dass die NI2-Schnittstelle 82 Pause-Frames empfangen hat und dass 28 Pause-Frames an die Schnittstelle HI23 übertragen wurden, die bekanntermaßen an Blade 3 angeschlossen ist.

### Überlegungen zum Design

Ein FlexPod ermöglicht eine flexible Konfiguration und Einrichtung von Speicher- und Datennetzwerken. Flexibilität bringt auch zusätzliche Herausforderungen mit sich. Es ist wichtig, Best Practices-Dokumente und ein Cisco Validated Design (CVD) zu befolgen:

- CVD - FlexPod-Bereitstellungsleitfaden
- NetApp Storage Best Practices (nicht speziell für FlexPod) - Storage Connectivity Options and Best Practices für das Cisco Unified Computing System (UCS) mit NetApp Storage

### Überlegungen zur Port-Speed-Auswahl und zum Port-Channel

Ein häufiges Problem für TAC-Techniker ist die Überlastung von Links aufgrund der Auswahl von 1-Gbit-Ethernet anstelle von 10-Gbit-Ethernet, auf die in Best Practice-Dokumenten verwiesen wird. Als Beispiel ist die Leistung eines **einzelnen Datenflusses** bei zehn 1-Gbit-Verbindungen nicht besser als bei einer 10-Gbit-Verbindung. Im Port-Channel kann ein einzelner Datenfluss über eine einzelne Verbindung geleitet werden.

Geben Sie den Befehl **show port-channel load-balance (Port-Channel-Lastenausgleich) ein, um zu erfahren, welche Lastenausgleichsmethode auf Nexus und/oder FIs NX-OS verwendet wird**. Der Administrator kann außerdem herausfinden, welche Schnittstelle in einem Port-Channel als ausgehende Schnittstelle für ein Paket oder einen Frame ausgewählt wird. Ein einfaches Beispiel für einen Frame im VLAN49 zwischen zwei Hosts ist hier dargestellt:

```
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
```

## Speicherspezifische Probleme

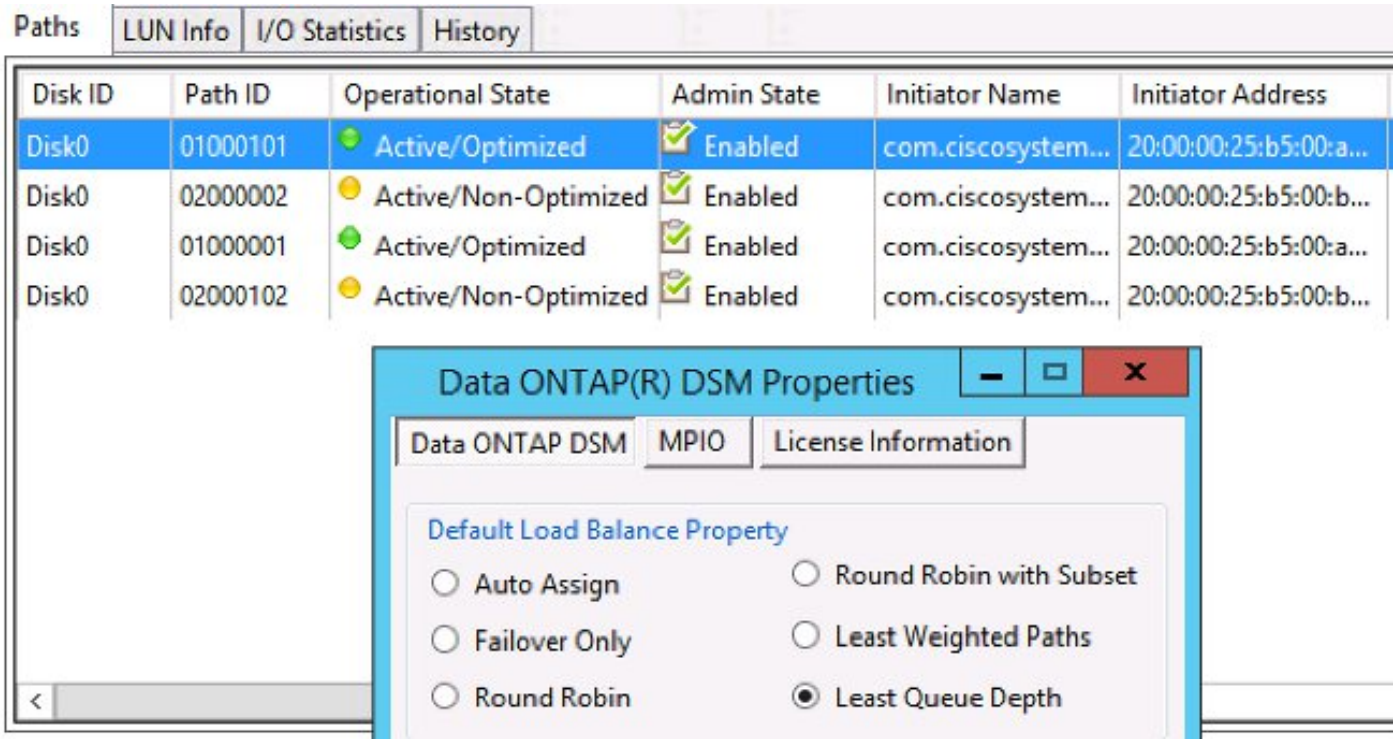
Die zuvor angesprochenen Probleme treten sowohl bei Daten- als auch bei Speichernetzwerken auf. Aus Gründen der Vollständigkeit werden auch Leistungsprobleme speziell für Storage Area Network (SAN) erwähnt. Speicherprotokolle wurden mit Ausfallsicherheit erstellt, und Multi-Pathing wird immer noch erweitert. Durch die Einführung von Technologien wie z. B. Asymmetric Logical Unit Assignment (ALUA) und Multi-Path IO (MPIO) werden Administratoren mehr Flexibilität und Optionen geboten.

### Speicherplatzierung

Ein weiterer Gesichtspunkt ist die Speicherplatzierung. Ein FlexPod-Design sieht vor, dass Storage an Nexus-Switches angeschlossen werden muss. Direkt angeschlossener Speicher entspricht nicht dem CVD. Designs mit direkt angeschlossenem Speicher werden unterstützt, wenn Best Practices befolgt werden. Gleichzeitig handelt es sich bei diesen Designs nicht ausschließlich um FlexPod.

### Optimale Pfadauswahl

Dies ist technisch gesehen kein Problem mit Cisco, da die meisten dieser Optionen für Cisco Geräte transparent sind. Es ist ein häufiges Problem, einen optimalen Pfad zu wählen und zu halten. Ein modernes gerätespezifisches Modul (DSM) kann über mehrere Pfade angeboten werden und muss anhand bestimmter Kriterien für Ausfallsicherheit und Lastenausgleich ein optimales Modul bzw. optimale Module auswählen. Dieser Screenshot zeigt vier Pfade, die NetApp DSM für Microsoft Windows und Load Balancing-Optionen zur Verfügung stehen.




Die empfohlenen Einstellungen sollten auf der Grundlage eines Gesprächs mit dem Storage-Anbieter ausgewählt werden. Diese Einstellungen können Leistungsprobleme beeinflussen. Ein typischer Test, den Sie vom TAC möglicherweise durchführen lassen, ist ein Lese-/Schreibtest, der nur in Fabric A oder Fabric B durchgeführt wird. Dadurch können Sie Leistungsprobleme in

der Regel auf Situationen beschränken, die im Abschnitt "Häufige Probleme" dieses Dokuments behandelt werden.

#### Gemeinsame Nutzung von VM- und Hypervisor-Datenverkehr

Dieser Punkt ist spezifisch für die Computing-Komponente, unabhängig vom Anbieter. Eine einfache Möglichkeit zum Einrichten eines Speichernetzwerks für Hypervisoren ist die Erstellung von zwei Host Bus Adaptern (HBAs), einem für jede Faser, sowie die Ausführung des Boot-LUN-Datenverkehrs und des VM-Speicherdatenverkehrs über diese beiden Schnittstellen. Es wird immer empfohlen, den Boot-LUN-Datenverkehr und den VM-Speicherverkehr zu teilen. Dies ermöglicht eine bessere Leistung und zusätzlich eine logische Trennung zwischen den beiden Arten von Datenverkehr. Ein Beispiel finden Sie im Abschnitt "Bekannte Probleme".

# Tipps zur Fehlerbehebung

## Das Problem beschränken

Wie bei jeder schnellen Fehlerbehebung ist es sehr wichtig, das Problem einzugrenzen und die richtigen Fragen zu stellen.

- Welche Geräte/Anwendungen/VM sind betroffen?
- Welcher Speichercontroller ist betroffen?
- Welche Pfade sind betroffen?
- Wie oft tritt das Problem auf (/nicht)?

## Cisco

#### Grenzwerte für Zähler

In dieser Dokumentschnittstelle werden ASIC-Warteschlangenzähler behandelt. Die Zähler geben auch zu einem bestimmten Zeitpunkt eine Übersicht. Daher ist es wichtig, die Zunahme der Zähler zu überwachen. Bestimmte Zähler können nicht standardmäßig gelöscht werden. Zum Beispiel die zuvor erwähnte Carmel-ASIC.

Um ein deutliches Beispiel zu geben, ist das Vorhandensein von CRC oder Rückwürfen auf einer Schnittstelle möglicherweise nicht ideal, aber es ist zu erwarten, dass ihre Werte nicht null sind. Die Zähler hätten zu einem bestimmten Zeitpunkt, möglicherweise während der Übergangsphase oder der Ersteinrichtung, ansteigen können. Daher ist es wichtig, die Erhöhung der Zähler zu beachten und wann sie zuletzt gelöscht wurden.

#### Überlegungen zur Kontrollebene

Es ist zwar hilfreich, Zähler zu überprüfen, aber es ist wichtig zu wissen, dass bestimmte Probleme auf der Datenebene möglicherweise keine einfache Reflektion über die Kontrollebenen und Tools finden. Wie bereits erwähnt, ist der Ethanalyzer ein sehr nützliches Tool, das sowohl

auf dem UCS als auch auf dem Nexus 5000 verfügbar ist. Es kann jedoch nur Datenverkehr auf Kontrollebene erfassen. Eine Erfassung des Datenverkehrs wird häufig vom TAC angefordert, insbesondere wenn nicht klar ist, wo der Fehler liegt.

### Erfassung von Datenverkehr

Eine zuverlässige Erfassung des Datenverkehrs auf den End-Hosts kann ein Leistungsproblem aufdecken und es recht schnell eingrenzen. Sowohl der Nexus 5000 als auch das UCS bieten Datenverkehr-SPAN. Insbesondere die Optionen des UCS, bestimmte HBAs und Fabric-Seiten mit SPAN zu verbinden, sind nützlich. Weitere Informationen zu den Funktionen zur Erfassung von Datenverkehr bei der Überwachung einer UCS-Sitzung finden Sie unter den folgenden Referenzen:

- UCS-Datenverkehrsanalyse für physische und virtuelle Adapter (Video)
- Konfigurationsleitfaden für die grafische Benutzeroberfläche von Cisco UCS Manager - Datenverkehrsüberwachung

## NetApp

NetApp bietet eine Reihe von Dienstprogrammen zur Fehlerbehebung bei den Speichercontrollern an, darunter:

- perfstat - ein sehr nützliches Dienstprogramm, das in der Regel für NetApp Support-Mitarbeiter ausgeführt wird
- systat - Bietet Informationen darüber, wie beschäftigt der Filer ist und was er macht - NetApp Support Library

Es gibt einige der gebräuchlichsten Befehle:

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

In der Ausgabe **sysstat -x2** gibt es einige Aspekte, die auf überladene NetApp-Arrays oder -Festplatten hinweisen könnten:

- Dauerhafte **CP-**ty-Spalte mit vielen **Werten:** oder **F**
- Dauerhafte **HDD-**Standspalte über **20 %**

In diesem Artikel wird die Konfiguration von NetApp beschrieben: NetApp Ethernet Storage - Best Practices.

- VLAN-Tagging
- VLAN-Trunking
- Jumbo-MTU
- IP-Hashing
- FlowControl deaktivieren

## VMware

ESXi bietet Secure Shell (SSH)-Zugriff, über den Sie eine Fehlerbehebung durchführen können.

Zu den nützlichsten Tools, die Administratoren zur Verfügung gestellt werden, gehören esxtop und perfmon.

- esxtop - Ähnlich wie Linux/BSD top können Benutzer Parameter überwachen, die in Echtzeit Performance-bezogene Parameter sind.
  Verwenden von esxtop zur Identifizierung von Problemen mit der Speicherleistung für ESX/ESXi
- perfmon - Ermöglicht Benutzern die Fehlerbehebung bei Microsoft Windows Virtual Machines (VM)
  Erfassen der Windows Perfmon-Protokolldaten zur Diagnose von Leistungsproblemen bei virtuellen Systemen
- Diagnosepakete auf ESXi sammeln - Diagnoseinformationen für VMware ESX/ESXi mithilfe des vSphere Client (653) sammeln
- VMware vSwitch Load Balancing-Anforderungen für Cisco Server der B-Serie - Routen auf Basis von IP-Hash werden von Cisco UCS B200 M1/M2 Blade-Servern mit UCS Fabric Interconnects der Serie 6100 nicht unterstützt.

# Bekannte Probleme und Verbesserungen

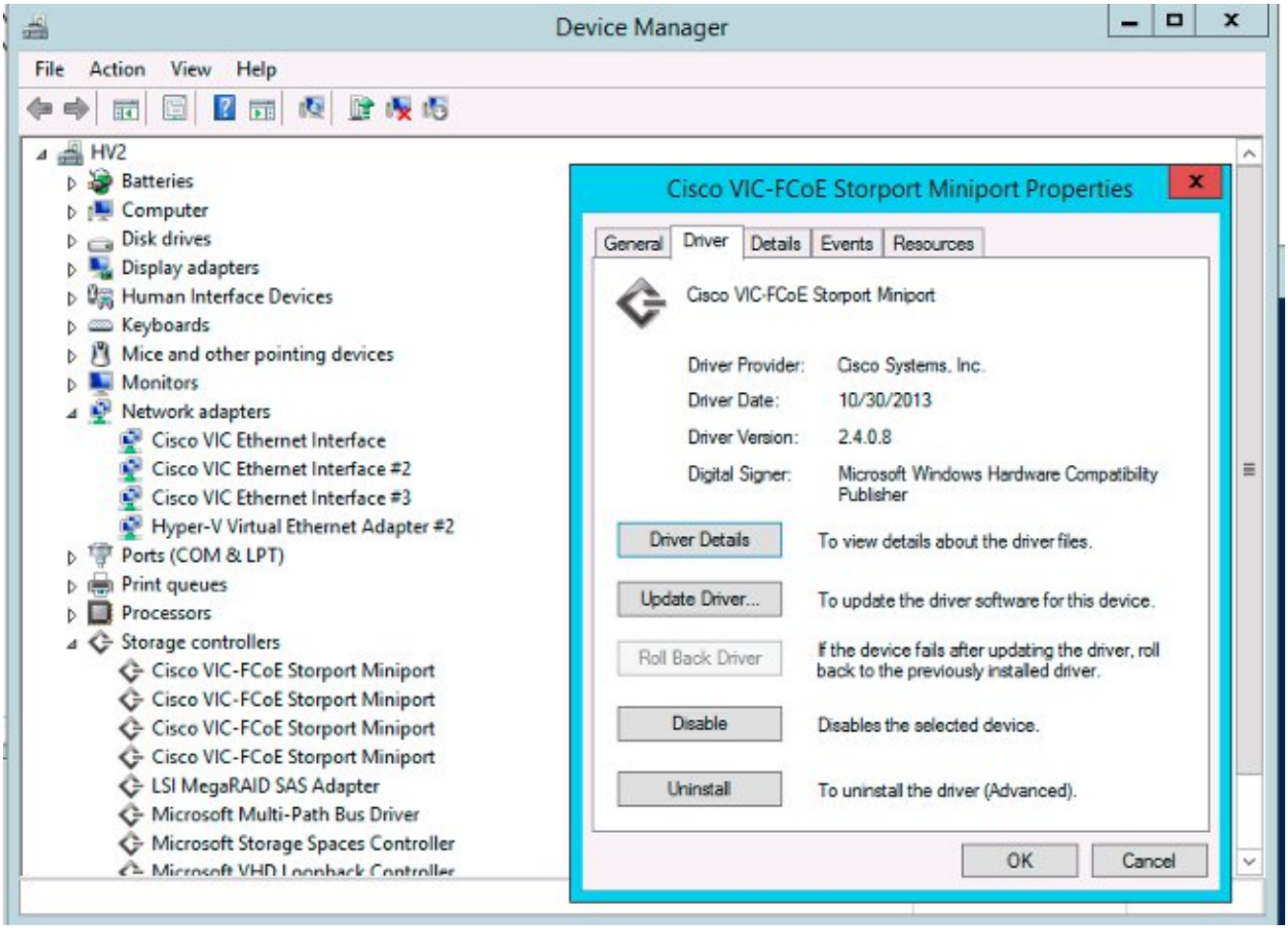
- Cisco Bug ID CSCuj86736 - bei passiven Twinax-Kabeln können CRC-Fehler zunehmen. Dies wird verursacht, wenn der Nexus 5000 DFE nicht optimiert. Geben Sie den Befehl **show hardware internal carmel eye** ein, um zu überprüfen, ob der Parameter "Eye height" über 100 mv liegt. Dies wurde in den Versionen 5.2(1)N1(7) und 7.0(4)N1(1) behoben.
- Cisco Bug-ID CSCuo76425 - ähnlich dem vorherigen Fehler und auch auf den UCS Fabric Interconnects vorhanden. Dies ist in Version 2.2(3a) behoben.
- Cisco Bug-ID CSCuo76425 - identisch mit Bug CSCuj86736 mit Ausnahme von UCS Fabric Interconnect.
- Cisco Bug ID CSCup40056 - Timing-Problem, das durch die Freigabe des Boot-Datenverkehrs mit dem VM-Datenverkehr verursacht wird, beschrieben in Unified Computing System Virtual Machine Live Migration schlägt mit virtuellen Fibre Channel-Adaptern fehl.
- Langsame Erkennung und Vermeidung von Abfluss - häufig werden FC und FCoE durch langsame Ableitung beeinträchtigt. NX-OS Release 7.0(0)N1(1) bietet neue Möglichkeiten zur Erkennung und Vermeidung. Weitere Informationen zu dieser Funktion finden Sie im Konfigurationshandbuch für NX-OS-Schnittstellen der Cisco Nexus 5500-Serie sowie im Leitfaden zur Erkennung von und zur Vermeidung von Überlastungen bei langsamer Entwässerung.
- Cisco Bug ID CSCuj81245 - eine Einschränkung in PALO-basierten Karten (VIC1240 und andere), die FC-Abbrüche verursacht.
- Cisco Bug ID CSCuh61202 - Nach dem Upgrade auf Version 2.1(3) wird die UCS-Firmware-FC abgebrochen, und es sind mehrere andere Probleme zu erkennen.
- Cisco Bug ID CSCtw91018 - Eine Mischung aus MTU-Einstellungen für VNICs auf einem einzigen, PALO-basierten Adapter kann bei einigen Datenverkehrsklassen zum Ausfall führen.
- Cisco Bug-ID CSCuq40256 - führt dazu, dass PFC auf Verbindungen von Fabric Interconnect bis hin zu Server-Adaptern deaktiviert wird. Dies führt zu einer Reihe von Problemen, die mit Fibre Channel-Ausfällen und Out-of-Order-Frames beginnen, die auf Speicherseite gemeldet werden. Es können Speicherunterbrechungen und andere Leistungsprobleme gemeldet

werden.

## TAC-Tickets

In vielen Fällen wird der TAC-Techniker Sie bitten, einige grundlegende Informationen zu sammeln, bevor eine Untersuchung gestartet werden kann.

- Topologiediagramm - enthält Portnummern und Leitungsgeschwindigkeiten, unbedingt erforderlich.
- Technischer UCSM-Support - Visual Guide zum Sammeln von Dateien des technischen Supports (B- und C-Serie).
- Technischer Support für das UCS-Chassis für ein Chassis, bei dem Probleme auftreten - siehe vorheriger Link.
- Technischer Nexus 5000-Support und andere Netzwerkgeräte zwischen dem UCS und NetApp - Umleitung der Ausgabe des Befehls show tech-support details.
- Ausgabe des Befehls **show queueing interface** auf beiden FIs
```
connect nxos A|B
show queuing interface | no-more
show interface priority-flow-control | no-more
show interface flowcontrol | no-more.
```
- Host-Treiberversionen auf ESXi führen aus: Geben Sie die folgenden Befehle ein: vmkload_mod -s enicvmkload_mod -s fnic
- Linux -
```
dmesg | egrep -i 'enic|fnic'
```
- Windows - Überprüfen Sie die Treiberversion im "Gerätemanager". Ein Beispiel aus Windows 2012 R2 zeigt drei Cisco VIC Ethernet-Schnittstellen und vier VIC FCoE Mini-Port-Schnittstellen (die auch für Fibre Channel, nicht nur FCoE, verantwortlich sind) und Version 2.4.0.8 des FCoE-
Treibers.




# Feedback

Über die Feedback-Schaltfläche können Sie Feedback zu diesem Dokument oder zu Ihren Erfahrungen abgeben. Wir werden dieses Dokument laufend aktualisieren, sobald Entwicklungen eintreten und Feedback eingeht.