Problemi di prestazioni comuni di FlexPod

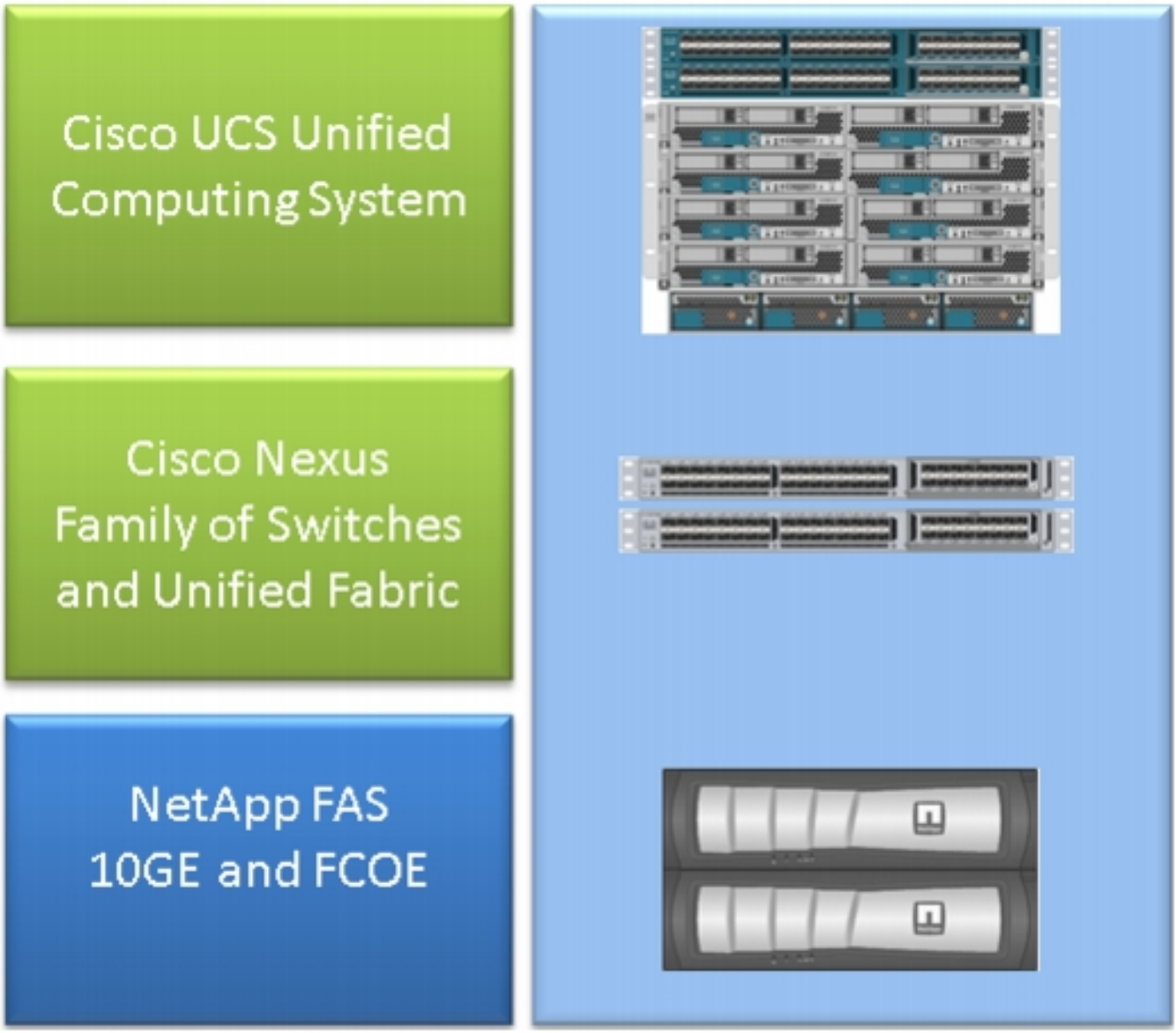
Sommario

Introduzione

Questo documento descrive i problemi di prestazioni comuni negli ambienti FlexPod, fornisce un metodo per risolvere i problemi e fornisce le fasi di mitigazione. È stato concepito come punto di partenza per i clienti che desiderano risolvere i problemi di prestazioni in un ambiente FlexPod. Questo documento è stato redatto in seguito ai problemi riscontrati dal team TAC (Data Center Solutions Technical Assistance Center) negli ultimi mesi.

Panoramica concettuale di FlexPod

Un FlexPod è costituito da un computer UCS (Unified Computing System) collegato tramite switch Nexus allo storage NetApp e alle reti IP.



Il FlexPod più comune è costituito da uno chassis Cisco UCS serie B collegato tramite Fabric Interconnect (FI) agli switch Nexus 5500 e ai filer NetApp. Un'altra soluzione, chiamata FlexPod Express, utilizza uno chassis UCS serie C collegato agli switch Nexus 3000. In questo documento viene descritto il FlexPod più comune.

Considerazioni sulle prestazioni

In ambienti complessi con più parti responsabili, come in genere si vede in un FlexPod, è necessario considerare più aspetti per risolvere il problema. I problemi di prestazioni tipici delle reti IP e di layer 2 derivano da:

- Perdita di pacchetti o frame: la perdita di bit di dati causa un effetto negativo sulle prestazioni delle applicazioni.
- Buffer: se un pacchetto o un frame trascorre troppo tempo in una coda o in un buffer, alcune implicazioni relative alle prestazioni possono essere viste dalle applicazioni, soprattutto in caso di reti di storage. I problemi relativi a latenza, riordino e normalizzazione rientrano in questa categoria.
- Problemi di mancata corrispondenza MTU e frammentazione - un problema comune quando si raggiungono prestazioni superiori. In questa categoria rientrano i problemi relativi alla frammentazione e all'incoerenza dell'MTU.

Ambiente

Èimportante conoscere l'ambiente per il quale vengono misurate le prestazioni. Per circoscrivere correttamente il problema, è necessario porre domande relative al tipo di storage e al protocollo, nonché al sistema operativo e alla posizione del server interessato. Il minimo indispensabile è un diagramma della topologia che delinei la connettività.

Misurazione

Ènecessario sapere cosa viene misurato e come viene misurato. Alcune applicazioni, così come la maggior parte dei fornitori di storage e hypervisor, forniscono misurazioni di qualche tipo che indicano le prestazioni e lo stato del sistema. Queste misurazioni sono un buon punto di partenza in quanto non sostituiscono la maggior parte delle metodologie di risoluzione dei problemi.

Ad esempio, la misurazione della latenza di archiviazione NFS (Network File System) nell'hypervisor potrebbe indicare un calo delle prestazioni, che tuttavia di per sé non implica la rete. Nel caso di un NFS, un semplice ping tra l'host e la rete IP di archiviazione NFS potrebbe indicare se la causa è la rete.

Previsione

Questo punto non può essere sottolineato abbastanza, soprattutto quando si apre una richiesta TAC. Per indicare prestazioni insoddisfacenti, occorre indicare il parametro misurato. Questo include il valore previsto **e** testato. È consigliabile visualizzare i dati precedenti **e** la metodologia di test utilizzata per ottenere tali dati.

Ad esempio: La latenza di 10 ms raggiunta durante il test, con una sola scrittura da un singolo iniziatore a un singolo LUN (Logical Unit Number), potrebbe non essere indicativa della latenza prevista per un sistema completamente caricato.

Problemi di prestazioni in un FlexPod

Poiché questo documento è destinato a servire da riferimento per la maggior parte degli ambienti FlexPod, vengono descritti solo i problemi più frequenti riscontrati dal team TAC responsabile delle soluzioni per data center.

Problemi comuni

In questa sezione vengono illustrati i problemi comuni di storage e delle reti IP/Layer 2.

Perdita di frame e pacchetti

La perdita di frame e pacchetti è il fattore più frequente che influisce sulle prestazioni. Uno dei punti più comuni in cui cercare le indicazioni di un problema è a livello di interfaccia. Dalla CLI di Nexus 5000 o UCS Nexus Operating System (NX-OS), accedere all'**interfaccia show | sec "è attivo" | egrep ^(Eth|fc)|discard|drop|**comando **CRC**. Per le interfacce attive, elenca il nome ed elimina contatori e rilasci. Analogamente, viene visualizzata una panoramica completa quando si immette il comando **show interface counters error** che mostra le statistiche degli errori per tutte le interfacce.

Mondo Ethernet

Èimportante sapere che i contatori con valore diverso da 0 potrebbero non indicare un problema. In alcuni scenari, tali contatori potrebbero essere stati generati durante la configurazione iniziale o in precedenti modifiche operative. Un aumento dei contatori dovrebbe essere monitorato.

Èinoltre possibile raccogliere i contatori dal livello ASIC, che potrebbe essere più indicativo. In particolare, per l'errore CRC (Cyclic Redundancy Check) sulle interfacce, un comando preferito di TAC è **show hardware internal carmel crc**. Carmel è il nome dell'ASIC responsabile dell'inoltro a livello di porto.

Un output simile può essere ottenuto dagli switch serie 6100 FI o Nexus 5600 per ciascuna porta. Per FI 6100, l'ASIC Gatos, immettere questo comando:

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

Per Nexus 5600, da bigsur ASIC, immettere questo comando:

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

Il comando per carmel ASIC indica dove sono stati ricevuti i pacchetti CRC e dove sono stati inoltrati, e soprattutto se sono stati calpestati o meno.

Poiché il funzionamento di Nexus 5000 e UCS NX-OS è cut-through, i frame in modalità di commutazione con FCS (Frame Check Sequence) errato vengono sottoposti a staging solo prima dell'inoltro. È importante scoprire da dove provengono i fotogrammi danneggiati.

```
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 |
```

Nell'esempio vengono mostrati i pacchetti con stomped provenienti da Eth 1/17 e Eth 1/18, che è un uplink al Nexus 5000. Si può supporre che quei fotogrammi siano stati successivamente inviati a Eth 1/34, come Eth 1/17 + Eth 1/18 rx Stomp = Eth 1/34 tx Stomp.

Un'immagine simile della serie Nexus 5000 mostra:

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

Questo output mostra i CRC ricevuti su due collegamenti e contrassegnati come stomps prima dell'inoltro. Per ulteriori informazioni, vedere la Guida alla risoluzione dei problemi di Nexus 5000.

Mondo Fibre Channel

Un modo semplice per cercare le perdite (errori, errori, CRC, esaurimento del credito B2B) è tramite il comando **show interface counters fc**.

Questo comando, disponibile su Nexus 5000 e Fabric Interconnect, fornisce una buona indicazione di ciò che accade nel mondo Fibre Channel.

Ad esempio:

```
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
(...)
```

L'interfaccia non è occupata e l'output mostra che non si sono verificati errori o scarti.

Sono state inoltre evidenziate transizioni creditizie B2B da 0; a causa degli ID bug Cisco CSCue80063 e CSCut08353, questi contatori non sono attendibili. Funzionano correttamente su Cisco MDS, ma non sulle piattaforme UCS Nexus5k. Inoltre, è possibile verificare l'ID bug Cisco CSCsz95889.

Analogamente a carmel nel mondo Ethernet per Fibre Channel (FC), la struttura fc-mac può essere utilizzata. Ad esempio, per la porta fc2/1, immettere il comando **show hardware internal fc-mac 2 port 1 statistics**. I contatori presentati sono in formato esadecimale.

```
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
```

Nell'output vengono visualizzati 15 scarti sull'input. Corrisponde a FCP_CNTR_PIF_RX_DROP che conta fino a 0xf (15 in decimali). Queste informazioni possono essere nuovamente correlate alle informazioni di FWM (Forwarding Manager).

```
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
```

In questo modo, tuttavia, viene indicato all'amministratore il numero di rilasci e il numero ASIC corrispondente. È necessario eseguire una query per ottenere informazioni sul motivo dell'eliminazione di ASIC.

```
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 questo caso, il traffico è stato interrotto dall'Access Control List (ACL) in entrata, generalmente nello zoning FC.

MTU non corrispondente

Negli ambienti FlexPod è importante supportare l'impostazione end-to-end della MTU (Maximum Transition Unit) per le applicazioni e i protocolli in cui è richiesta. Nella maggior parte degli ambienti, si tratta di Fibre Channel over Ethernet (FCoE) e di frame jumbo.

Inoltre, in caso di frammentazione, si può prevedere un calo delle prestazioni. In caso di protocolli come NFS (Network File System) e iSCSI (Internet Small Computer System Interface), è importante verificare e provare le MTU (Maximum Transmission Unit) IP end-to-end e MSS (Maximum Segment Size) TCP.

Sia che si tratti di risolvere i problemi dei frame Jumbo o di FCoE, è importante ricordare che per funzionare correttamente entrambi gli ambienti necessitano di una configurazione coerente e di un contrassegno CoS (Class of Service).

Nel caso di UCS e Nexus, un comando utile per convalidare l'impostazione MTU per interfaccia, per gruppo QoS è **show queuing interface | in coda|qos-group|MTU**.

Visualizzazione dell'MTU sulle piattaforme Nexus 5000 e UCS

Un aspetto noto sia di UCS che di Nexus è la visualizzazione delle MTU sull'interfaccia. Questo output mostra un'interfaccia configurata per mettere in coda i frame Jumbo e FCoE:

```
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)
```

Allo stesso tempo, il comando **show interface** visualizza 1500 byte:

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

Se confrontato con le informazioni dell'ASIC carmel, l'ASIC mostra la capacità MTU di una data porta.

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

Questa mancata corrispondenza MTU è attesa sulle piattaforme sopra menzionate, e potrebbe fuorviare i neofiti.

Configurazione completa

L'unico modo per garantire prestazioni adeguate è una configurazione completa e coerente. La configurazione dei frame jumbo e i passaggi per il lato Cisco, oltre a VMware ESXi, sono descritti in UCS con esempio di configurazione dell'MTU jumbo end-to-end di VMware ESXi.

Esempio di configurazione di uplink UCS FCoE: mostra una configurazione UCS e Nexus 5000. Vedere l'Appendice A nel documento di riferimento per una descrizione della configurazione di base di Nexus 5000.

La configurazione della connettività FCoE per un blade Cisco UCS è incentrata sulla configurazione UCS per FCoE. Esempio di configurazione Nexus 5000 NPIV FCoE con FCoE

NPV Attached UCS focalizzato sulla configurazione Nexus.

Test di frame jumbo end-to-end

La maggior parte dei sistemi operativi moderni offre la possibilità di testare una corretta configurazione jumbo frame con un semplice test ICMP (Internet Control Message Protocol).

Calcolo

9000 byte - Intestazione IP senza opzioni (20 byte) - Intestazione ICMP (8 byte) = 8972 byte di dati

Comandi nei sistemi operativi più diffusi

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
```

Problemi correlati al buffer

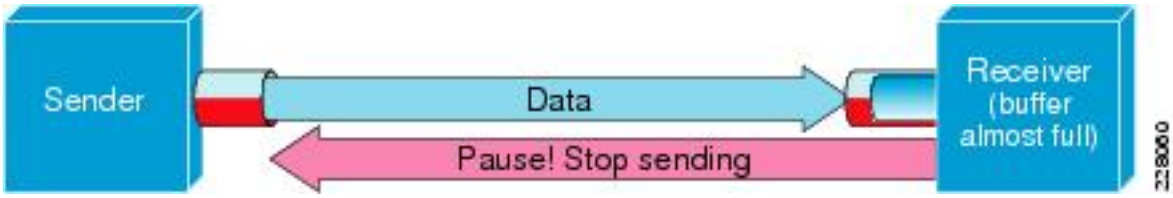
Il buffering e altri problemi relativi alla latenza sono tra le cause comuni di riduzione delle prestazioni nell'ambiente FlexPod. Non tutti i problemi segnalati come latenza derivano da problemi di buffer effettivi, alcune misurazioni potrebbero indicare una latenza end-to-end. Ad esempio, nel caso di NFS, il periodo di tempo indicato potrebbe essere necessario per la corretta lettura/scrittura nello storage e non per l'effettiva latenza di rete.

La congestione è la causa più comune per la memorizzazione nel buffer. Nel mondo del layer 2, la congestione può causare buffer e anche code di gocce di frame. Per evitare cadute durante i periodi di congestione, sono stati introdotti i frame di pausa IEEE 802.3x e il controllo del flusso di priorità (PFC). Entrambi si affidano alla richiesta del punto finale di tenere le trasmissioni per un breve periodo di tempo mentre la congestione dura. Questo problema può essere causato da una congestione della rete (sovraccaricare la rete ricevuta con una quantità di dati) o dal passaggio di un frame con priorità, come nel caso di FCoE.

Controllo flusso - 802.3x

Per verificare quali interfacce hanno il controllo del flusso abilitato, immettere il comando **show interface flowcontrol**. È importante seguire le raccomandazioni del fornitore dello storage per quanto riguarda l'abilitazione del controllo del flusso.

Di seguito è riportata un'illustrazione del funzionamento del controllo del flusso 802.3x.

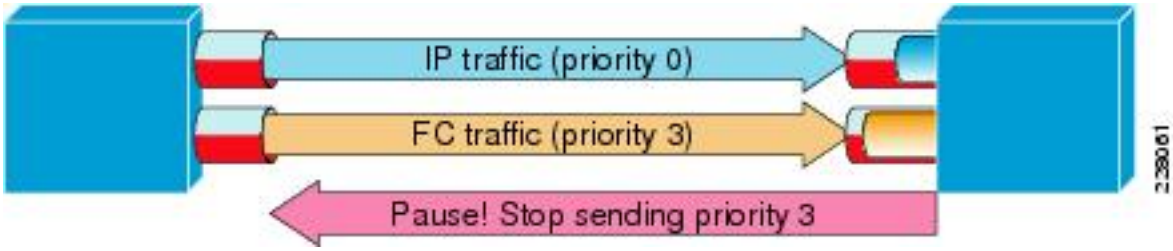


PFC - 802.1Qbb

PFC non è richiesto per tutte le impostazioni, ma è consigliato per la maggior parte di esse. Per verificare le interfacce con PFC abilitato, usare il comando **show interface priority-flow-control | i** Il comando può essere eseguito su UCS NX-OS e Nexus 5000.

Le interfacce tra FI e Nexus 5000 dovrebbero essere visibili in tale elenco. In caso contrario, è necessario verificare la configurazione QoS. Per trarre vantaggio dalla funzionalità PFC, la qualità del servizio deve essere uniforme e completa. Per verificare il motivo per cui il PFC non è presente su un'interfaccia specifica, immettere il comando **show system internal dcbx log interface ethernet x/y** per ottenere il log DCBX (Data Center Bridging Capabilities Exchange Protocol).

Di seguito è riportata un'illustrazione che illustra il funzionamento dei frame di pausa con PFC.



Il comando **show interface priority-flow-control** permette all'amministratore di osservare il comportamento delle classi per-QoS dei frame di pausa delle priorità.

Di seguito è riportato un esempio:

```
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)
```

Questo output mostra che, nella seconda classe, il dispositivo stava solo trasmettendo (TX) un frame PPP.

In questo caso, la porta Ethernet 1/1 è rivolta verso l'IOM e, anche se la porta generale non avrà il PFC abilitato, potrebbe elaborare i frame PPP per le porte FEX.

```
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 questo caso, sono coinvolte interfacce FEX.

```
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
```

Le porte FEX interessate possono essere controllate anche mediante **show fex X detail** dove X è il numero dello chassis.

```
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
```

Per ulteriori informazioni sui meccanismi di pausa, consultare i seguenti documenti.

- Operazioni Fibre Channel over Ethernet
- White paper su Unified Fabric - Fibre Channel over Ethernet (FCoE)

Accodamento ignorati

Sia Nexus 5000 che UCS NX-OS tengono traccia dei rigetti in entrata dovuti all'accodamento per gruppo QOS. Ad esempio:

```
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)
```

L'eliminazione in ingresso *deve* avvenire solo nelle code configurate per consentire le eliminazioni.

Gli scarti delle code in ingresso possono verificarsi per i motivi seguenti:

- SPAN (Switched Port Analyzer)/sessione di monitoraggio abilitata su alcune interfacce (vedere l'ID bug Cisco CSCur25521)
- Pressione di un'altra interfaccia; i frame di pausa sono in genere visualizzati quando abilitati
- Traffico indirizzato alla CPU

Problema del driver

Cisco fornisce due driver di sistema operativo per UCS, enic e fnic. Enic è responsabile della connettività Ethernet, mentre fnic è responsabile della connettività Fibre Channel e FCoE. È **molto importante** che i driver enic e fnic corrispondano esattamente a quanto specificato nella matrice di interoperabilità UCS. I problemi introdotti da driver errati vanno dalla perdita di pacchetti e dall'aggiunta di latenza a un processo di avvio più lungo o alla totale mancanza di connettività.

Informazioni scheda

Una scheda di rete fornita da Cisco può fornire una buona misurazione del traffico trasmesso, nonché delle cadute. In questo esempio viene illustrato come connettersi allo chassis X, al server Y e all'adattatore 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.
```

Da qui, l'amministratore può accedere alla funzionalità Monitoring Center for Performance (MCP).

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

La funzione MCP permette di monitorare l'uso del traffico per interfaccia logica (LIF).

```
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
```

Lo chassis 1, il server 1 e la scheda 1 dispongono di due schede di interfaccia di rete virtuali (VNIC, Virtual Network Interface Card) associate alle interfacce virtuali (Virtual Ethernet o Virtual Fibre Channel) 834 e 836. Tali schede hanno i numeri 2 e 3. Le statistiche per LIF 2 e 3 possono essere controllate come mostrato di seguito:

```
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
```

Èimportante notare che all'amministratore di UCS vengono fornite le colonne totale e delta (tra due esecuzioni successive di stati di vita), nonché il carico di traffico corrente per LIF e informazioni su eventuali errori che si sono verificati.

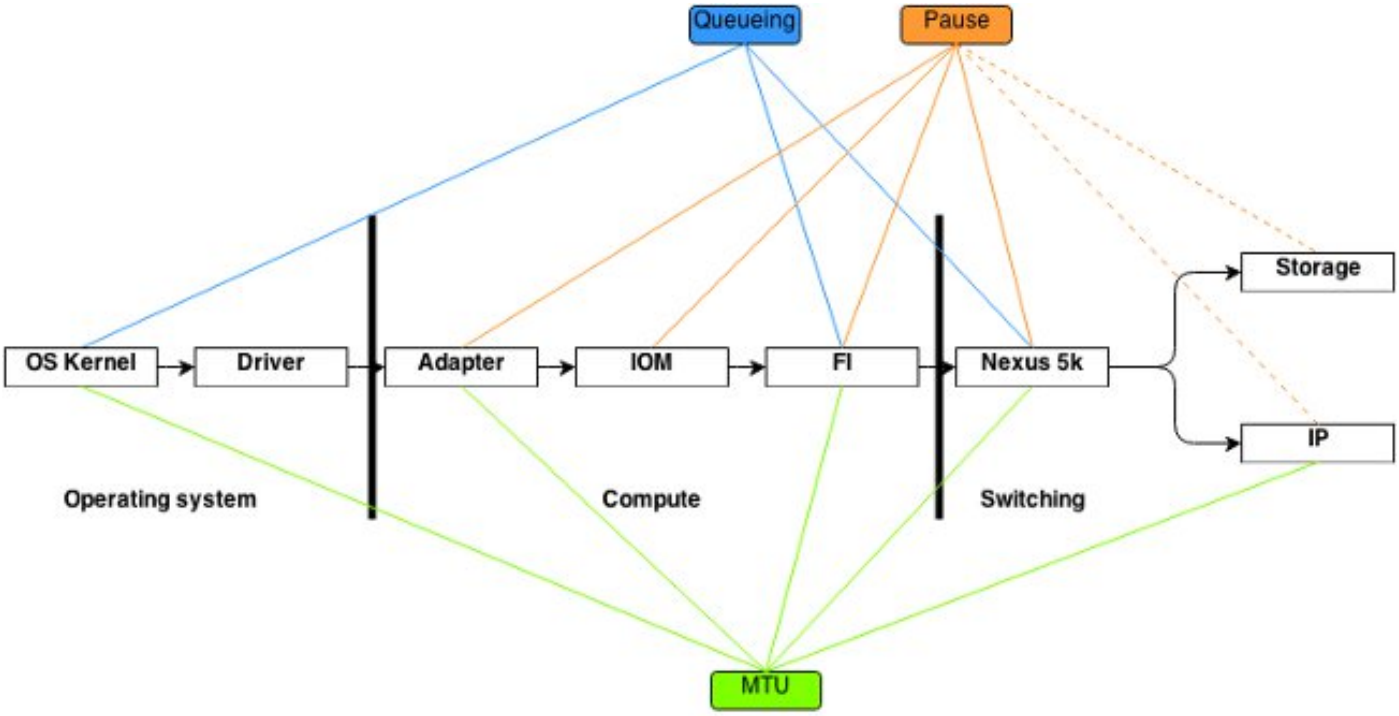
Nell'esempio precedente vengono mostrate le interfacce senza errori con un carico di lavoro minimo. Nell'esempio viene mostrato un server diverso.

```
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
```

Due interessanti bit di informazioni mostrano che 96 frame sono stati scartati dalla scheda di rete a causa della mancanza di buffer o di buffer disabilitati e, inoltre, per i segmenti TCP di offload del segmento (TSO) in fase di elaborazione.

Flusso di pacchetti logico

Il diagramma qui illustrato delinea il flusso logico dei pacchetti in un ambiente FlexPod.



Questo diagramma è inteso come un'analisi dei componenti di un frame che passano attraverso l'ambiente FlexPod. Non riflette la complessità di nessuno dei blocchi ed è semplicemente un modo per memorizzare dove particolari funzionalità devono essere configurate e verificate.

Modulo di ingresso/uscita

Come mostrato nel diagramma di flusso dei pacchetti logico, il modulo di input/output (IOM) è un componente al centro di tutte le comunicazioni che passano attraverso l'UCS. Per connettersi all'IOM nello chassis X, immettere il comando **connect iom x**.

Di seguito sono riportati altri comandi utili:

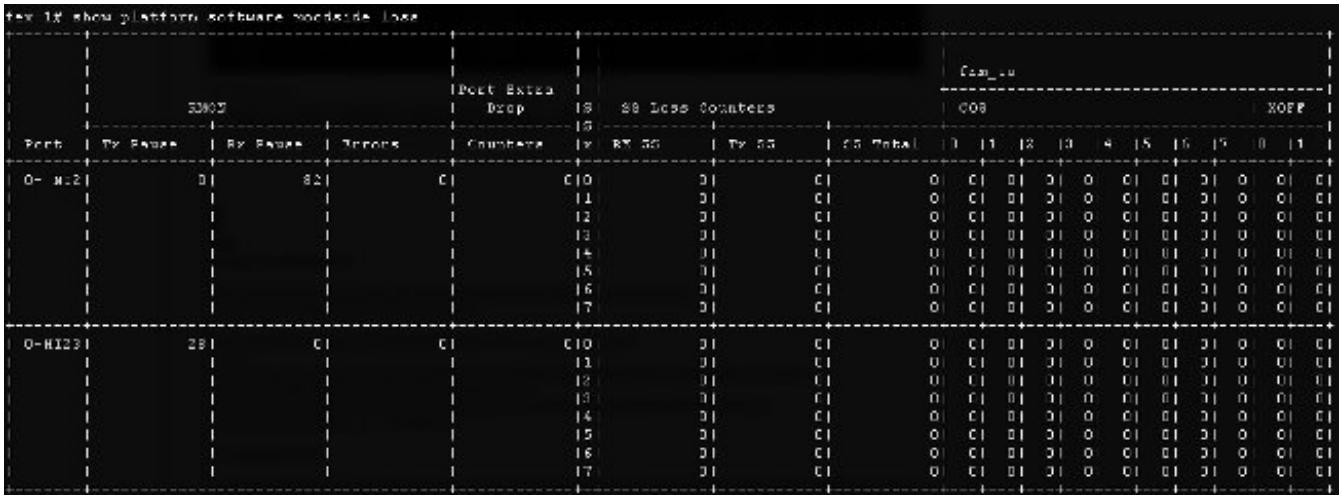
- Informazioni sulla topologia - il comando **show platform software [woodside|redwood] sts** mostra informazioni topologiche dal punto di vista dell'IOM.

Mostra le interfacce di rete (NI) che conducono agli FI, in questo caso sono otto, con quattro attive. Inoltre, mostra le interfacce host (HI) che conducono, all'interno dello chassis, a determinati blade.

- Velocità del traffico: il comando **show platform software [woodside|redwood] rate** viene usato per controllare la velocità del traffico che passa attraverso le interfacce HI una volta che la topologia e il mapping dell'interfaccia Hi a blade sono noti.

```
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 |   5.63Kbps  |   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 |   |
+--------+-------------+-----------+-------------+-------------+-----------+-------------+---------+---+
```

- Perdita di traffico - immettere il comando **show platform software [woodside|redwood]** loss. L'esecuzione di questo comando azzera i contatori delle perdite. Permette di vedere i frame di pausa e le perdite per singola interfaccia.



A causa del modo in cui funziona l'infrastruttura sottostante, i contatori vengono visualizzati solo per le interfacce che hanno sperimentato perdite tra l'esecuzione dei due comandi. In questo esempio, l'interfaccia NI2 ha ricevuto 82 frame di pausa e 28 frame di pausa sono stati trasmessi all'interfaccia HI23, che si sa essere collegata al blade 3.

Considerazioni sulla progettazione

Un FlexPod consente una configurazione flessibile e la configurazione di storage e reti di dati. La flessibilità comporta anche ulteriori sfide. È di fondamentale importanza seguire i documenti relativi alle best practice e un Cisco Validated Design (CVD):

- CVD - Guida all'installazione di FlexPod
- Best practice per lo storage NetApp (non specifiche di Flexd) - Opzioni di connettività dello storage UCS (Cisco Unified Computing System) e best practice con lo storage NetApp

Selezione della velocità delle porte e considerazioni sul canale delle porte

Un problema comune rilevato dai tecnici TAC è l'sovrautilizzo dei collegamenti a causa della selezione di una rete Ethernet da 1 Gbit anziché da 10 Gbit, a cui si fa riferimento nei documenti relativi alle best practice. Ad esempio, le prestazioni del **flusso singolo** non saranno migliori su dieci collegamenti da 1 Gbit rispetto a un collegamento da 10 Gbit. Nel canale della porta un singolo flusso può passare su un singolo collegamento.

Per conoscere il metodo di bilanciamento del carico utilizzato su Nexus e/o sul sistema operativo NX di FI, immettere il comando **show port-channel load-balance**. L'amministratore può anche scoprire quale interfaccia di un canale della porta verrà scelta come interfaccia in uscita di un pacchetto o di un frame. Di seguito è riportato un semplice esempio di frame sulla VLAN49 tra due host:

```
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
```

Problemi specifici di storage

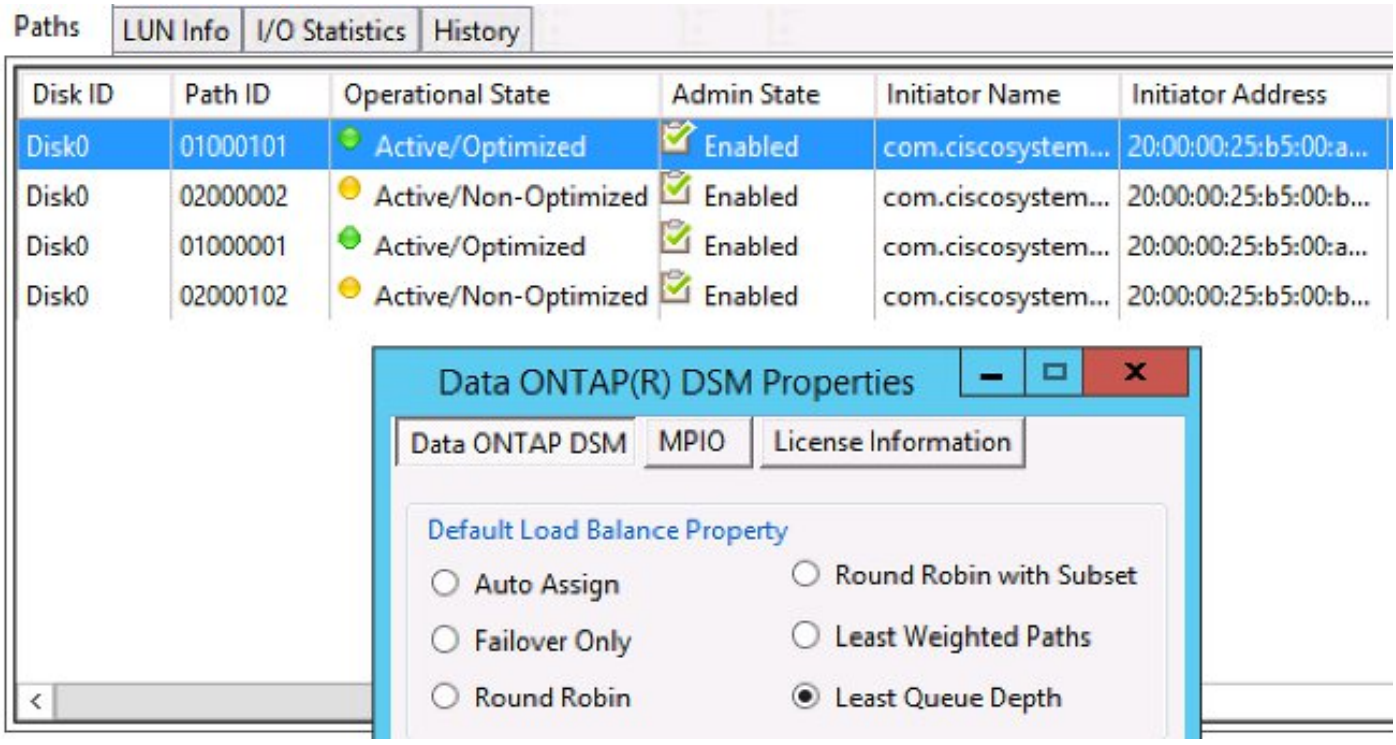
I problemi discussi in precedenza sono comuni alle reti di dati e di storage. Per motivi di completezza, vengono inoltre citati i problemi di prestazioni specifici delle reti SAN (Storage Area Network). I protocolli di storage sono stati creati con resilienza e il multi-pathing è ancora migliorato. Con l'avvento di tecnologie quali ALUA (Asymmetric Logical Unit Assignment) e MPIO (Multi-path IO), agli amministratori vengono offerte maggiore flessibilità e opzioni.

Posizionamento storage

Un'altra considerazione è il posizionamento dello storage. Un design FlexPod richiede il collegamento dello storage sugli switch Nexus. Lo storage collegato direttamente non è conforme a CVD. Se vengono seguite le best practice, sono supportati i progetti con storage a collegamento diretto. Allo stesso tempo, questi progetti non sono strettamente FlexPod.

Selezione ottimale del percorso

Tecnicamente non è un problema di Cisco, in quanto la maggior parte di queste opzioni è trasparente per i dispositivi Cisco. È un problema comune scegliere e attenersi a un percorso ottimale. Un moderno modulo DSM (Device Specific Module) può essere presentato con più percorsi e deve sceglierne uno o più ottimali, in base a determinati criteri per fornire resilienza e bilanciamento del carico. In questa schermata vengono mostrati quattro percorsi disponibili per NetApp DSM per Microsoft Windows e le opzioni di bilanciamento del carico.



Le impostazioni consigliate devono essere scelte in base a una discussione con il fornitore dello storage. Tali impostazioni possono influire sui problemi di prestazioni. Un tipico test che TAC potrebbe richiedere è un test di lettura/scrittura solo per il fabric A o il fabric B. In questo modo è possibile limitare i problemi di prestazioni alle situazioni descritte nella sezione "Problemi comuni" di questo documento.

Condivisione del traffico tra VM e hypervisor

Questo punto è specifico del componente di elaborazione, indipendentemente dal fornitore. Un modo semplice per configurare una rete di storage per gli hypervisor dal punto di vista del calcolo consiste nel creare due HBA (Host Bus Adapter), uno per ciascuna fibra, ed eseguire sia il traffico delle LUN di avvio che il traffico di storage delle macchine virtuali (VM) su queste due interfacce. Si consiglia sempre di dividere il traffico delle LUN di avvio e il traffico dello storage delle VM. In questo modo è possibile ottenere prestazioni migliori e inoltre effettuare una divisione logica tra i due tipi di traffico. Per un esempio, vedere la sezione "Problemi noti".

Suggerimenti per la risoluzione dei problemi

Riduzione del problema

Come nel caso di una rapida risoluzione dei problemi, è molto importante circoscrivere il problema e porre le domande giuste.

- Quali dispositivi/applicazioni/macchine virtuali sono interessati (o meno)?
- Quali controller di storage sono interessati (o non sono interessati)?
- Quali percorsi sono interessati (o meno)?
- Con quale frequenza viene visualizzato il problema?

Cisco

Limitazioni dei contatori

In questa interfaccia documento vengono illustrati i contatori di accodamento ASIC. Poiché i contatori forniscono una vista in un determinato momento, è importante controllare l'aumento dei contatori. Alcuni contatori non possono essere cancellati in base alla progettazione. Per esempio, l'ASIC carmelo menzionato in precedenza.

Per fare un esempio chiaro, la presenza di CRC o di scarti su un'interfaccia potrebbe non essere ideale, ma è probabile che i valori siano diversi da zero. I contatori potrebbero essere aumentati in un determinato momento, probabilmente durante la transizione o la configurazione iniziale. Per questo è importante notare l'aumento dei contatori e quando è stata l'ultima volta che sono stati azzerati.

Considerazioni sui piani di controllo

Anche se è utile esaminare i contatori, è importante sapere che alcuni problemi relativi al piano dati potrebbero non trovare una facile riflessione per controllare i contatori e gli strumenti del piano dati. L'etanalyzer è uno strumento molto utile disponibile sia su UCS che su Nexus 5000. Tuttavia, può solo catturare il traffico del control plane. Il TAC richiede spesso una cattura del traffico, soprattutto quando non è chiaro dove stia la colpa.

Acquisisci traffico

Un'acquisizione affidabile del traffico effettuata sugli host terminali può far luce su un problema di prestazioni e restringerlo abbastanza rapidamente. Sia Nexus 5000 che UCS offrono traffic SPAN. In particolare, le opzioni UCS di SPANing su particolari HBA e lati di struttura sono utili. Per ulteriori informazioni sulle funzionalità di acquisizione del traffico quando si monitora una sessione su UCS, vedere i seguenti riferimenti:

- Analisi del traffico UCS per adattatori fisici e virtuali (video)
- Guida alla configurazione dell'interfaccia utente di Cisco UCS Manager - Monitoraggio del traffico

NetApp

NetApp offre una serie completa di utility per la risoluzione dei problemi dei controller di storage, tra cui:

- perfstat: un'utilità molto utile, generalmente eseguita per il personale di supporto NetApp
- systat - fornisce informazioni sull'utilizzo del filer e sulle relative attività - Libreria di supporto NetApp

I comandi più comuni sono:

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

Di seguito sono riportati alcuni elementi da cercare nell'output **sysstat -x 2** che potrebbero indicare un sovraccarico dell'array o dei dischi NetApp:

- Colonna ty **CP** sostenuto con **:** o **F**
- Colonna **util disco rigido** sostenuto superiore al **20%**

In questo articolo viene descritto come configurare NetApp: Procedure ottimali per lo storage Ethernet NetApp .

- Tagging VLAN
- Trunking VLAN
- MTU jumbo
- Hashing IP
- Disabilita controllo flusso

VMware

ESXi fornisce l'accesso SSH (Secure Shell), tramite il quale è possibile risolvere i problemi. Tra gli strumenti più utili forniti agli amministratori vi sono esxtop e perfmon.

- esxtop - come Linux/BSD top, consente agli utenti di monitorare i parametri relativi alle prestazioni in tempo reale
  Utilizzo di esxtop per identificare i problemi di prestazioni dello storage per ESX / ESXi
- perfmon - consente agli utenti di risolvere i problemi relativi a macchine virtuali (VM) Microsoft Windows

Raccolta dei dati di registro di Windows Perfmon per diagnosticare i problemi di prestazioni della macchina virtuale
- Raccolta del bundle diagnostico su ESXi - Raccolta delle informazioni diagnostiche per VMware ESX/ESXi mediante il client vSphere (653)
- Requisito di bilanciamento del carico VMware vSwitch per server Cisco serie B - Il routing basato su hash IP non è supportato con i server blade Cisco UCS B200 M1/M2 che utilizzano interconnessioni fabric UCS serie 6100

Problemi noti e miglioramenti

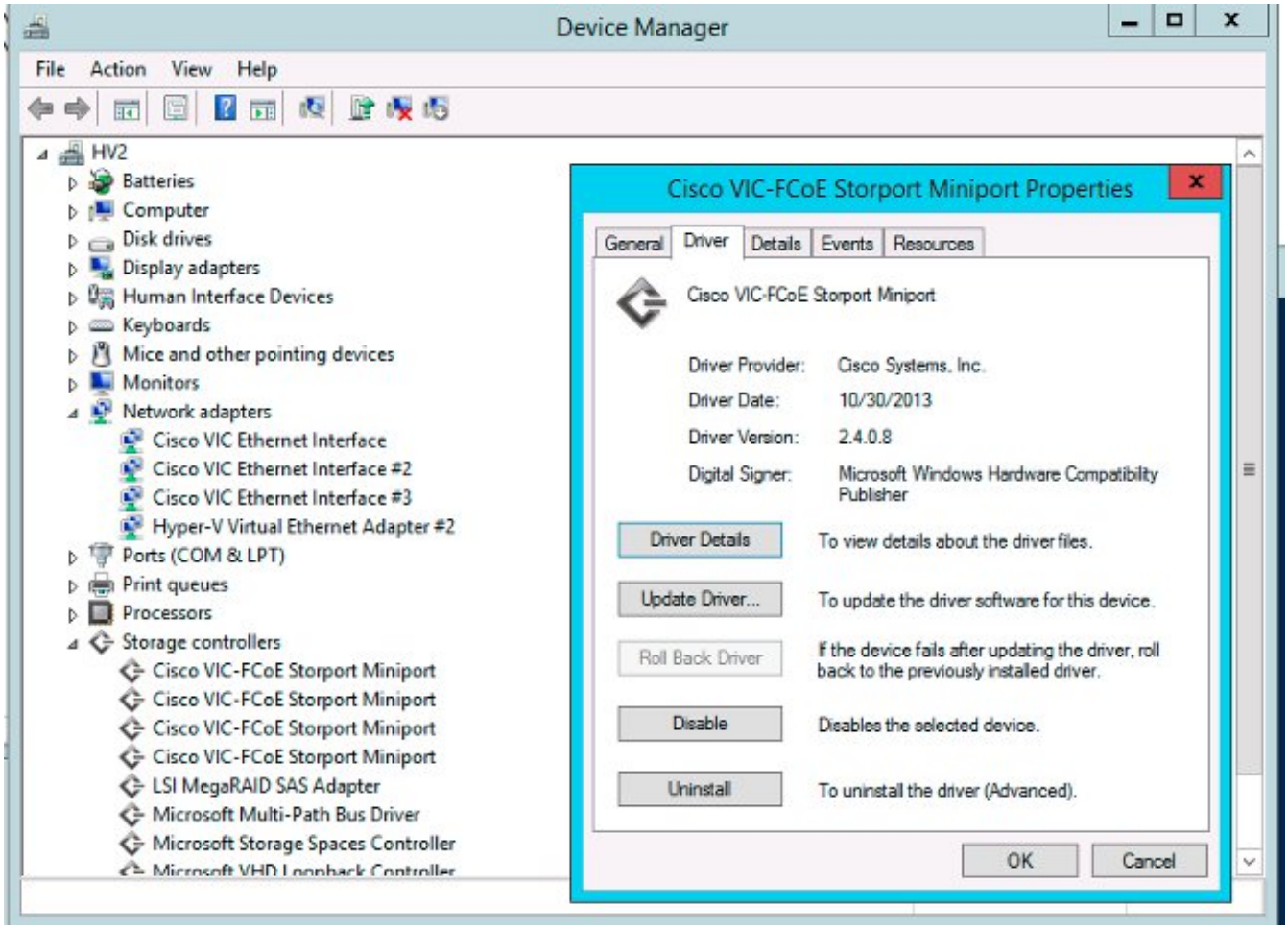
- ID bug Cisco CSCuj86736 - con cavi twinax passivi, gli errori CRC potrebbero aumentare. Ciò si verifica quando Nexus 5000 non ottimizza DFE. Immettere il comando **show hardware internal carmel eye** per verificare che il parametro "Eye height" sia superiore a 100 mv. Questa condizione è stata risolta nelle release 5.2(1)N1(7) e 7.0(4)N1(1).
- ID bug Cisco CSCuo76425 - simile al bug precedente e presente anche sulle interconnessioni dell'infrastruttura UCS. Questo problema è risolto nella release 2.2(3a).
- ID bug Cisco CSCuo76425 - uguale al bug CSCuj86736 ad eccezione di UCS Fabric Interconnect.
- Cisco bug ID CSCup40056 - Problema di tempo causato dalla condivisione del traffico di avvio con il traffico della VM descritto in Unified Computing System Virtual Machine Live Migration non riuscito con le schede Fibre Channel virtuali.
- Rilevamento e prevenzione del drenaggio lento: molto spesso FC e FCoE sono influenzati dal drenaggio lento. NX-OS release 7.0(0)N1(1) introduce i mezzi per rilevarla ed evitarla. Per ulteriori informazioni, fare riferimento alla Guida alla configurazione delle interfacce NX-OS di Cisco Nexus serie 5500 e a Slow Drain Device Detection and Congestion Avoidance.
- ID bug Cisco CSCuj81245 - esiste una limitazione nelle schede basate su PALO (VIC1240 e altre) che causa l'interruzione della connessione FC.
- ID bug Cisco CSCuh61202 - dopo l'aggiornamento alla release 2.1(3), è possibile rilevare gli errori FC del firmware UCS e molti altri problemi.
- ID bug Cisco CSCtw91018 - una combinazione di impostazioni MTU per le VNIC su una singola scheda basata su PALO può causare la fame per alcune classi di traffico.
- ID bug Cisco CSCuq40256 - disabilita il PFC sui collegamenti tra l'interconnessione fabric e le schede server. Ciò causerà una serie di problemi che iniziano con interruzioni Fibre Channel e frame non in ordine riportati sul lato storage. È possibile che vengano segnalate disconnessioni dello storage e altri problemi di prestazioni.

Casi TAC

In molti casi, il tecnico di TAC ti chiederà di raccogliere alcune informazioni di base prima di avviare un'indagine.

- Diagramma topologico - che include i numeri di porta e la velocità della linea, assolutamente necessario.
- Supporto tecnico UCSM - Guida visiva per la raccolta dei file del supporto tecnico (serie B e C).
- Supporto tecnico UCS per uno chassis che presenta problemi - vedere il collegamento precedente.

- Supporto tecnico per Nexus 5000 e altri dispositivi di rete tra UCS e NetApp - Reindirizzamento dell'output del comando show tech-support details.
- Output del comando **show queueing interface** su entrambi gli FI.
```
connect nxos A|B
show queuing interface | no-more
show interface priority-flow-control | no-more
show interface flowcontrol | no-more.
```
- Versioni dei driver host su ESXi eseguono - immettere i seguenti comandi: vmkload_mod -s enicvmkload_mod -s fnic
- Linux-
```
dmesg | egrep -i 'enic|fnic'
```
- Windows - verificare la versione del driver in "Gestione periferiche". Un esempio di Windows 2012 R2 mostra tre interfacce Cisco VIC Ethernet e quattro interfacce miniport VIC FCoE (responsabili anche di Fibre Channel, non solo FCoE) e la release 2.4.0.8 del driver fnic.



Feedback

Utilizza il pulsante per inviare commenti e suggerimenti su questo documento o sulle tue esperienze. Il presente documento verrà costantemente aggiornato in base ai nuovi sviluppi e dopo la ricezione dei commenti.