FlexPod の一般的なパフォーマンスの問題

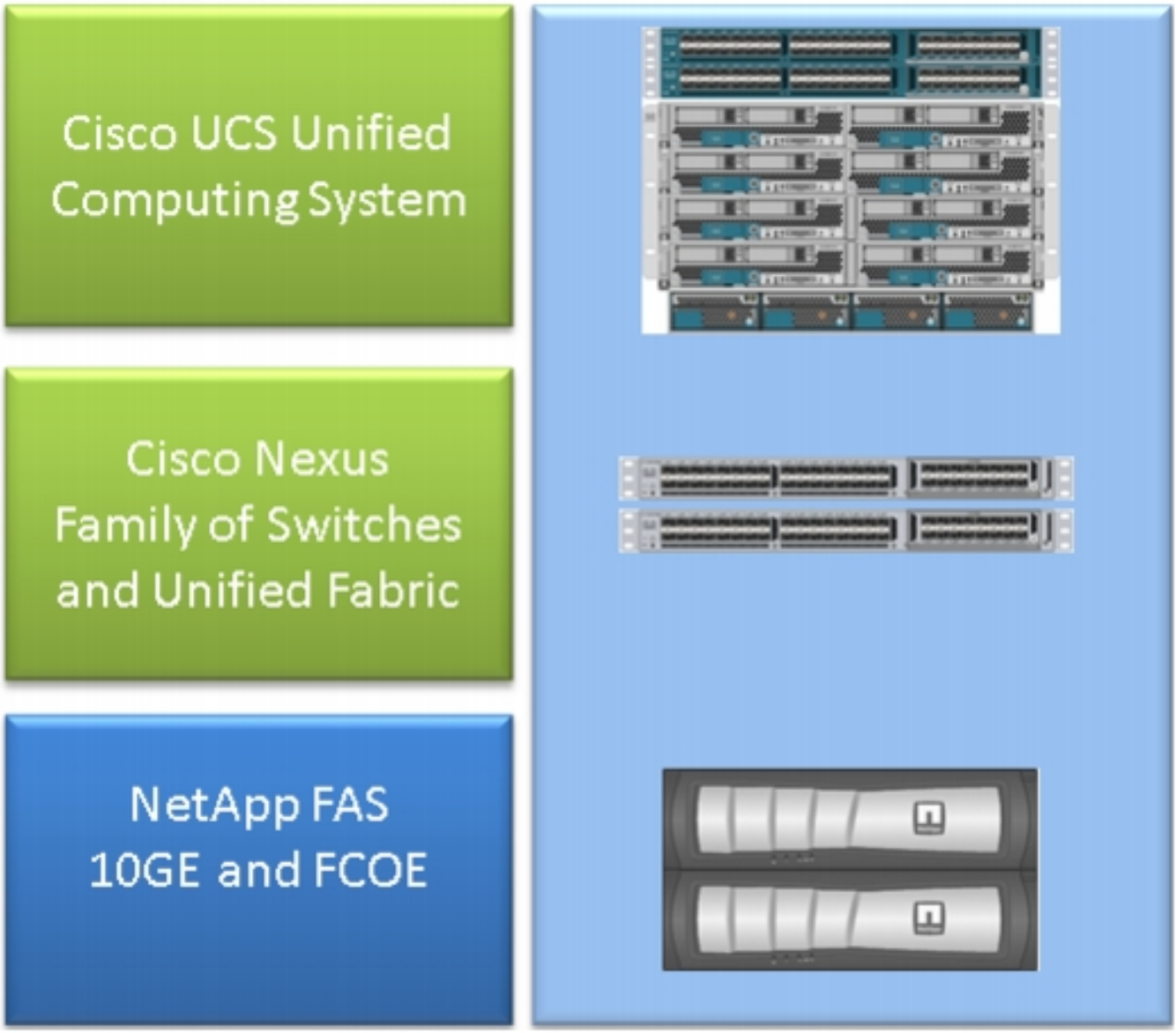
目次

概要

このドキュメントでは、FlexPod 環境の一般的なパフォーマンスの問題について説明し、問題の解決方法を紹介し、軽減手順を示します。 このドキュメントは、FlexPod 環境のパフォーマンス問題の解決法をさがしているお客様が開始点にできるよう準備されています。 このドキュメントには、ここ数か月の間にデータセンター ソリューション テクニカル アシスタンス センター（TAC）チームによって確認された問題の結果が記載されています。

FlexPod の概念の概要

FlexPod は、Nexus スイッチ経由で NetApp ストレージと IP ネットワークに接続されたユニファイド コンピューティング システム（UCS）コンピュータで構成されます。



最も一般的な FlexPod は、ファブリック インターコネクト（FI）から Nexus 5500 スイッチ経由で NetApp ファイラに接続された Cisco UCS B シリーズ シャーシで構成されます。 FlexPod Express と呼ばれる別のソリューションでは、Nexus 3000 スイッチに接続された UCS C シリーズ シャーシが使用されます。 このドキュメントでは、最も一般的な FlexPod について説明します。

パフォーマンスに関する考慮事項

FlexPod でよく見られる複数の要素で構成された複雑な環境では、問題を解決するために複数の側面を考慮する必要があります。 レイヤ 2 ネットワークと IP ネットワークにおける一般的なパフォーマンスの問題の原因を以下に示します。

- パケットまたはフレームの損失：データ ビットの損失はアプリケーションのパフォーマンスに悪影響を及ぼします。
- バッファリング：パケットまたはフレームがキュー/バッファ内に長く留まりすぎると、アプリケーション、特に、ストレージ ネットワーキングの場合に、特定のパフォーマンスに影響が出ます。 遅延、並べ替え、およびノーマライザの問題がこのカテゴリに分類されます。
- MTU の不一致の問題とフラグメンテーション：より高いパフォーマンスに到達したときによく見られる問題。 フラグメンテーションと MTU の不一致に関する問題がこのカテゴリに入ります。

環境

パフォーマンスを測定する環境を理解しておくことが重要です。 問題を適切に絞り込むためには、ストレージ タイプやプロトコルだけでなく、影響を受けるサーバのオペレーティング システム（OS）や口ケーションに関する疑問を提起する必要があります。 接続の概要を示すトポロジ図が最低限必要です。

測定

測定対象と測定方法を理解しておく必要があります。 特定のアプリケーションだけでなく、ほとんどのストレージ ベンダーとハイパーバイザ ベンダーもシステムのパフォーマンス/ヘルスを示すある種の測定結果を公表しています。 これらの測定結果は、ほとんどトラブルシューティング手法の代わりにはなりませんが、手がかりにはなります。

たとえば、ハイパーバイザにおけるネットワーク ファイル システム（NFS）ストレージの遅延の測定結果がパフォーマンスの低下を示していたとしても、それだけではネットワークが関係していることにはなりません。 この NFS のケースでは、ホストから NFS ストレージ IP ネットワークへの単純な ping によってネットワークが原因かどうかが明らかになります。

基準

この点は、特に TAC ケースを開くときに、あまり重視されません。 パフォーマンスが不十分であることを示すには、測定したパラメータを開示する必要があります。 これには期待値とテスト値が含まれます。 理想的には、過去のデータとそのデータを取得した際のテスト方法を示す必要があります。

以下に例を示します。 単一のイニシエータから単一の論理ユニット番号（LUN）への書き込みのみのテストで 10 ms の遅延が確認されたとしても、それがフル装備のシステムに対して想定される遅延になるわけではありません。

FlexPod のパフォーマンスの問題

このドキュメントは大部分の FlexPod 環境の参考になるように作られているため、データセンター ソリューションを担当している TAC チームによって最も頻繁に確認された問題のみが記載されています。

一般的な問題

ここでは、ストレージと IP/レイヤ 2 ネットワークに共通する問題について説明します。

フレームとパケットの損失

フレームとパケットの損失は、パフォーマンスに影響を与える、最もよく見られる要因です。 問題の兆候を探す一般的な場所の 1 つがインターフェイス レベルです。 Nexus 5000 または Cisco Nexus Operating System（NX-OS）CLI から、**show interface | sec "is up" | egrep ^(Eth|fc)|discard|drop|CRC** コマンドを入力します。 インターフェイスがアップしている場合は、その名前が列挙され、カウンタを破棄しドロップします。 同様に、すべてのインターフェイスのエラー統計情報を表示する **show interface counters errors** コマンドを入力すると、適切な要約が表示されます。

イーサネット環境

0 以外のカウンタが問題を示していない場合があることを認識しておくことが重要です。 特定のシナリオでは、このようなカウンタが初期セットアップまたは過去の動作変更で増加した可能性があります。 カウンタの増加をモニタする必要があります。

より示唆的な ASIC レベルでカウンタを収集することもできます。 特に、インターフェイス上の巡回冗長検査（CRC）エラーの場合に、TAC でよく使用されるコマンドが **show hardware internal carmel crc** です。 Carmel はポート レベル転送を担当している ASIC の名前です。

同様の出力を 6100 シリーズ FI または Nexus 5600 スイッチからポート単位で収集することができます。 FI 6100 の場合は、gatos ASIC から、次のコマンドを入力します。

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

Nexus 5600 の場合は、bigsur ASIC から、次のコマンドを入力します。

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

carmel ASIC 用のコマンドは、CRC パケットが受信された場所と転送された場所に加えて、さらに重要なことに、それらがストンプされたかどうかを表示します。

Nexus 5000 と UCS NX-OS の両方の処理がカットスルーされるため、不正なフレーム チェック シーケンス（FCS）があるスイッチング モード フレームに対して転送前にストンプだけが行われます。 破損したフレームの発生場所を見つけることが重要です。

```
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 |
```

この例は、Nexus 5000 へのアップリンクである、Eth 1/17 と Eth 1/18 からのストンプ済みパケットを示しています。 これらのフレームは、Eth 1/17 + Eth 1/18 rx Stomp = Eth 1/34 tx Stomp のように後で Eth 1/34 に送信されることが想定できます。

Nexus 5000 の同様の表示を以下に示します。

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

この出力は、CRC が 2 つのリンク上で受信され、転送前にストンプとしてマークされたことを示しています。 詳細については、『Nexus 5000 Troubleshooting Guide』を参照してください。

ファイバ チャネル環境

ドロップ (破棄、エラー、CRC、B2B クレジット枯渇) を検索する簡単な方法は、**show interface counters fc** コマンドを経由する方法です。

このコマンドは、Nexus 5000 とファブリック インターコネクト上で使用でき、ファイバ チャネルの世界で何が起きたかを示す適切な指標を提供します。

次に、例を示します。

```
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
(...)
```

このインターフェイスはビジーではなく、この出力は破棄やエラーが起きていないことを示して

います。

加えて、0 からの B2B クレジット遷移が強調表示されています。 Cisco Bug ID の CSCue80063 と CSCut08353 が原因で、これらのカウンタは信頼できません。 また、Cisco MDS では正常に機能しますが、Nexus5k プラットフォームの UCS では正常に機能しません。 Cisco Bug ID CSCsz95889 も確認してみてください。

ファイバ チャネル (FC) のイーサネット環境の carmel と同様に、fc-mac ファシリティを使用できます。 たとえば、ポート fc2/1 に対して、**show hardware internal fc-mac 2 port 1 statistics** コマンドを入力します。 表示されるカウンタは 16 進数形式です。

```
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
```

この出力は、入力で 15 回破棄があったことを示しています。 これは、FCP_CNTR_PIF_RX_DROP が 0xf (10 進数の 15) になっていることと一致します。 この情報は、再度、FWM (フォワーディング マネージャ) 情報に関連付けることができます。

```
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
```

ただし、これにより、ドロップ合計数と対応する ASIC 番号が管理者に通知されます。 ドロップした ASIC の原因に関する情報を問い合わせる必要があります。

```
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]
```

この場合は、トラフィックが FC 環境 - ゾーン分割でよく見られる入力アクセス コントロール リスト (ACL) によってドロップされました。

MTU の不一致

FlexPod 環境では、アプリケーションとプロトコルのエンドツーエンド最大転送単位 (MTU) 設定を提供することが重要です。 ほとんどの環境では、これが Fibre Channel over

Ethernet（FCoE）とジャンボ フレームになります。

加えて、フラグメンテーションが発生した場合に、パフォーマンスの低下が予想されます。 ネットワーク ファイル システム（NFS）や Internet Small Computer System Interface（iSCSI）などのプロトコルの場合は、エンドツーエンド IP 最大伝送単位（MTU）と TCP 最大セグメント サイズ（MSS）をテストして証明することが重要です。

ジャンボ フレームと FCoE のどちらをトラブルシューティングする場合でも、その両方が正常に動作するためには環境全体で一貫した設定とサービス クラス（CoS）マーキングが必要なことを覚えておくことが重要です。

UCS と Nexus の場合、インターフェイス単位と QoS グループ単位の MTU 設定の検証を支援するコマンドは **show queuing interface | i queuing|qos-group|MTU** です。

Nexus 5000 と UCS プラットフォームでの MTU の表示

UCS と Nexus のよく知られている側面は、インターフェイス上での MTU の表示です。 この出力は、ジャンボ フレームと FCoE をキューに入れるように設定されたインターフェイスを示しています。

```
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)
```

同時に、**show interface** コマンドは 1500 バイトを表示します。

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

carmel ASIC 情報と比較した場合、ASIC は特定のポートの MTU 機能を表示します。

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

この表示上の MTU の不一致は前述のプラットフォームで想定されていることであり、初心者は戸惑う可能性があります。

エンドツーエンドの構成

適切なパフォーマンスを保証するための唯一の手段は、エンドツーエンドの一貫した構成です。 Cisco 側と VMware ESXi のジャンボ フレーム構成と手順については、「VMware ESXi エンドツーエンド ジャンボ MTU 設定を使用した UCS の例」を参照してください。

「UCS FCoE アップリンクの構成例」に、UCS と Nexus 5000 の構成が記載されています。 基本的な Nexus 5000 構成の概要については、参考資料の付録 A を参照してください。

「Cisco UCS ブレード用の FCoE 接続の設定」では、FCoE 用の UCS 構成に焦点が当てられています。 「Nexus 5000 と UCS 間の FCoE を使用した NPIV と NPV の設定例」では、Nexus 構成に焦点が当てられています。

エンドツーエンドのジャンボ フレームのテスト

最近のほとんどのオペレーティング システムが、単純な Internet Control Message Protocol（ICMP）テストを使用して、適切なジャンボ フレーム構成をテストできる機能を備えています。

計算

9000 バイト - オプションなしの IP ヘッダー（20 バイト）- ICMP ヘッダー（8 バイト）= 8972 バイトのデータ

主要なオペレーティング システムのコマンド

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
```

バッファ関連の問題

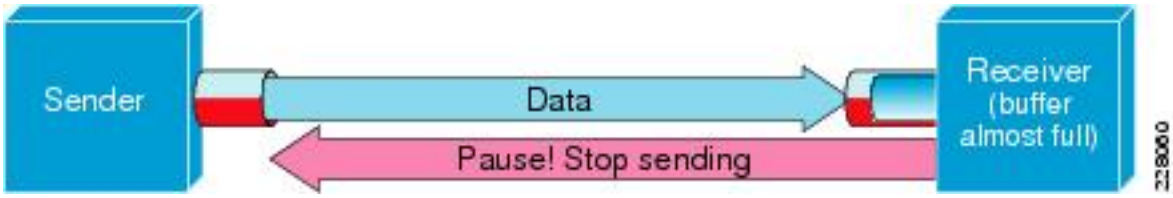
バッファリングやその他の遅延関連の問題も、FlexPod 環境における一般的なパフォーマンス低下の原因です。 遅延として報告されたすべての問題が実際のバッファリングの問題から発生しているわけではなく、さまざまな測定結果がエンドツーエンド遅延を示唆しています。 たとえば、NFS の場合は、報告された期間は、実際はネットワーク遅延ではなく、ストレージに対して正しく読み書きするのに必要な期間かもしれません。

輻輳がバッファリングの最も一般的な原因です。 レイヤ 2 環境では、輻輳によってフレームのバッファリングとテール ドロップが発生する可能性があります。 輻輳期間中のドロップを避けるために、IEEE 802.3x の一時停止フレームとプライオリティ フロー制御（PFC）が導入されました。 この両方が、輻輳が続いている間は伝送を短時間保留にするようエンド ポイントに要求します。 これは、FCoE の場合と同様に、ネットワーク輻輳（多量のデータで受信側が混乱する）が原因で、または、優先順位付きフレームを渡す必要があるという理由で発生します。

フロー制御 - 802.3x

どのインターフェイスでフロー制御が有効になっているかを確認するには、show interface flowcontrol コマンドを入力します。 フロー制御の有効化に関するストレージ ベンダーの推奨事項に従うことが重要です。

802.3x フロー制御の動作を示すイラストを以下に示します。

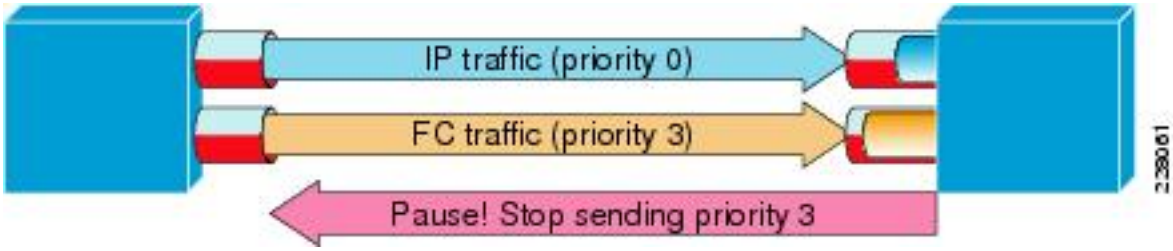


PFC - 802.1Qbb

PFC はすべてのセットアップに必要ではありませんが、ほとんどのセットアップに推奨されています。 どのインターフェイスで PFC が有効になっているかを確認するために、**show interface priority-flow-control | i On** コマンドを UCS の NX-OS と Nexus 5000 で実行できます。

FI と Nexus 5000 間のインターフェイスがそのリストに表示されるはずです。 そうでない場合は、QoS 構成を確認する必要があります。 PFC を利用するためには、QoS を一貫したエンドツーエンドにする必要があります。 PFC が特定のインターフェイス上でアップしていない理由を確認するには、**show system internal dcbx log interface ethernet x/y** コマンドを入力して、Data Center Bridging Capabilities Exchange Protocol (DCBX) ログを入手します。

一時停止フレームが PFC でどのような役割を果たすかを示すイラストを以下に示します。



show interface priority-flow-control コマンドを使用すれば、管理者は、プライオリティー時停止フレームの QoS クラス単位の動作を確認することができます。

次に例を示します。

```
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)
```

この出力は、2 つ目のクラスで、デバイスが PPP フレームを送信 (TX) しているところを示しています。

この場合は、イーサネット 1/1 が IOM に対向するポートで、ポート全体では PFC が有効になりませんが、FEX ポート向けの PPP フレームを処理します。

```
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
```

この場合は、FEX インターフェイスが関与します。

```
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
```

関与する FEX ポートは、**show fex X detail** 経由でチェックすることもできます。ここで、X はシャーシ番号です。

```
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
```

一時停止メカニズムの詳細については、次のドキュメントを参照してください。

- Fibre Channel over Ethernet の動作
- ユニファイド ファブリック ホワイト ペーパー - Fibre Channel over Ethernet (FCoE)

キューイングの破棄

Nexus 5000 と UCS NX-OS の両方が QoS グループ単位でキューイングが原因の入力破棄を追跡します。 次に、例を示します。

```
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)
```

入力破棄は、ドロップを許可するように設定されたキュー内でのみ起きるはずです。

入力キューイング破棄は次の原因で発生する可能性があります。

- 一部のインターフェイスでスイッチド ポート アナライザ (SPAN) /モニタリング セッション (Cisco Bug ID CSCur25521 を参照) が有効になっている。

- 別のインターフェイスからのバック プレッシャ、これが有効になっている場合に一時停止フレームがよく検出される。
- CPU にパントされたトラフィック。

ドライバの問題

Cisco は、UCS 用の 2 つのオペレーティング システム ドライバとして enic と fnic を提供しています。 enic はイーサネット接続を担当し、fnic はファイバ チャネルと FCoE 接続を担当します。 enic ドライバと fnic ドライバが UCS 相互運用性マトリックスで指定したとおりに動作することが**非常に重要**です。 不正なドライバによって引き起こされる問題は、パケット損失や遅延の増大からブート プロセスの長期化や接続の完全な欠如にまで及びます。

アダプタ情報

シスコが提供しているアダプタは、通過するまたはドロップされるトラフィックに関する適切な測定結果を提供できます。 次の例は、シャーシ X、サーバ Y、およびアダプタ Z への接続方法を示しています。

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

ここから、管理者は、Monitoring Center for Performance (MCP) ファシリティに口グインできます。

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

MCP ファシリティを使用すれば、論理インターフェイス (LIF) 単位のトラフィックの使用量をモニタすることができます。

```
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
```

シャーシ 1、サーバ 1、およびアダプタ 1 には、仮想インターフェイス (仮想イーサネットまたは仮想ファイバ チャネル) の 834 と 836 に関連付けられた 2 枚の仮想ネットワーク インターフェイス カード (VNIC) が実装されています。 これらのカードに番号の 2 と 3 が割り当てられています。 LIF 2 と 3 に関する統計情報は次のように確認することができます。

```
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
```

UCS の管理者には、合計列と差分（lifstats の連続した 2 回の実行の間）列だけでなく、LIF 単位の現在のトラフィック負荷と発生したエラーに関する情報も提示されることに注意してください。

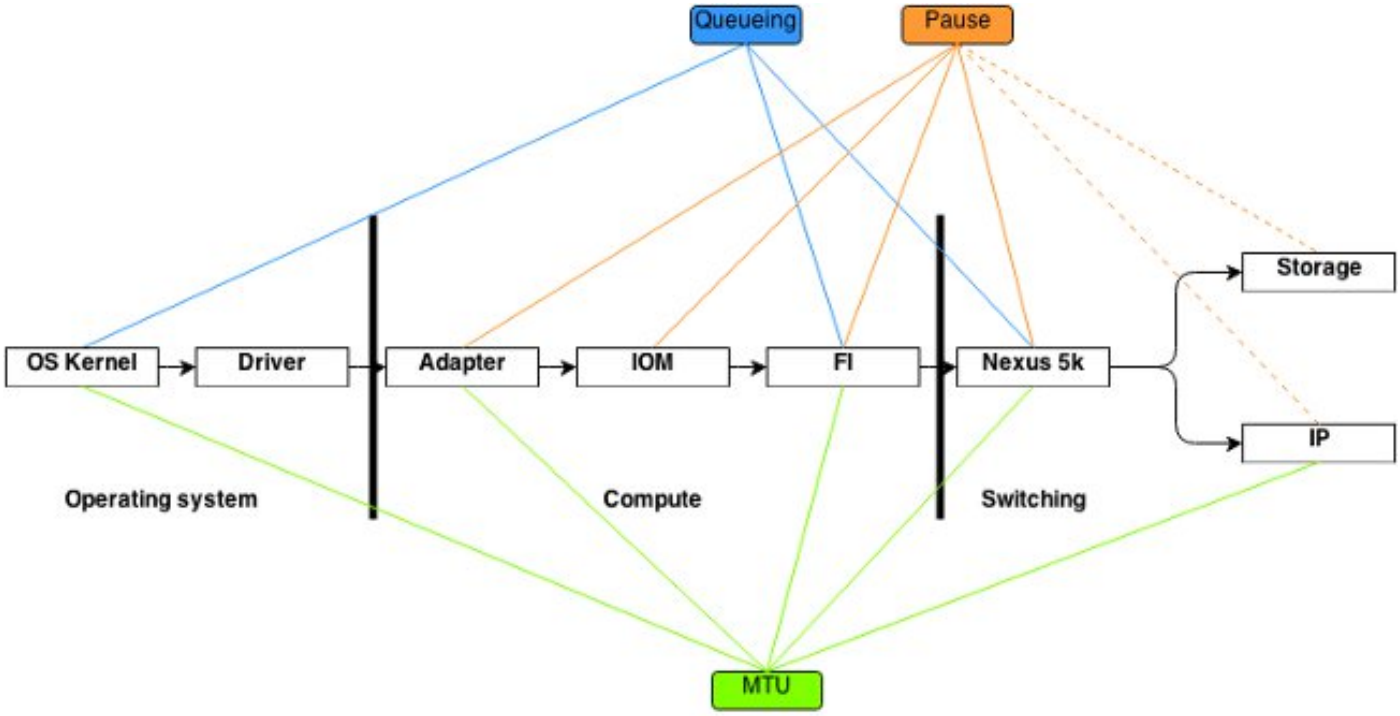
前の例は、エラーが発生していない、非常に負荷の少ないインターフェイスを示しています。 次の例は、別のサーバを示しています。

```
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
```

2 つの興味深い情報の断片により示されているのは、バッファが不足しているか、バッファリングが無効になっていることに加えて、TCP Segment Offloading（TSO）セグメントが処理中のためにアダプタで 96 個のフレームがドロップされたことです。

論理パケット フロー

次の図は、FlexPod 環境における論理パケット フローの概要を示しています。



この図は、フレームが FlexPod 環境を通過する途中、どのコンポーネントを通るのかを示しています。 これは、ブロックの複雑さを示しているわけではなく、特定の機能を設定して確認する場所を記憶するためのものです。

入出力モジュール

論理パケット フロー図に示すように、入出力モジュール（IOM）は UCS を通過するすべての通信の真ん中に位置するコンポーネントです。 シャーシ X 内の IOM に接続するには、**connect iom x** コマンドを入力します。

その他の有用なコマンドを以下に示します。

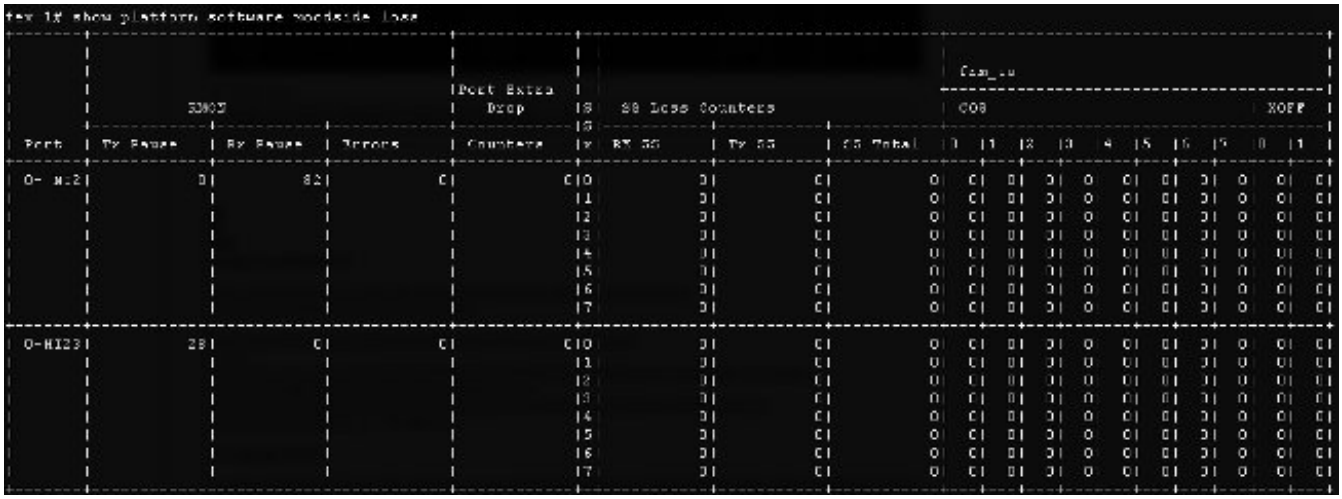
- トポロジ情報：**show platform software [woodside|redwood] sts** コマンドは IOM の観点からトポロジ情報を表示します。

具体的には、FI につながるネットワーク インターフェイス (NI) を表示します。この場合は、8 本のインターフェイスが存在し、そのうちの 4 本がアップしています。 加えて、シャーシ内の特定のブレードにつながるホスト インターフェイス (HI) も表示します。

- トラフィック レート：**show platform software [woodside|redwood] rate** コマンドは、トポロジーおよび HI インターフェイスとブレードのマッピングが判明してから、HI インターフェイスを通過するトラフィックのレートをチェックするために使用されます。

```
fex-1# show platform software woodside rate
+--------+-------------+-------------+-------------+-------------+-------------+-------------+-------+-------+---+
| Port   || Tx Packets | Tx Rate     | Tx Bit      || Rx Packets | Rx Rate     | Rx Bit      |Avg Pkt|Avg Pkt|   |
|        ||            | (pkts/s)    | Rate        ||            | (pkts/s)    | Rate        | (Tx)  | (Rx)  |Err|
+--------++------------+-------------+-------------++------------+-------------+-------------+-------+-------+---+
| 0-BI   ||          13 |           2 |   2.64Kbps  ||           9 |           1 |   3.49Kbps  |   106 |   222 |   |
| 0-CI   ||          17 |           3 |  11.79Kbps  ||          18 |           3 |  13.53Kbps  |   413 |   450 |   |
| 0-NI3  ||          37 |           7 |  66.37Kbps  ||          42 |           8 |  12.44Kbps  |  1101 |   165 |   |
| 0-NI2  ||         245 |          49 | 682.12Kbps  ||          74 |          14 |  19.96Kbps  |  1720 |   148 |   |
| 0-NI1  ||           1 |           0 |   1.74Kbps  ||          11 |           2 |   5.56Kbps  |  1072 |   296 |   |
| 0-NI0  ||           6 |           1 |   5.22Kbps  ||          12 |           2 |   6.63Kbps  |   524 |   325 |   |
| 0-HI31 ||          11 |           2 |   4.65Kbps  ||           1 |           0 | 560.00 bps  |   244 |   332 |   |
| 0-HI27 ||          12 |           2 |   5.28Kbps  ||           1 |           0 | 560.00 bps  |   255 |   332 |   |
| 0-HI26 ||           1 |           0 | 632.00 bps  ||           0 |           0 |   0.00 bps  |   376 |     0 |   |
| 0-HI23 ||          67 |          13 |  16.20Kbps  ||         245 |          49 | 680.94Kbps  |   131 |  1717 |   |
| 0-HI19 ||          29 |           5 |   8.20Kbps  ||          28 |           5 |  61.60Kbps  |   157 |  1355 |   |
| 0-HI18 ||          12 |           2 |   5.15Kbps  ||           2 |           0 | 640.00 bps  |   248 |   182 |   |
+--------++------------+-------------+-------------++------------+-------------+-------------+-------+-------+---+
```

- トラフィック損失：**show platform software [woodside|redwood] loss** コマンドを入力します。 このコマンドを実行すると、損失カウンタが 0 になります。 これにより、インターフェイス単位の一時停止フレームとドロップを確認することができます。



基盤となるインフラストラクチャの機能のため、カウンタは 2 つのコマンドの実行間で損失が発生したインターフェイスに関してのみ表示されます。 この例では、NI2 インターフェイスで 82 個の一時停止フレームが受信され、28 個の一時停止フレームがインターフェイス HI23 (ブレード 3 に取り付けられている) に送信されたことを確認できます。

設計上の考慮事項

FlexPod は、ストレージとデータ ネットワークの柔軟な設定とセットアップを可能にします。 この柔軟性は新たな課題も生みます。 ベスト プラクティス ドキュメントと Cisco Validated Design (CVD) に準拠することが重要です。

- CVD - FlexPod 導入ガイド
- NetApp ストレージ ベスト プラクティス (Flexpod に固有ではない) - Cisco Unified Computing System (UCS) ストレージ コネクティビティのオプションと NetApp ストレージのベスト プラクティス

ポート速度の選択とポート チャネルに関する考慮事項

TAC エンジニアによって確認されている一般的な問題は、ベスト プラクティス ドキュメントで参照されている 10 Gbit イーサネットの代わりに 1 Gbit イーサネットを選択したことによって生じるリンクの過剰使用です。 端的な例として、**単一フロー**のパフォーマンスは 1 つの 10 Gbit リンクよりも 10 本の 1 Gbit リンクの方が低くなります。 ポート チャネルでは、単一フローを単一リンク経由で流すことができます。

Nexus または FI の NX-OS で使用されているロード バランシング方式を確認するには、**show port-channel load-balance** コマンドを入力します。 管理者は、パケットまたはフレームの発信インターフェイスとして選択されたポート チャネルのインターフェイスを確認することもできます。 2 つのホスト間の VLAN49 上のフレームの簡単な例を以下に示します。

```
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
```

ストレージ固有の問題

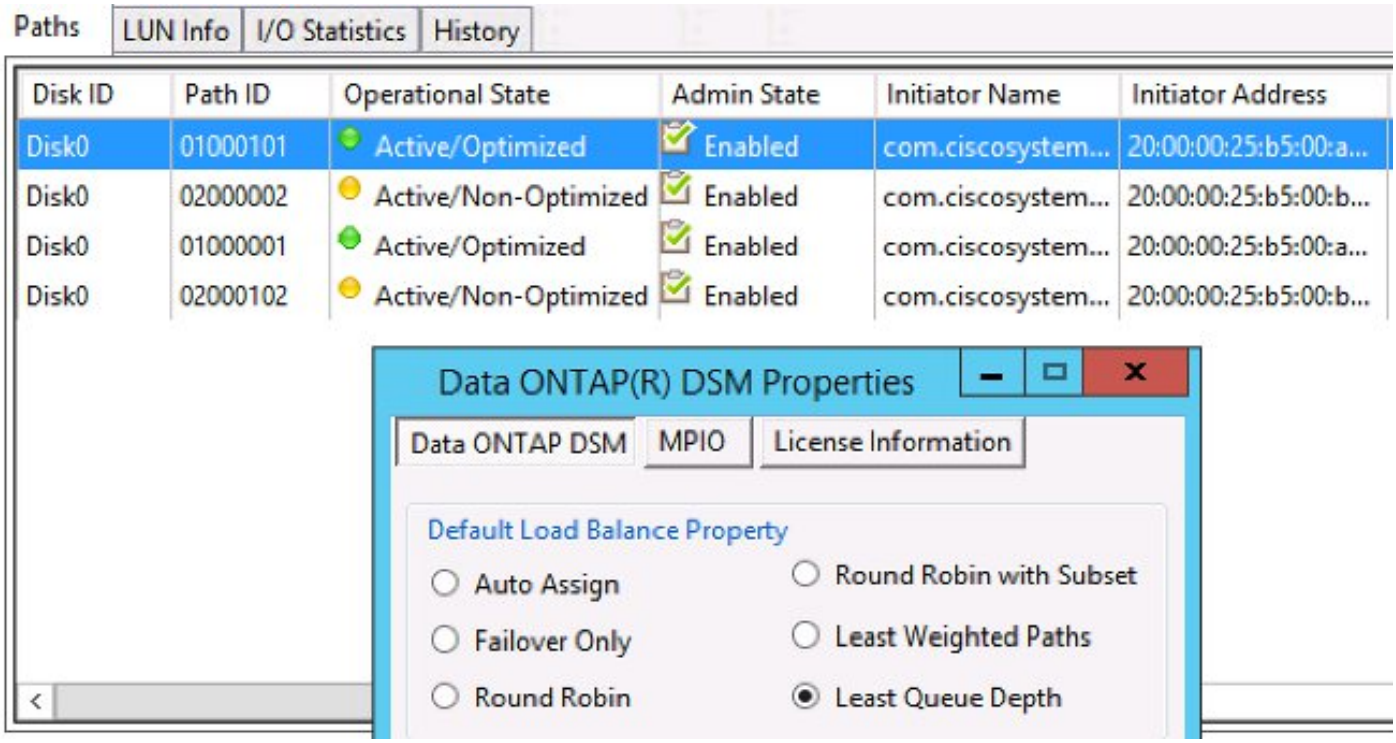
前述の問題は、データ ネットワーキングとストレージ ネットワーキングの両方に共通しています。 完全を期すために、ストレージ エリア ネットワーク（SAN）に固有のパフォーマンス問題についても説明します。 ストレージ プロトコルは復元力に基づいて構築されており、マルチパッシングも追加されています。 非対称論理ユニット割り当て（ALUA）やマルチパス I/O（MPIO）などのテクノロジーの出現によって、柔軟性の向上と新しいオプションが管理者にもたらされました。

ストレージの配置

もう 1 つの留意点がストレージの配置です。 FlexPod 設計では、ストレージを Nexus スイッチに接続するように規定されています。 直接接続されたストレージは CVD に反します。 ベスト プラクティスに従った場合は、直接接続されたストレージがある設計がサポートされます。 ただし、このような設計は、厳密には FlexPod ではありません。

最適なパスの選択

これは技術的には Cisco の問題ではありません。このようなオプションのほとんどがシスコ デバイスからは見えないためです。 最適なパスの選択と設置は一般的な問題です。 最近のデバイス固有のモジュール（DSM）は、複数のパスに存在させることができるため、復元力とロード バランシングを提供するための特定の基準に基づいて最適なものを選択する必要があります。 このスクリーンショットは、Microsoft Windows 用の NetApp DSM とロード バランシング オプションで使用可能な 4 つのパスを示しています。



推奨設定はストレージ ベンダーとの話し合いに基づいて選択する必要があります。 この設定がパフォーマンスの問題に影響を与える場合があります。 TAC が依頼する可能性がある一般的なテストは、ファブリック A またはファブリック B のみを経由する読み取り/書き込みテストです。 通常は、このテストを使用することによって、このドキュメントの「一般的な問題」の項に記載されている状況にパフォーマンス問題を絞り込むことができます。

VM とハイパーバイザのトラフィック共有

この点は、ベンダーに関係なく、計算コンポーネントに特有です。 計算の観点からストレージ ネットワークをハイパーバイザ用にセットアップする簡単な方法は、ファイバごとに 1 つずつの計 2 つのホスト バス アダプタ（HBA）を作成し、その 2 つのインターフェイス経由でブート LUN トラフィックと仮想マシン（VM）ストレージ トラフィックの両方を送ることです。 ブート LUN トラフィックと VM ストレージ トラフィックを分離することを常にお勧めします。 これにより、パフォーマンスが向上するうえ、2 種類のトラフィックを論理的に分離することが可能になります。 例については、「既知の問題」の項を参照してください。

トラブルシューティングのヒント

問題の絞り込み

迅速なトラブルシューティングの場合と同様に、問題を絞り込んで、適切な疑問を提起することが非常に重要です。

- どのデバイス/アプリケーション/VM が影響を受けるか（または受けないか）。
- どのストレージ コントローラが影響を受けるか（または受けないか）。
- どのパスが影響を受けるか（または受けないか）。
- 問題が発生する（または発生しない）頻度はどのくらいか。

Cisco

カウンタの制限

このドキュメント インターフェイスでは、ASIC キューイング カウンタについて説明します。 カウンタからはある時点の状況も把握できるため、カウンタの増加をモニタすることが重要になります。 特定のカウンタは設計上クリアすることができません。 たとえば、前述の carmel ASIC です。

的確な例を示すとすれば、インターフェイス上で CRC や破棄が発生することは望ましくないものの、それらの値が 0 以外になることは想定しておかなければならない場合があります。 カウンタは、遷移中または初期セットアップ中のある時点で増加する可能性があります。 そのため、カウンタの増加と最後にクリアされた時点に注意することが重要です。

コントロール プレーンに関する考慮事項

カウンタの確認は有効ですが、特定のデータ プレーン問題はコントロール プレーン カウンタやツールに簡単に反映されない場合があることを認識しておくことが重要です。 的確な例として、ethanalyzer は UCS と Nexus 5000 の両方で使用可能な非常に便利なツールです。 ただし、コントロール プレーン トラフィックしかキャプチャすることができません。 トラフィック キャプチャは、特に、障害が発生した場所が不明な場合に、TAC から依頼されることがあります。

トラフィックのキャプチャ

エンド ホストで実行される信頼できるトラフィック キャプチャは、パフォーマンス問題に光を当てて短期間で問題を絞り込むことができます。 Nexus 5000 と UCS の両方でトラフィック SPAN が提供されます。 具体的には、特定の HBA とファブリック側を SPAN する UCS のオプションが役に立ちます。 UCS 上のセッションをモニタする場合にトラフィック キャプチャ機能の詳細を確認するには、次のリファレンスを参照してください。

- 物理アダプタと仮想アダプタの UCS トラフィック分析（ビデオ）
- Cisco UCS Manager GUI コンフィギュレーション ガイド - トラフィックのモニタリング

NetApp

NetApp は、次のようなストレージ コントローラをトラブルシュートするためのユーティリティの完全なセットです。

- perfstat：非常に便利なユーティリティです。通常は NetApp サポート担当者が実行します。
- systat：ファイラの状態とファイラの動作に関する情報を提供します（NetApp Support Library）。

最も一般的なコマンドを以下に示します。

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

sysstat -x 2 の出力で、過負荷状態の NetApp アレイまたはディスクを示すものを以下に示します。

- 長期間同じ状態の **CP ty** 列、大量の **:** または **F** を含んでいる
- 長期間同じ状態の **HDD util** 列、**20%** を超えている

次の記事に、NetApp の設定方法が記載されています。 NetApp イーサネット ストレージのベスト プラクティス

- VLAN タギング
- VLAN トランキング
- ジャンボ MTU
- IP ハッシング
- フロー制御の無効化

VMware

ESXi は、トラブルシューティングが可能なセキュア シェル（SSH）アクセスを提供します。 管理者に提供される最も有益なツールが esxtop と perfmon です。

- esxtop：Linux/BSD の top と同様、リアルタイム パフォーマンス関連パラメータをモニタすることができる
  esxtop を使用した ESX/ESXi のストレージ パフォーマンス問題の特定
- perfmon：Microsoft Windows 仮想マシン（VM）をトラブルシュートすることができる
  仮想マシンのパフォーマンス問題を診断するための Windows Perfmon ログ データの収集
- ESXi 上の診断バンドルの収集：vSphere クライアント（653）を使用した VMware

ESX/ESXi の診断情報の収集
- Cisco B シリーズ サーバの VMware vSwitch ロード バランシング要件：IP ハッシュに基づくルートは、UCS 6100 シリーズ ファブリック インターコネクトを使用する Cisco UCS B200 M1/M2 ブレード サーバでサポートされない

既知の問題と機能強化

- Cisco Bug ID CSCuj86736：パッシブ twinax ケーブルを使用した場合に、CRC エラーが増加する可能性があります。 これは Nexus 5000 が DFE を最適化しない場合に発生します。 "Eye height" パラメータが 100 mv を超えていることを確認するには、**show hardware internal carmel eye** コマンドを入力します。 これは、リリース 5.2(1)N1(7) と 7.0(4)N1(1) で修正されています。
- Cisco Bug ID CSCuo76425：前のバグと同様に、UCS ファブリック インターコネクト上でも発生します。 これは、リリース 2.2(3a) で修正されています。
- Cisco Bug ID CSCuo76425：UCS ファブリック インターコネクトを除いて、CSCuj86736 と同じです。
- Cisco Bug ID CSCup40056：「ユニファイド コンピューティング システム仮想マシンのライブ マイグレーションが仮想ファイバ チャネル アダプタで失敗する」に記載されている VM トラフィックとブート トラフィックを共有した場合にタイミングの問題が発生します。
- 低速ドレインの検出と回避：FC と FCoE は頻繁に低速ドレインの影響を受けます。 NX-OS リリース 7.0(0)N1(1) で、これを検出して回避するための手段が導入されました。 機能の詳細については、「Cisco Nexus 5500 シリーズ NX-OS インターフェイス コンフィギュレーション ガイド」と「低速ドレイン デバイスの検出と輻輳の回避」を参照してください。
- Cisco Bug ID CSCuj81245：PALO ベースのカード（VIC1240 など）で FC の中断を引き起こす制限があります。
- Cisco Bug ID CSCuh61202：リリース 2.1(3) へのアップグレード後に、UCS ファームウェア FC が中断し、その他の複数の問題が確認されています。
- Cisco Bug ID CSCtw91018：単一の PALO ベースのアダプタ上で VNIC の MTU 設定が混在している場合に、一部のトラフィック クラスのスタベーションが発生する可能性があります。
- Cisco Bug ID CSCuq40256：ファブリック インターコネクトとサーバ アダプタ間のリンク上で PFC が無効になります。 これにより、ファイバ チャネルの中断をはじめ、順序が不規則なフレームまで、ストレージ側で報告されているさまざまな問題が発生します。 ストレージが接続解除され、他のパフォーマンス問題が報告される場合もあります。

TAC ケース

TAC エンジニアが、調査を開始する前にいくつかの基本情報の収集を依頼する場合があります。

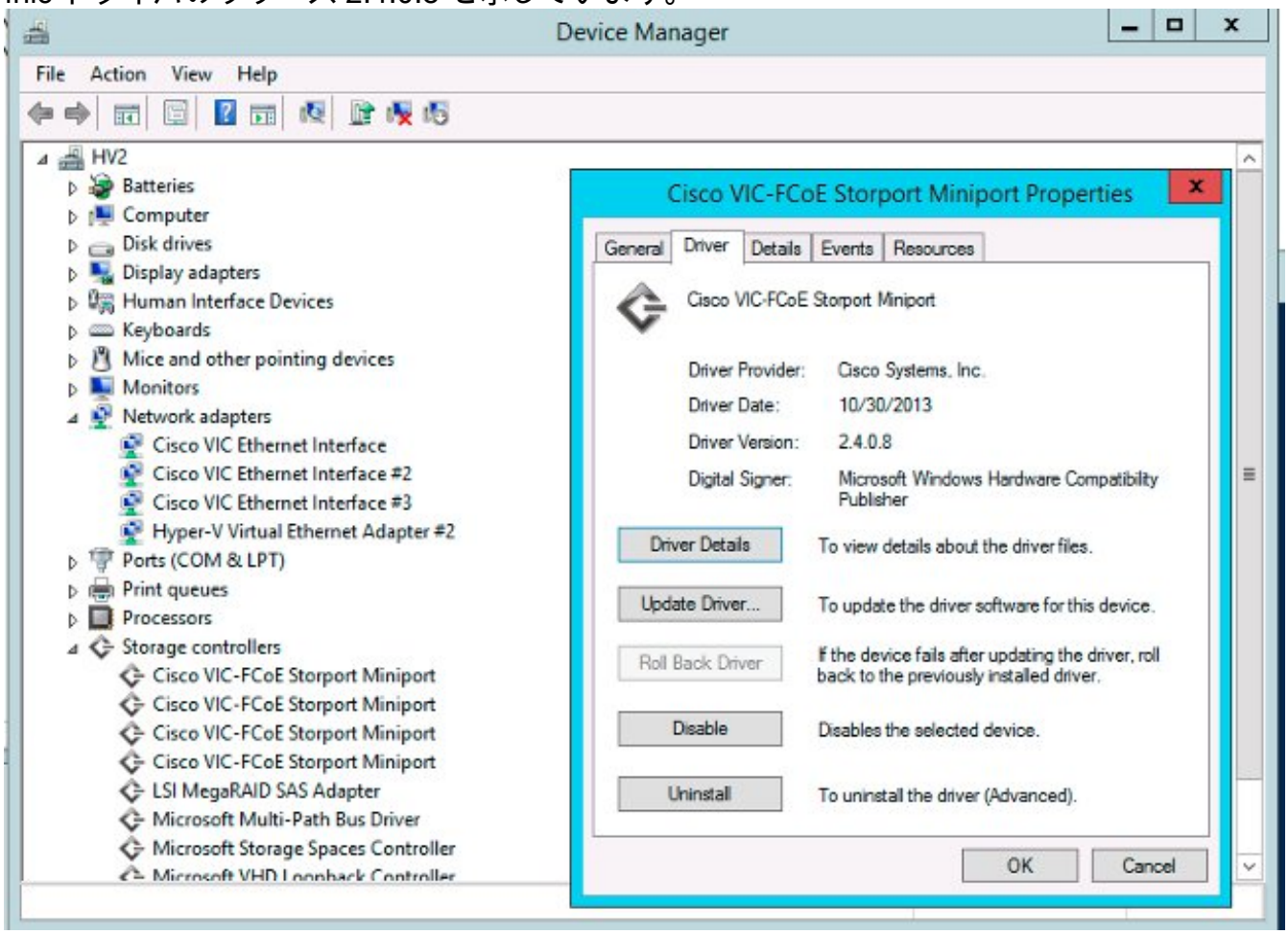
- トポロジ図：ポート番号と回線速度を含めます。これは、絶対に必要です。
- UCSM テクニカル サポート：テクニカル サポート ファイルを収集するためのビジュアル ガイド（B シリーズと C シリーズ）。
- 問題が発生したシャーシの UCS シャーシ テクニカル サポート：前のリンクを参照してください。
- Nexus 5000 テクニカル サポートおよび UCS と NetApp 間のその他のネットワーク デバイスの両方：show tech-support details コマンドの出力のリダイレクション。
- 両方の FI 上での **show queueing interface** コマンドの出力。

```
sysstat -M 2
```

- ESXi 上のホスト ドライバのバージョン：次のコマンドを入力します。 vmkload_mod -s enicvmkload_mod -s fnic
- Linux：

```
sysstat -M 2
```

- Windows：「デバイス マネージャ」のドライバ バージョンをチェックします。 Windows 2012 R2 の例では、3 つの Cisco VIC イーサネット インターフェイス、4 つの VIC FCoE ミニポート インターフェイス（FCoE だけでなく、ファイバ チャネルも担当する）、および fnic ドライバのリリース 2.4.0.8 を示しています。



フィードバック

このドキュメントに関するフィードバックまたはお客様の体験を提出する場合は、フィードバック ボタンを使用してください。 シスコでは、開発が行われたときとフィードバックの受信後にこのマニュアルを更新しています。