White paper

Propagação de congestionamentos e como evitá-la

Este documento descreve como a disseminação de congestionamento (também conhecida como drenagem lenta) pode impactar a rede da área de armazenamento (SAN), as medições usadas para descrever a severidade de cada tipo de congestionamento para o Connectrix- B-Series e a Série MDS, bem como as medidas preventivas que podem ser tomadas para evitar os efeitos da disseminação do congestionamento

Maio de 2019

Propagação de congestionamentos e como evitá-la | H17762.2 |

Revisões

Data	Descrição
Maio de 2019	Versão inicial

Agradecimentos

Este artigo foi produzido pelos seguintes membros da equipe de engenharia de armazenamento da Dell EMC:
Autores:
Alan Rajapa
Erik Smith

Sumário

1 Prefácio

Este documento descreve como a disseminação de congestionamento (também conhecida como drenagem lenta) pode impactar a rede da área de armazenamento (SAN), as medições usadas para descrever a severidade de cada tipo de congestionamento para o Connectrix- B-Series e a Série MDS, bem como as medidas preventivas que podem ser tomadas para evitar os efeitos da disseminação do congestionamento

Como parte de um esforço para melhorar e aumentar o desempenho e os recursos de sua linha de produtos, a Dell EMC lança periodicamente revisões de versões de seus produtos de hardware e software. Por isso, algumas funções descritas neste documento podem não ser compatíveis com todas as revisões de software ou hardware usadas no momento. Para obter as informações mais atuais sobre os recursos do produto,
consulte as notas da versão do produto.

Se um produto não funcionar corretamente ou conforme descrito neste documento, entre em contato com seu representante da Dell EMC.

Público

Este livro técnico é destinado à equipe de campo da Dell EMC, inclusive consultores de tecnologia, arquitetos de armazenamento e administradores e operadores envolvidos na aquisição, no gerenciamento, na operação ou no projeto de um ambiente de armazenamento em rede que contém dispositivos da EMC e do host.

Documentação relacionada

todas as notas da versão e documentação relacionadas podem ser encontradas em https://dell.com/support. Clique em **Suporte por produto**, informe o nome do produto e clique em **Documentação**.

Matriz de suporte da Dell EMC e Interoperabilidade de navegação do E-Lab

Para obter as informações mais atualizadas, sempre consulte a *Matriz de suporte da Dell EMC,* disponível por meio do E-Lab Interoperability Navigator (ELN) em https://www.dellemc.com/pt-br/products/interoperability/elab.htm#tab0=2

Onde obter ajuda

As informações sobre licenciamento, suporte e produtos da Dell EMC podem ser obtidas no site de suporte on-line da Dell EMC, conforme descrito a seguir.

Obs.: Para abrir um chamado por meio do site de suporte on-line da Dell EMC, você deve ter um contrato de suporte válido. Entre em contato com o representante de vendas da Dell EMC para obter detalhes de como adquirir um contrato de suporte válido ou para tirar dúvidas sobre sua conta.

Informações do produto

Para obter informações sobre documentação, notas da versão, atualizações de software ou sobre produtos, licenciamento e serviços da Dell EMC, visite o site de suporte on-line da Dell EMC (registro obrigatório) em:
https://www.dell.com/support

Suporte técnico

A Dell EMC oferece uma série de opções de suporte.

Suporte por produto -

A Dell EMC oferece informações consolidadas e específicas de produto no site:
https://support.dell.com/products
As páginas da Web de suporte por produto oferecem links rápidos para documentação, white papers, sugestões (como artigos da base de conhecimento frequentemente usados) e downloads, bem como conteúdo mais dinâmico, como apresentações, discussões, entradas relevantes do fórum de atendimento ao cliente e um link para o bate-papo on-line da Dell EMC.

Bate-papo on-line da Dell EMC

Abra uma sessão de bate-papo ou de mensagem instantânea com um engenheiro de suporte da Dell EMC.

Suporte eLicensing

Para ativar seus direitos e obter os arquivos de licença, acesse o Centro de serviços em https://dell.com/support, conforme descrito na carta do LAC (License Authorization Code, código de autorização de licenciamento) enviada a você por e-mail.

2 Visão geral

O objetivo deste white paper é:

1. Descrever como a propagação do congestionamento (também conhecida como drenagem lenta) pode afetar sua rede da área de armazenamento (SAN),

2. Definir as métricas usadas para descrever cada severidade e tipo de congestionamento para os produtos Connectrix - B-Series e Série MDS.

3. Descrever as medidas preventivas que podem ser usadas para evitar os efeitos da propagação do congestionamento e

4. Demonstrar como usar as informações acima para detectar, impedir e remediar a propagação do congestionamento devido a superatribuição.

PRÉ-REQUISITOS

Obs.:

Este documento presume que as versões de software a seguir estão sendo usadas. As etapas podem ser diferentes em versões mais antigas.
Consulte o apêndice para obter detalhes que descrevem como habilitar os recursos necessários.

1. A Dell EMC Unisphere para PowerMax e VMAX está instalado e em execução, e o array foi registrado para coletar dados de desempenho.
   https://www.dell.com/support/products/27045_Unisphere-for-/Documentation/?source=promotion

2. GUIs de gerenciamento de SAN estão instaladas.
   a. Para fabrics Brocade: Connectrix Manager Data Center Edition (CMCNE) 14.x ou superior
      Download:
      https://www.dell.com/search/?text=CMCNE%2014&searchLang=pt_BR&facetResource=DOWN
      Guia do administrador:
      https://www.dell.com/search/?text=CMCNE%2014% 20admin%20guide&searchLang=en_US
   b. Para fabrics Cisco: Cisco Data Center Network Manager(DCNM) 10.x ou superior
      Download:
      https://www.dell.com/support/search/?text=DCNM%2010&facetResource=DOWN

      Guia do administrador:
      https://www.cisco.com/c/en/us/support/cloud-systems-management/prime-data-center-network-manager/products-installation-guides-list.html

3. O microcódigo de switches de SAN deve uma destas opções:
   a. Brocade: Fabric O.S 7.4.1d ou superior
      Download:
      https://www.dell.com/support/search/?text=Brocade%20FOS%20download&searchLang=en_US&facetResource=DOWN

   b. Cisco NX-OS 6.2(13) ou superior
      Download:
      https://www.dell.com/support/search/?text=NX-OS%20download

4. Todas as licenças de monitoramento de desempenho necessárias estão instaladas.

    a. A Brocade exige uma licença do MAPS:
    https://docs.broadcom.com/docs/53-1005239-04

    b. A Cisco requer a licença do pacote do servidor DCNM-SAN:
    https://www.cisco.com/c/en/us/support/cloud-systems-management/prime-data-center-network-manager/products-installation-guides-list.html

    c. PowerMAX e VMAX exigem uma eLicense do Dell EMC Unisphere. Consulte a página 21 do PDF a seguir
    para obter mais detalhes
    https://www.dell.com/collateral/TechnicalDocument/docu88904.pdf

3 O que é a propagação do congestionamento?

A transporte de dados de e para um storage array exige que todos os dados sejam entregues ao destino em tempo hábil. Isso é válido, especialmente, para protocolos de armazenamento baseados em Block que usam SCSI (por exemplo, Fibre Channel - FCP). Embora as razões exatas para isso fiquem fora do escopo deste white paper, mais detalhes podem ser encontrados na seção “Congestionamento e contrapressão” do livro técnico *Conceitos e protocolos do armazenamento em rede*: (https://www.dellemc.com/pt-br/products/interoperability/elab.htm#tab0=1hardware/technical-documentation/h4331-networked-storage-cncpts-prtcls-sol-gde.pdf).

Assim como qualquer outro protocolo de rede, o Fibre Channel (FC) precisa garantir o fornecimento de dados em tempo hábil, em uma ampla variedade de situações comuns de congestionamento de rede. O mecanismo usado pelo FC se concentra na prevenção de perda de quadros por meio do controle de fluxo buffer para buffer. Por esse motivo, o FC é considerado um “protocolo sem perdas”.

Embora os mecanismos de controle de fluxo usados por cada protocolo sejam um pouco diferentes, o FC e outros protocolos sem perdas (por exemplo, DCB Ethernet e Infiniband) impedem o overflow de buffer nas duas extremidades do link, permitindo que o transmissor determine quando o receptor na outra extremidade está se aproximando da capacidade. Quando essa determinação for feita, uma porta interromperá a transmissão de dados até que a outra extremidade do link indique que ele está pronto para receber dados adicionais. Embora um transmissor esteja nesse estado, não é possível transmitir os quadros e; portanto, dizemos que ele está sofrendo congestionamento. Se um transmissor experimenta o congestionamento por um período longo o suficiente, esse congestionamento pode se propagar para a origem. Esse fenômeno é conhecido como propagação de congestionamentos e um exemplo é exibido na sequência de diagramas a seguir.

A Figura 1 é um exemplo de uma SAN que não está sofrendo congestionamento. Os Hosts 1 e 2 estão executando comandos de leitura para o array.
Como o array e o host são conectados em 16 Gbps e a largura de banda de ISL é suficiente (ou seja, 32 GB), não há nenhum congestionamento na SAN.

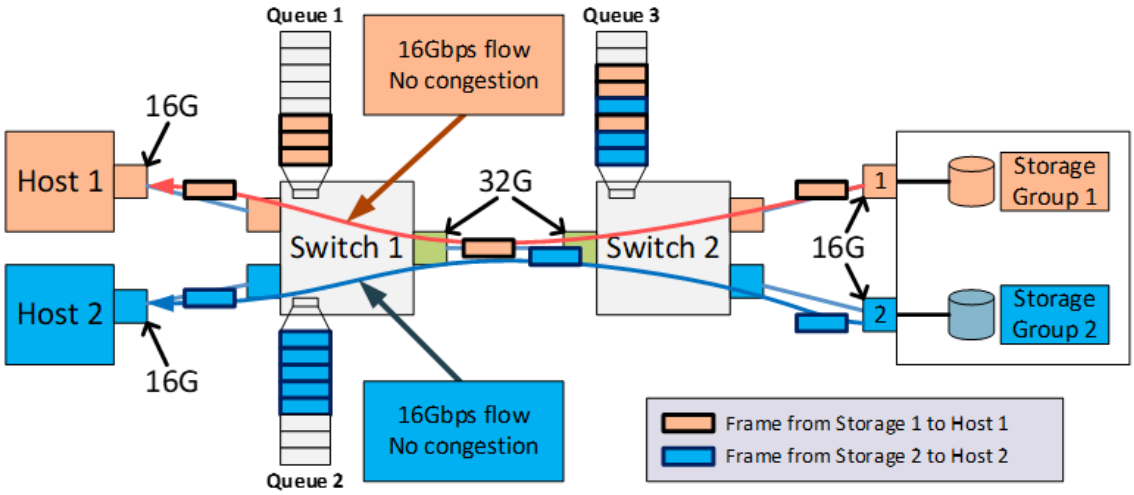


Figura 1 sem congestionamento

A Figura 2 mostra um exemplo de uma SAN que está sofrendo propagação de congestionamento devido a superatribuição. Observe que a única diferença entre as duas figuras é que na Figura 3, a interface no Host 1 foi configurada para ser executada a 4 Gbps em vez de 16 Gbps. Assim que isso for feito, se a interface do array transmitir dados a uma taxa maior que a velocidade do HBA conectado (ou seja, 4G), o Host 1 não poderá receber os dados na taxa que está sendo transmitida, e o impacto imediato é o enfileiramento de quadros. Conforme o preenchimento da Fila 1, o congestionamento se espalha de volta para a origem dos dados. Como o Host 1 e o Host 2 estão compartilhando o mesmo Inter Switch Link (ISL), esse congestionamento impacta o “fluxo inocente” entre o Host 2 e o Armazenamento 2, reduzindo o throughput de 16 Gbps para 4 Gbps.

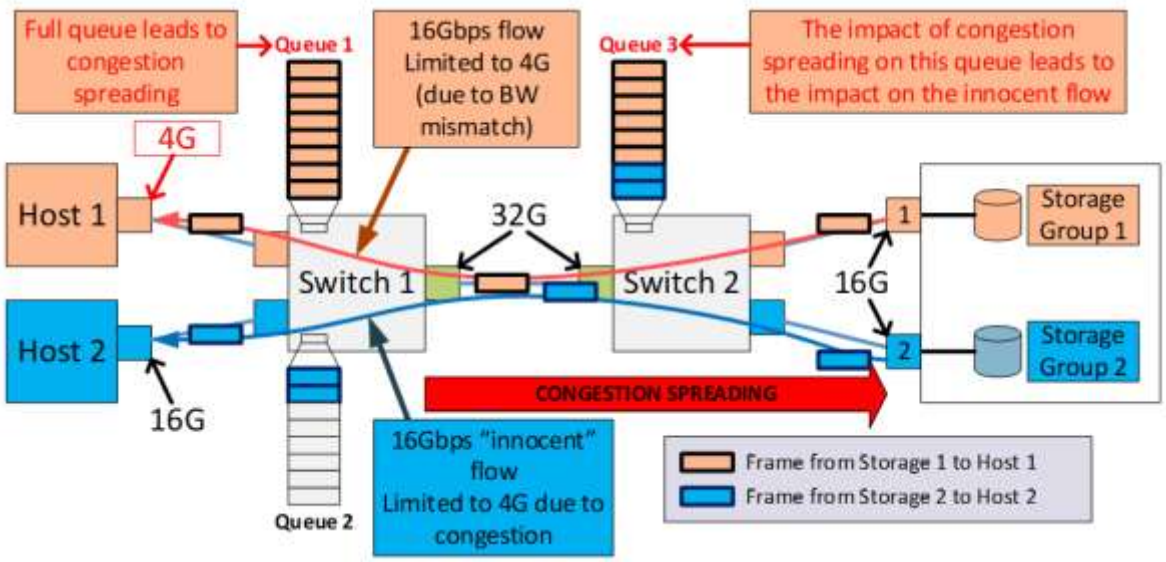


Figura 2 Congestionamento

As informações adicionais sobre congestionamento e propagação do congestionamento podem ser encontradas na seção “Congestionamento e contrapressão” do livro técnico *Conceitos e protocolos do armazenamento em rede*: (https://www.dell.com/collateral/hardware/technical-documentation/h4331-networked-storage-cncpts-prtcls-sol-gde.pdf). É importante observar que a superatribuição é apenas uma das possíveis causas da propagação do congestionamento. Outras causas serão explicadas nas seções a seguir.

- Taxa de congestionamento (taxa C)

A taxa de congestionamento ou a taxa C é um valor calculado que pode ajudar a detectar quando a propagação do congestionamento está ocorrendo. Por exemplo, a Figura 3 mostra um host (ou seja, o Host 1) que é capaz de receber dados a uma taxa de 4 Gbps, mas está recebendo dados de uma interface de armazenamento que é capaz de transmitir dados a 16 Gbps.

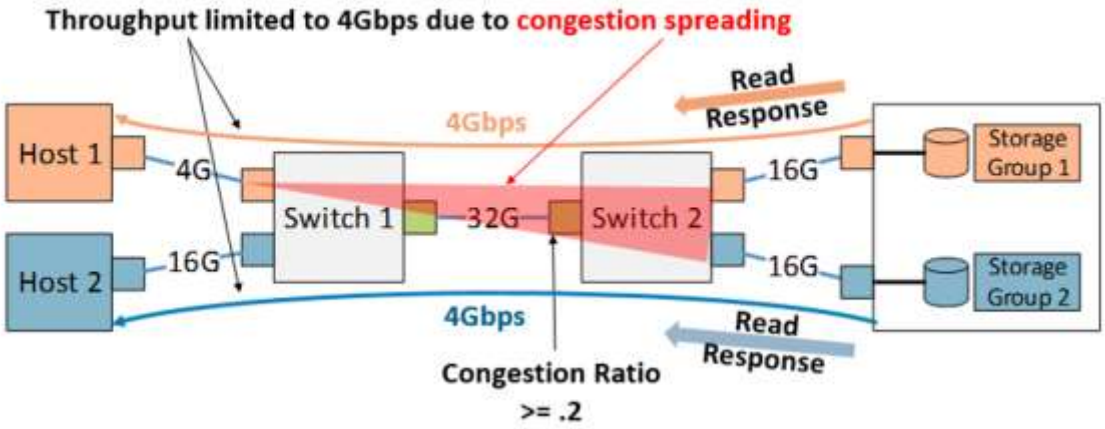


Figura 3 Taxa de congestionamento

Parte do que torna esses problemas difíceis de detectar e solucionar é que, do ponto de vista da interface 4G no comutador 1, tudo está certo. A interface do comutador está transmitindo quadros o mais rápido que o link permite. No entanto, já que o armazenamento está transmitindo dados na taxa permitida pelo link (ou seja, 16 Gbps), haverá 12 Gbps (16 Gbps - 4 Gbps) de largura de banda que será transmitida pelo array e precisará ser enfileirada em algum lugar. Esse enfileiramento geralmente acontece na fabric e é a causa da propagação do congestionamento. Conforme mencionado acima, um método que pode ser usado para detectar a presença da propagação do congestionamento é o cálculo da taxa de congestionamento. Para isso, adote o contador “Tempo gasto com crédito de transmissão zero” e divida-o pelo contador de quadros transmitidos, e você terá o número (normalmente entre 0 e 1). Se esse número for maior que. 2, você terá congestionamento. A propósito, esse número precisa ser calculado de acordo com a interface; portanto, talvez seja melhor simplesmente criar um script do processo para verificar esse valor.

4 Propagação do congestionamento devido à superatribuição

O estudo de caso a seguir baseia-se na propagação do congestionamento devido à superatribuição. A topologia deste estudo de caso é exibida na Figura 4 abaixo. Neste estudo de caso, você aprenderá sobre as ferramentas e técnicas que estão disponíveis atualmente para ajudar a detectar e evitar que esse problema ocorra.

Obs.: A propagação do congestionamento é um problema extremamente difícil de detectar e resolver. Isso se deve, principalmente, à incapacidade da geração atual de ferramentas de gerenciamento de oferecer uma indicação clara de que o problema está ocorrendo, bem como de permitir que você forneça apenas uma orientação real sobre como resolver o problema. Como resultado, a solução desses problemas exige que o usuário final entenda qual é a situação e, em seguida, saiba como usar as ferramentas que estão disponíveis atualmente para extrair conclusões dos dados limitados.

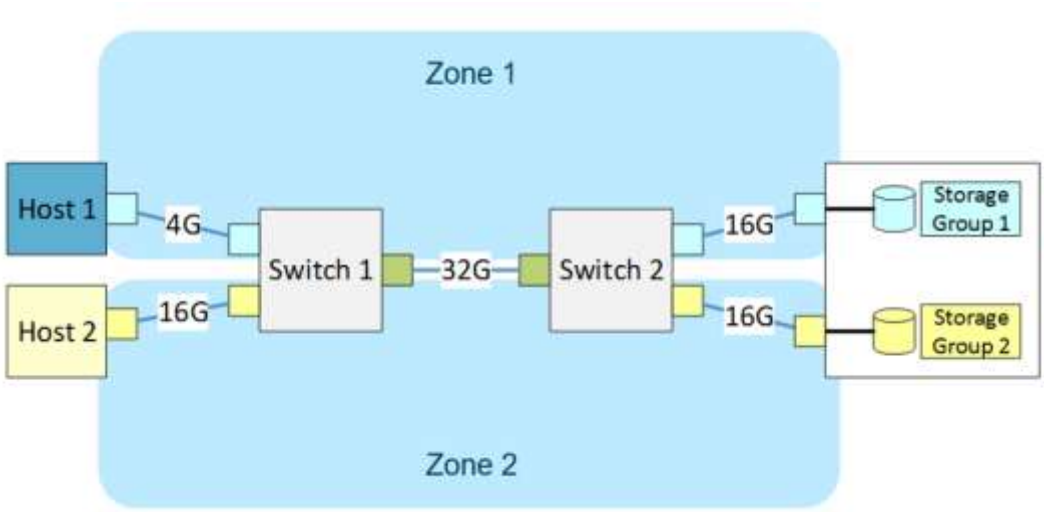


Figura 4 Topologia de caso do estudo de caso de superatribuição

- Cenário:

O Usuário 1 tinha um aplicativo existente em funcionamento no Host 2 (HBA de 16G) que executava a E/S em vários tamanhos de block e de fila e em padrões de E/S. Esse aplicativo está sendo executado por bastante tempo nesse ambiente e não experimentou nenhum problema até o momento. No início deste mês, o Usuário 2 decidiu carregar um aplicativo no Host 1 (HBA de 4G) para fins de teste. Inicialmente, tudo estava bem no ambiente em relação ao desempenho e à latência. No entanto, o Usuário 1 recentemente começou a observar problemas de desempenho com o aplicativo.

- Visão geral da solução de problemas

Para solucionar os problemas em geral, primeiro você deve entender como os elementos são executados e estão configurados quando estão em suas condições ideais de funcionamento. Como você sabe, uma SAN tem muitas peças móveis que compõem o ecossistema, portanto é muito importante criar um perfil de ambiente que consiste em perfis dos três principais componentes que compõem uma SAN: Aplicativo(s), SAN fabric e armazenamento.

A criação desses perfis da linha de base em vários componentes em seu ambiente fornecerá a capacidade necessária para identificar facilmente os problemas quando eles surgirem. Deve-se observar que a criação desses perfis não é uma tarefa feita uma única vez. Você deve reunir constantemente os dados da linha de base durante toda a vida útil de seu ambiente para não apenas solucionar problemas, mas também planejar ainda mais crescimento e expansão.

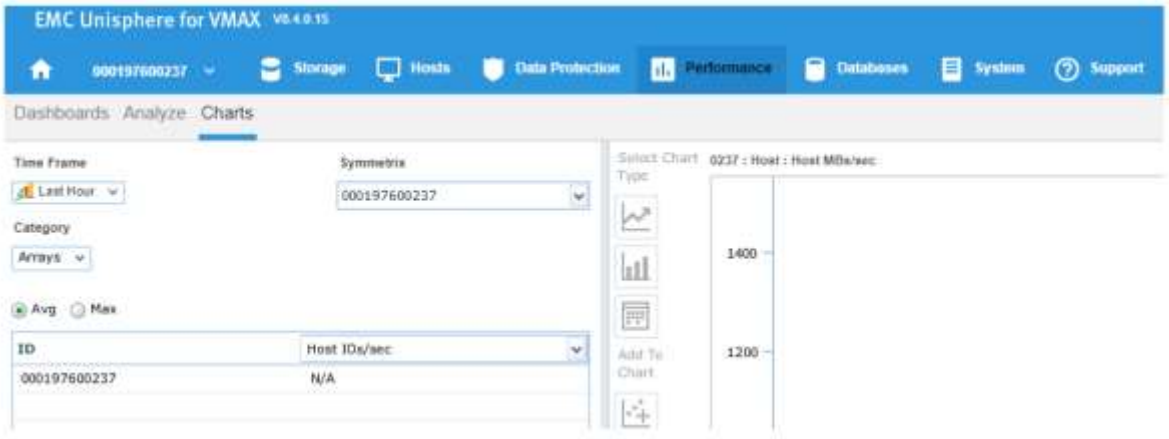
Nas próximas seções, mostraremos que você deve coletar essas estatísticas da linha de base do seu storage array para que, quando ocorrer um problema conforme declarado no cenário acima, esteja bem equipado para encontrar a causa-root.

Linha de base do aplicativo

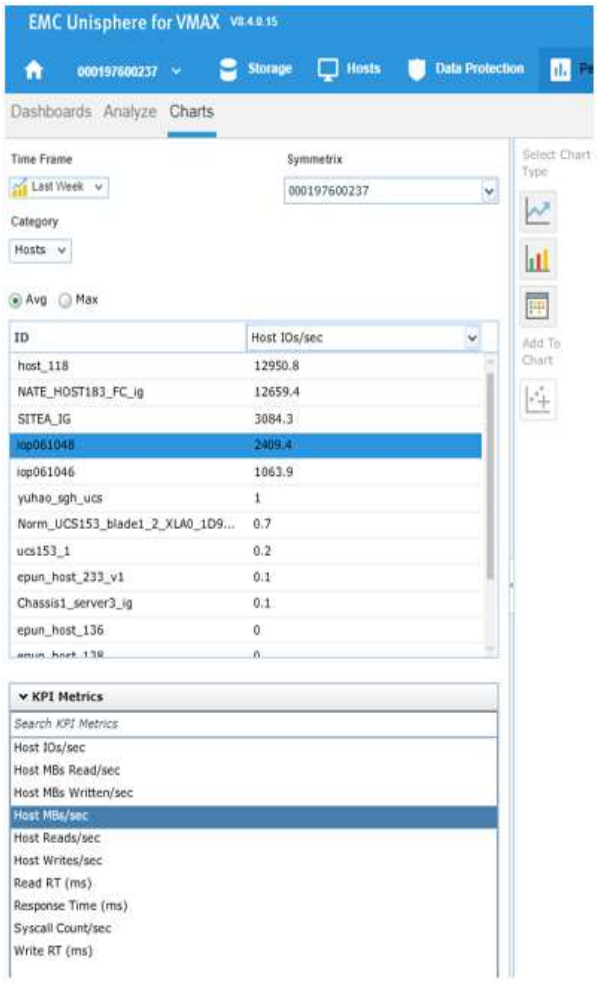
Com o Dell EMC PowerMax e VMAX, quando você habilita o monitoramento de desempenho, pode voltar ao histórico (até um ano desde a habilitação do recurso), para que possa entender o que o perfil do aplicativo tinha em termos de IOPs e tempos de resposta antes de qualquer alteração ter sido post. A existência desse perfil básico do aplicativo permitirá que você use os gráficos gerados e determine facilmente onde pode haver problemas.

Gerando gráficos de perfil de base do aplicativo

No Dell EMC Unisphere, clique em **Desempenho > Gráficos**



Selecione um **intervalo de tempo**. Esse intervalo pode ser qualquer momento ANTES de você perceber o problema de desempenho. No menu drop-down Categoria, selecione **Hosts > Hosts.**



Selecione o Host em questão. Para as **Métricas de KPI**, geraremos sete gráficos diferentes. Repita esta seção para cada KPI medição. Se você clicar em todas as medições de uma só vez, o gráfico as colocará em um único gráfico.

a. ES/s do host
b. MB/s do host
c. Leituras/s do host
d. Gravações/s do host
e. RT de leitura (ms)
f. Tempo de resposta (ms)
g. RT de gravação (ms)

Na Figura 5, “E/S e MB/s do host”, estamos analisando as E/S os MB/s do host. Nestes gráficos, podemos ver quando e por quanto tempo o aplicativo executou a maioria das E/Ss e utilizou a largura de banda completa do link, bem como os pontos mais baixos.

Obs: na legenda, você perceberá que existem dois hosts, mas, atualmente, estamos mostrando apenas a E/S de um deles porque o outro não está gerando nenhuma E/S.

Figura 5 E/S e MB/s do host

Os gráficos mostrados na Figura 6, “Leituras e gravações/s”, fornecem uma análise do tipo de E/S que o aplicativo gera. Com base nesses gráficos, podemos determinar qual porcentagem de E/S do aplicativo é leitura ou gravação. Nesse caso, podemos confirmar que o aplicativo compõe cerca de 70/30 em termos de leituras/gravações.

Figura 6 Leituras e gravações/s

A Figura 7 e a Figura 8 provavelmente são os gráficos mais relevantes a serem usados para solucionar problemas. Eles oferecem uma análise dos tempos de resposta entre as leituras e as gravações, que nos permitem entender a latência que o aplicativo está experimentando. Isso é extremamente útil quando precisamos solucionar problemas de desempenho; já que, se houver um aumento nos tempos de resposta, podemos correlacionar o pico aos eventos específicos usando os gráficos anteriores.

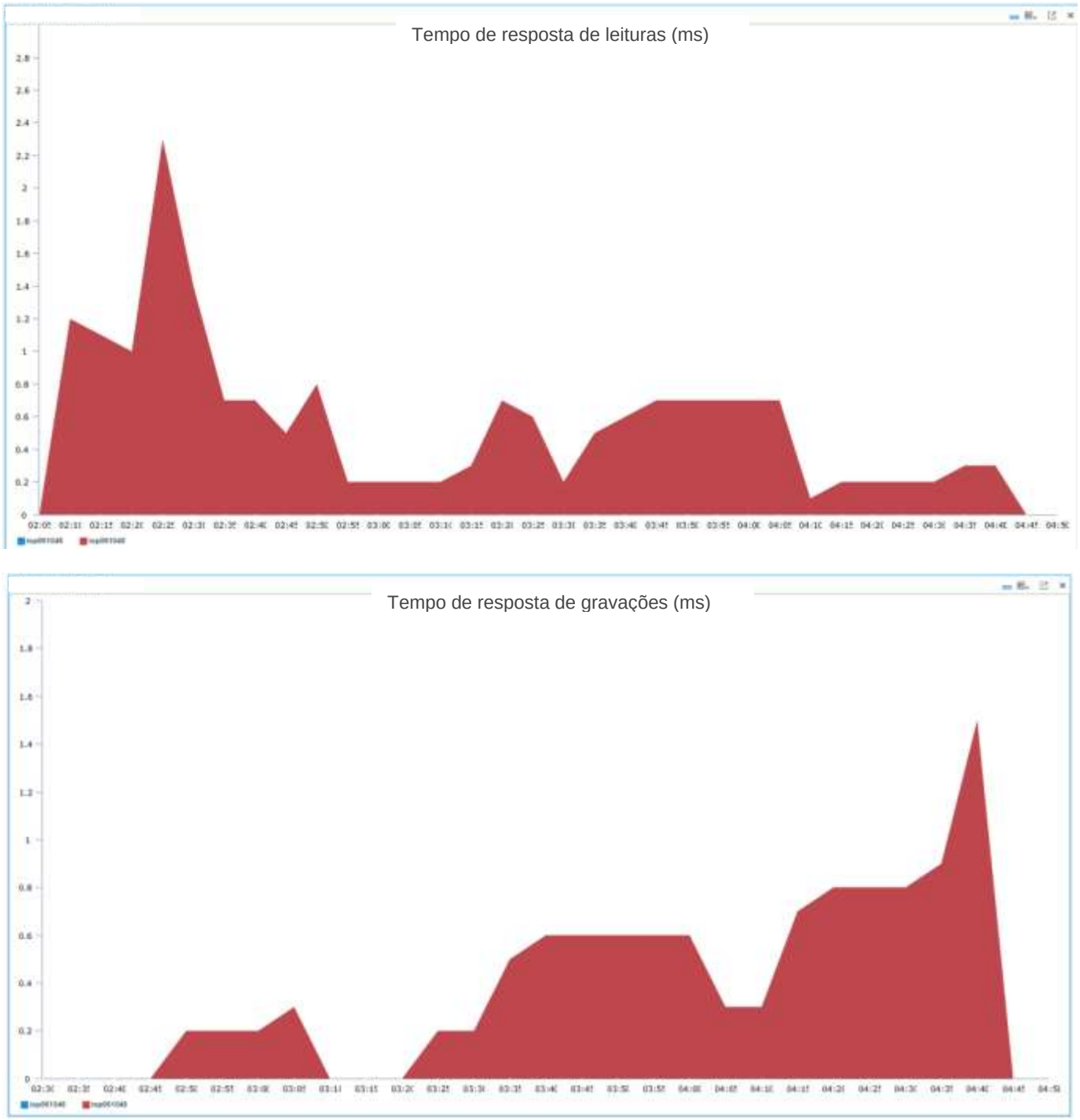


Figura 7 RT de leituras e gravações (ms)

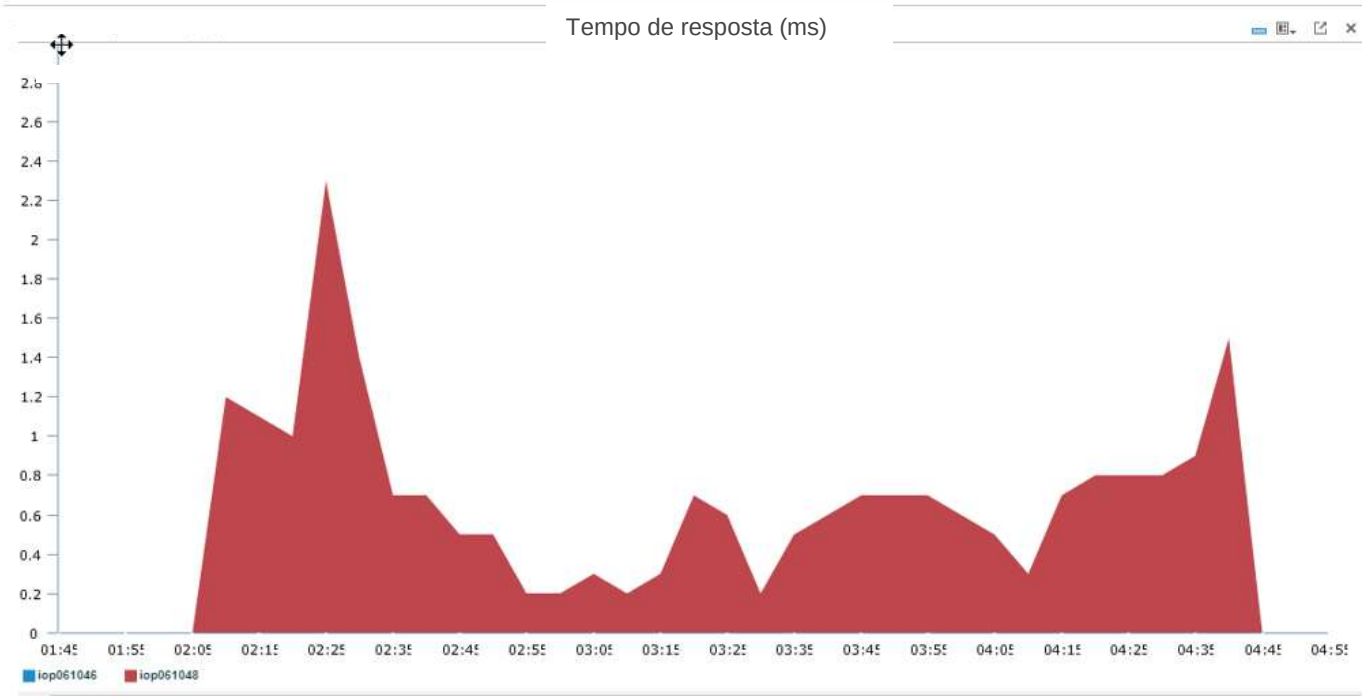


Figura 8 Tempos de resposta (ms)

Agora que temos um perfil de aplicativo do Usuário 1, sabemos que são geradas cerca de 70/30 leituras/gravações com tempos de resposta de, aproximadamente, 0,7 ms, com tempo máximo de 2,3 ms.

Conforme discutido na seção do cenário, recentemente adicionamos um novo host que iniciava um problema de desempenho no ambiente, então vamos analisar como podemos solucionar esse problema.

Como estamos tendo um problema de desempenho em nosso ambiente, precisaremos implementar os recursos disponíveis na SAN que podem nos ajudar a determinar quando esses tipos de problemas surgem.

Como sabemos quais são os tempos de resposta médios, com base no perfil do aplicativo, sabemos que esse problema de desempenho é maior do que os tempos de resposta esperados.

Alertas de propagação do congestionamento da SAN do Connectrix

Nesta seção, vamos analisar o tipo dos eventos de congestionamento que relatamos no comutador da SAN. Certifique-se de que você concluiu a ativação desses recursos no ambiente de acordo com os Pré-requisitos.

4.1.1 Brocade

1. Certifique-se de ter pelo menos o **Tráfego de porta principal** e o **Crédito de buffer-para-buffer igual a zero** em seu painel de indicadores. Você pode clicar na chave inglesa no canto superior esquerdo para adicioná-los caso não os tenha.

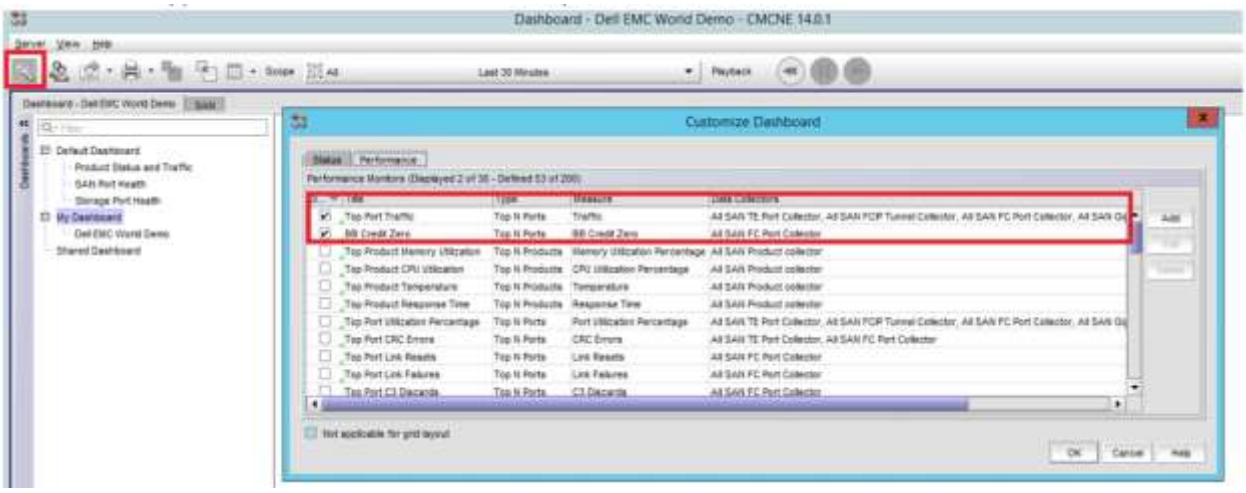


Figura 9 Painel de indicadores do CMCNE

2. Quando ocorre a propagação do congestionamento devido à superatribuição, como no exemplo da Figura 9, “Painel de indicadores do CMCNE”, os seguintes alertas no painel de indicadores do CMCNE são exibidos:
    a. Porta F altamente utilizada
    b. Crédito buffer-para-buffer igual a zero

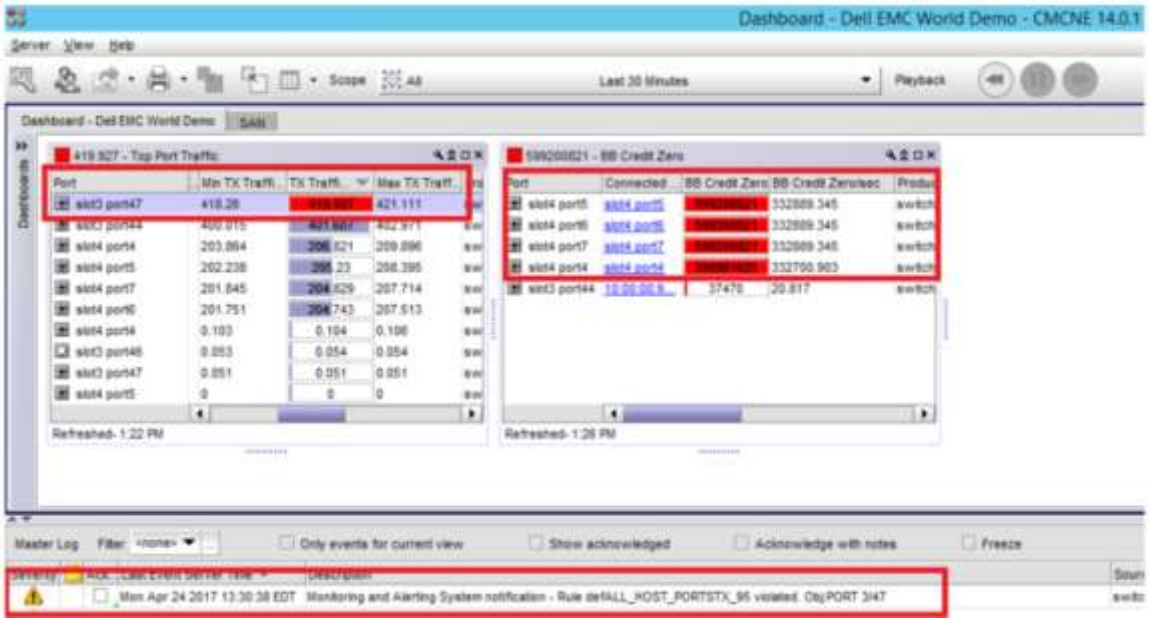
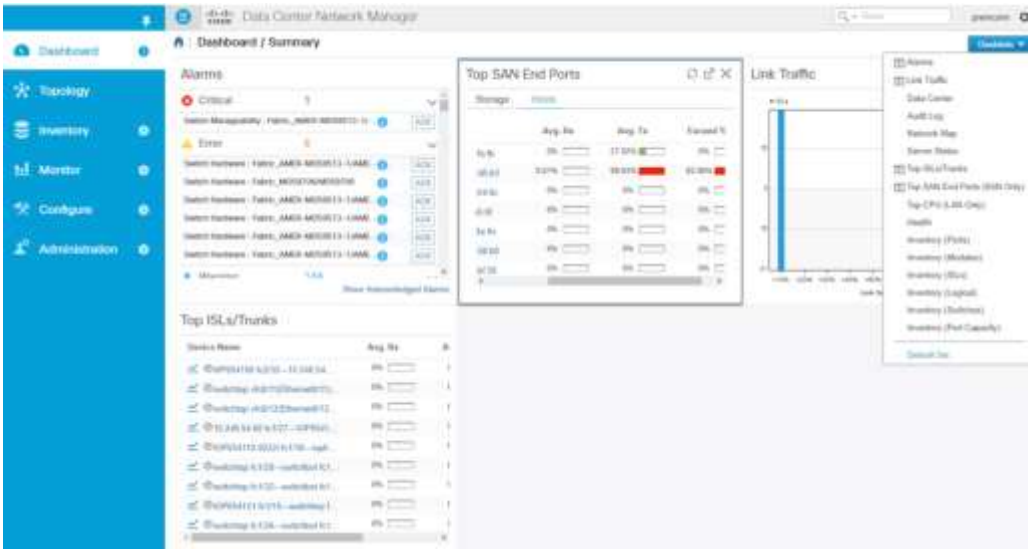


Figura 10 Painel de indicadores do CMCNE exibindo alertas

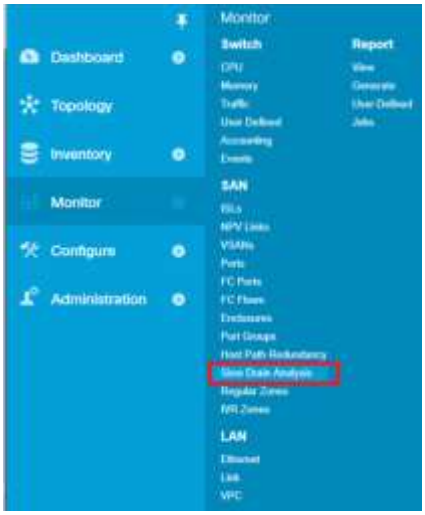
3. A combinação desses dois eventos — a porta F altamente utilizada e os créditos de buffer-para-buffer iguais a zero nos ISLs — podem indicar que você tem um possível problema de desempenho que precisa ser investigado. Consulte o capítulo Correção para obter as etapas sobre o que analisar.

4.1.2 Cisco

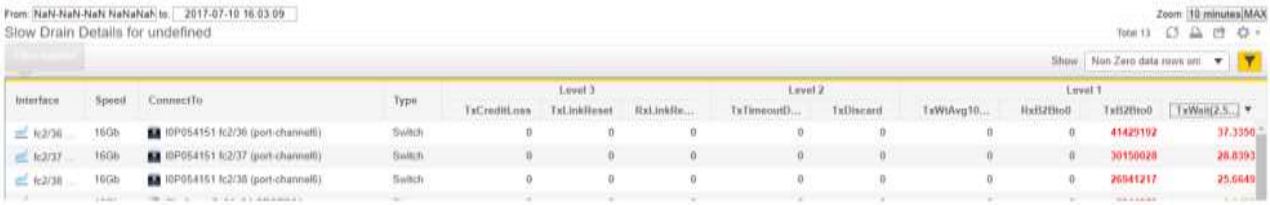
1. No painel de controle do DCNM, as **principais portas da extremidade da SAN** devem ser exibidas como uma dashlet. Caso contrário, você pode adicioná-lo no menu drop-down. Nas **principais portas de extremidade da SAN,** você verá que um ou mais relatórios de dispositivos têm mais de 90% de utilização. O DCNM terá os limites padrão que fazem com que ele sinalize uma porta amarela ou vermelha quando começar a exceder a utilização padrão. Esse alerta por si só não significa necessariamente que há um problema de desempenho na SAN. Também será necessário procurar outros alertas em fabrics.



2. Se você vir uma porta F altamente utilizada, execute a ferramenta de análise de drenagem lenta. Clique em **Monitor > SAN > Slow Drain Analysis.**



3. Execute a ferramenta de análise de drenagem lenta por 10 minutos. Quando o relatório for concluído, você perceberá que há uma grande quantidade do contador do TxWait sendo incrementado durante o tempo em que o relatório estava em execução. A combinação desses alertas e a porta F altamente utilizada indica que há congestionamento na SAN ocorrendo devido a superatribuição.



A combinação desses dois eventos — a porta F altamente utilizada e os créditos de buffer-para-buffer iguais a zero nos ISLs — podem indicar que você tem um possível problema de desempenho que precisa ser investigado. Consulte o capítulo Correção: para obter as etapas sobre o que analisar.

ALERTAS DE PROPAGAÇÃO DE CONGESTIONAMENTO DO UNISPHERE

Nesta seção, falaremos sobre como usar o Unisphere para PowerMAX e VMAX para correlacionar os eventos do comutador SAN ao storage array.

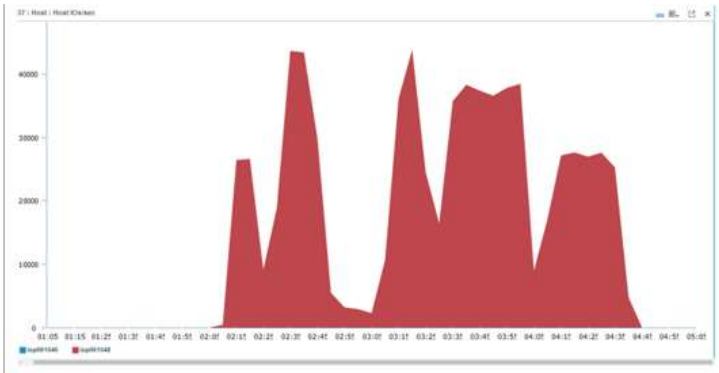
Certifique-se de ter concluído a ativação desses recursos no ambiente de acordo com a seção pré-requisitos.

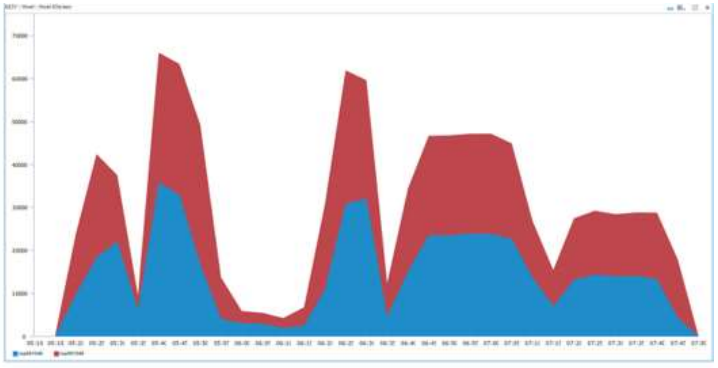
1. Use as etapas da Gerando gráficos de perfil de base do aplicativo. Gere os mesmos sete gráficos, adicionando o host do Usuário 2 à combinação (porque essa era uma das alterações recentes no ambiente antes de o problema de desempenho ter ocorrido). Analise os dados.

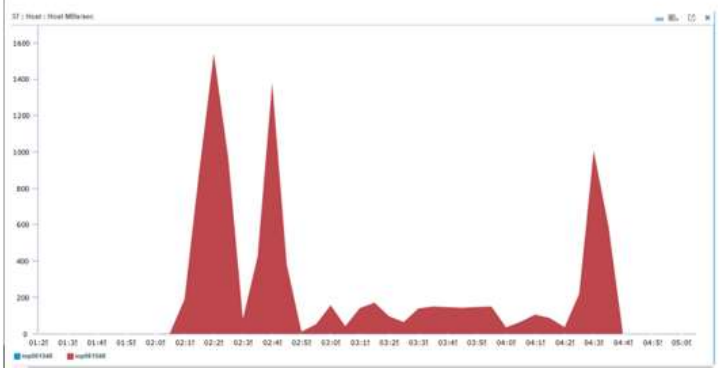
Lembre-se de que nosso perfil de base do aplicativo é igual a 70/30 leituras/gravações com tempos médios de resposta de aproximadamente 0,7 ms e tempo máximo de resposta de 2,3 ms.

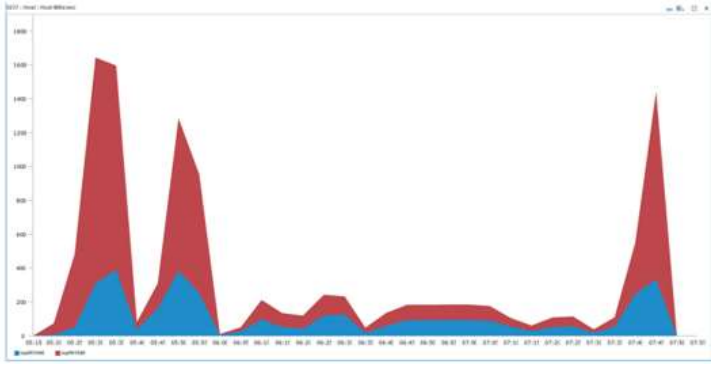
Na Figura 11 abaixo, quando comparamos as E/S/s e os MB/s, não vemos uma indicação de um problema. Na verdade, se você compará-lo ao gráfico do aplicativo da linha de base original, estamos adotando mais IOPs.

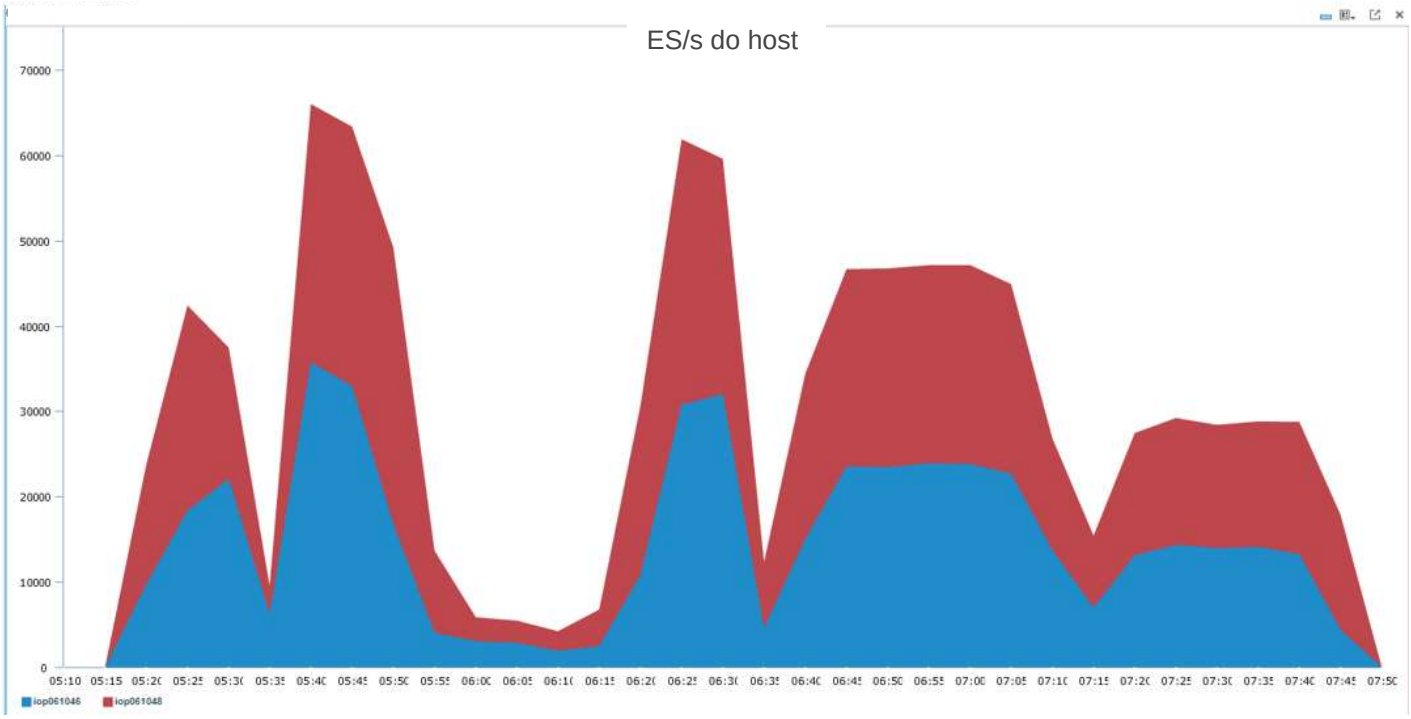
Além disso, você pode ver que há alguns pontos em que estamos chegando perto da transmissão de dados (destacada abaixo). Esses pontos entrarão em cena posteriormente.










ES/s do host
70000
60000
50000
40000
30000
20000
10000
0
iop061046
iop061048

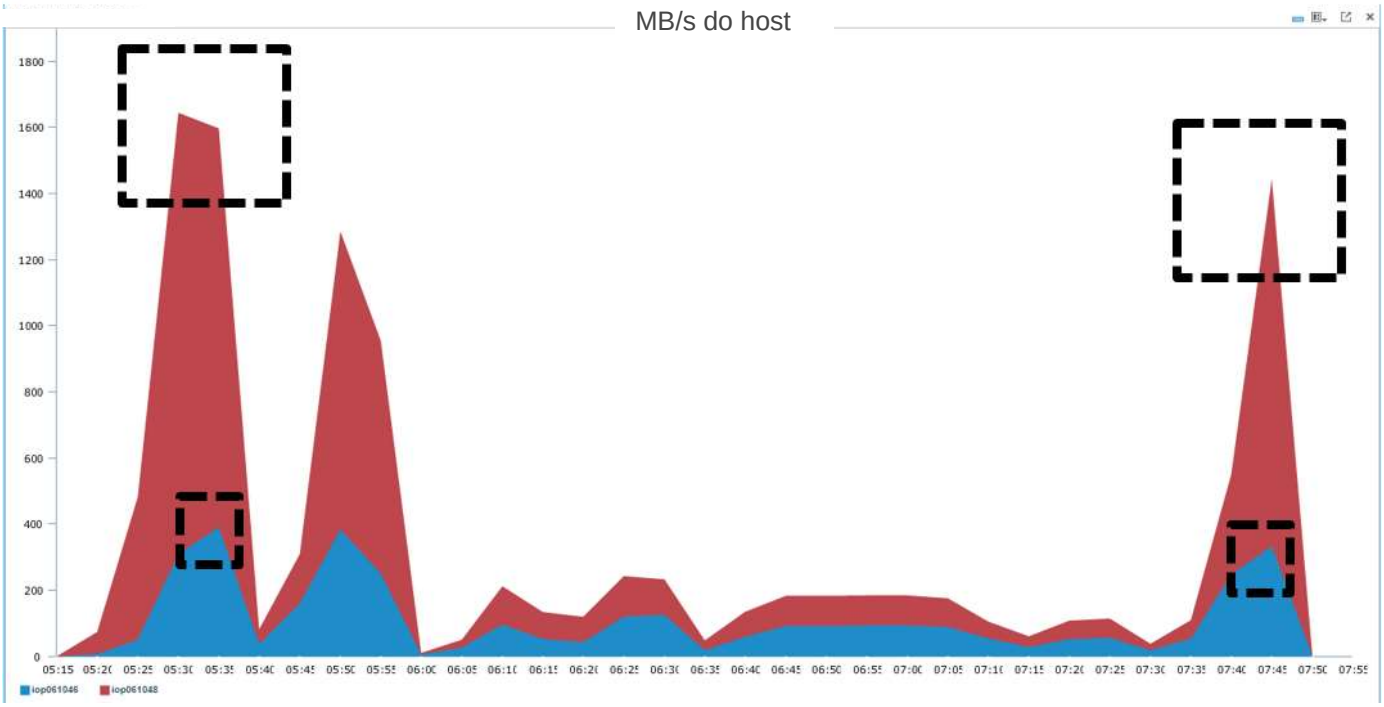
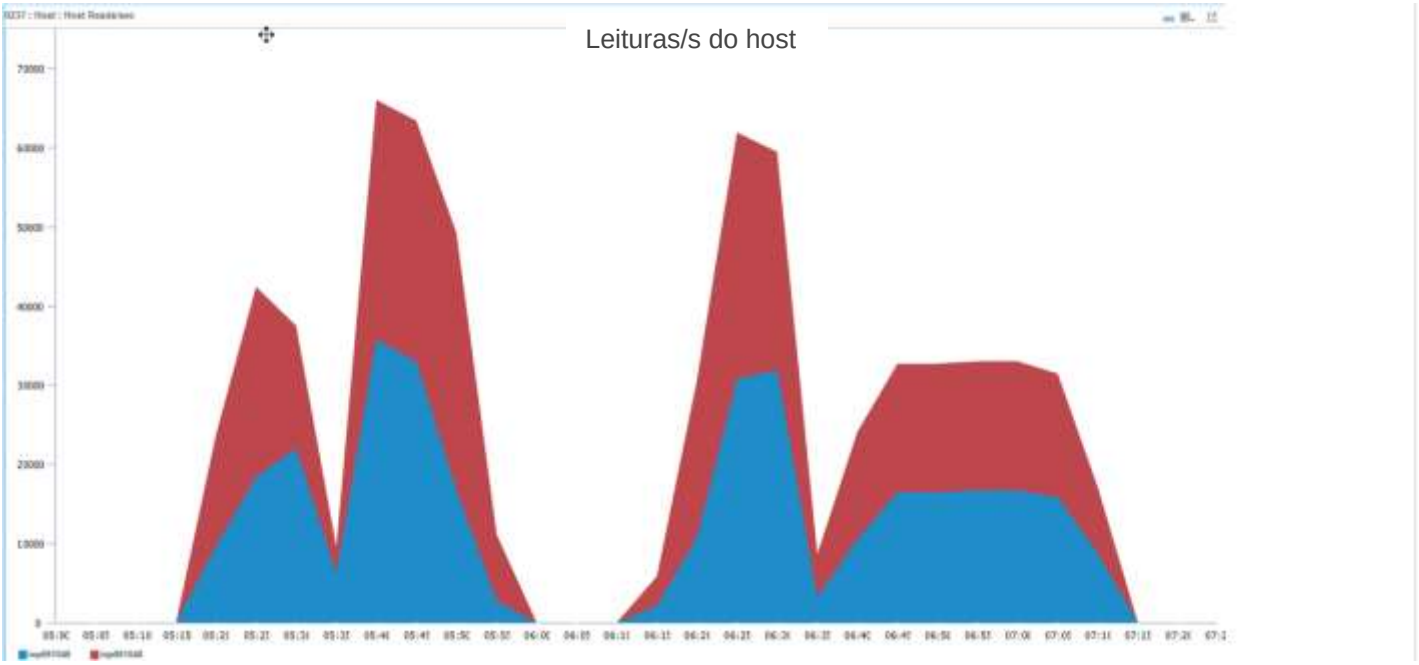


Figura 11 E/S e MB/s do host

A *Figura 12* mostra a comparação em leituras/gravações entre os dois servidores, pois você pode ver que não há muita diferença entre o perfil de E/S entre os eles neste ponto.



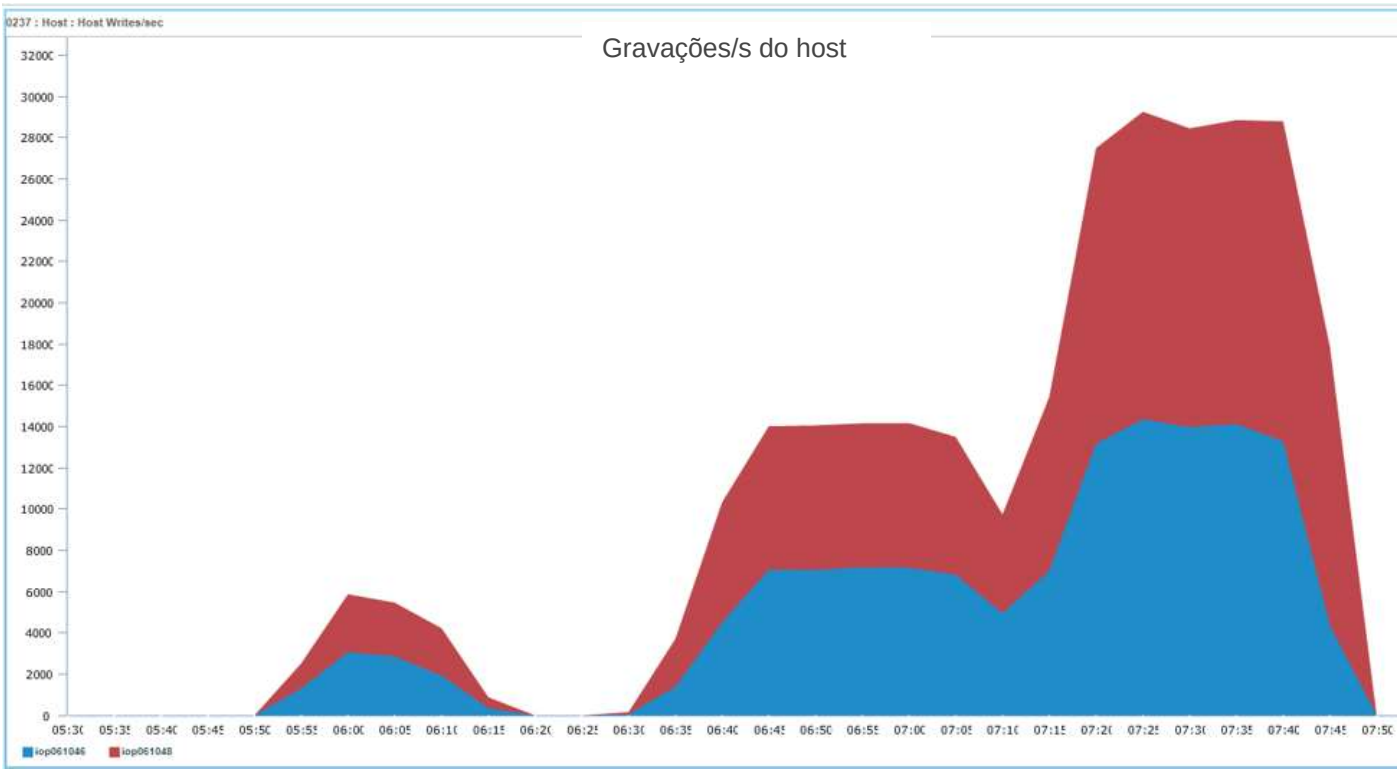
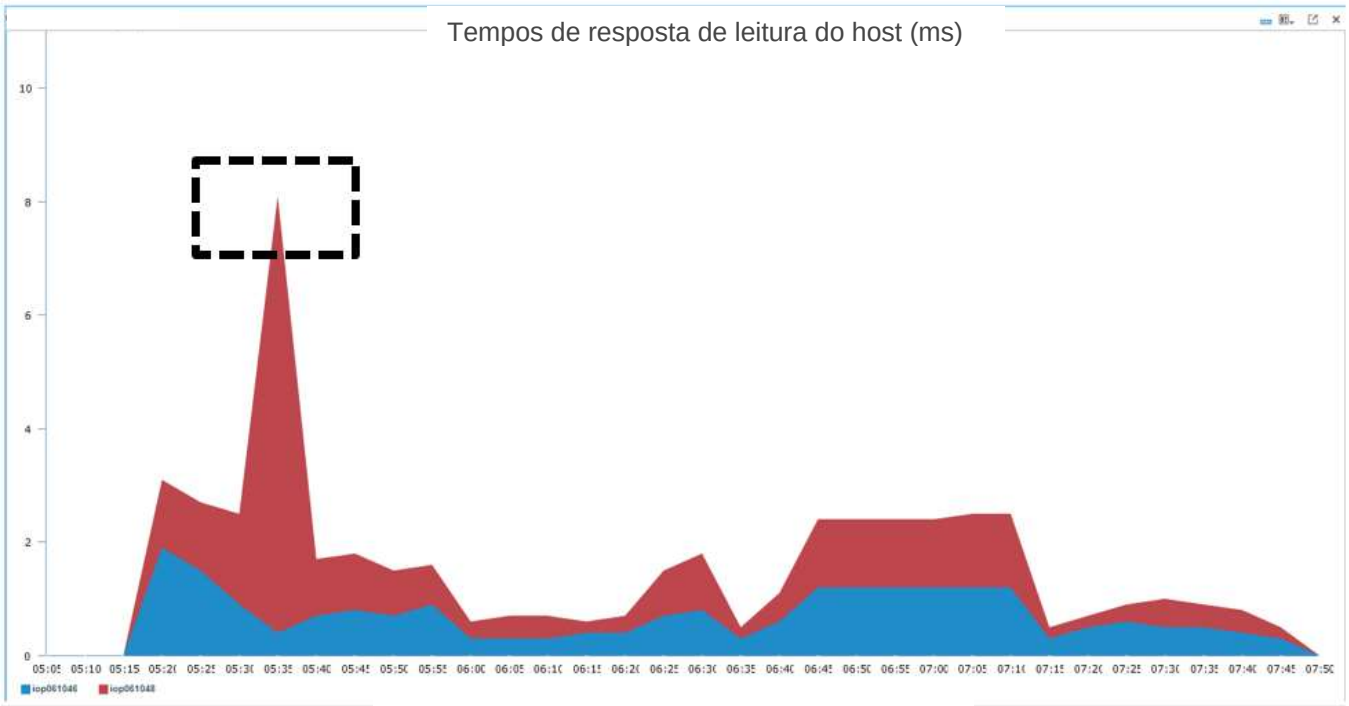


Figura 12 Leituras e gravações/s do host

A Figura 13 apresenta as informações mais relevantes. Ao manter nosso perfil de aplicativos em mente, notamos tempos de resposta médios de aproximadamente 0,7 ms e máx. de 2,3 ms. No gráfico a seguir, podemos ver que há um enorme aumento nos tempos de resposta, em que entramos no intervalo de 8 ms, e que nossos tempos de resposta médios gerais também aumentaram.

Analisando a Figura 11, podemos ver que os altos tempos de resposta se correlacionam de volta para quando os dois servidores estavam próximos da transmissão de dados.
Normalmente, em Fibre Channel, você precisará de E/S de blocks grandes (mais de 128 KB) para saturar um link.



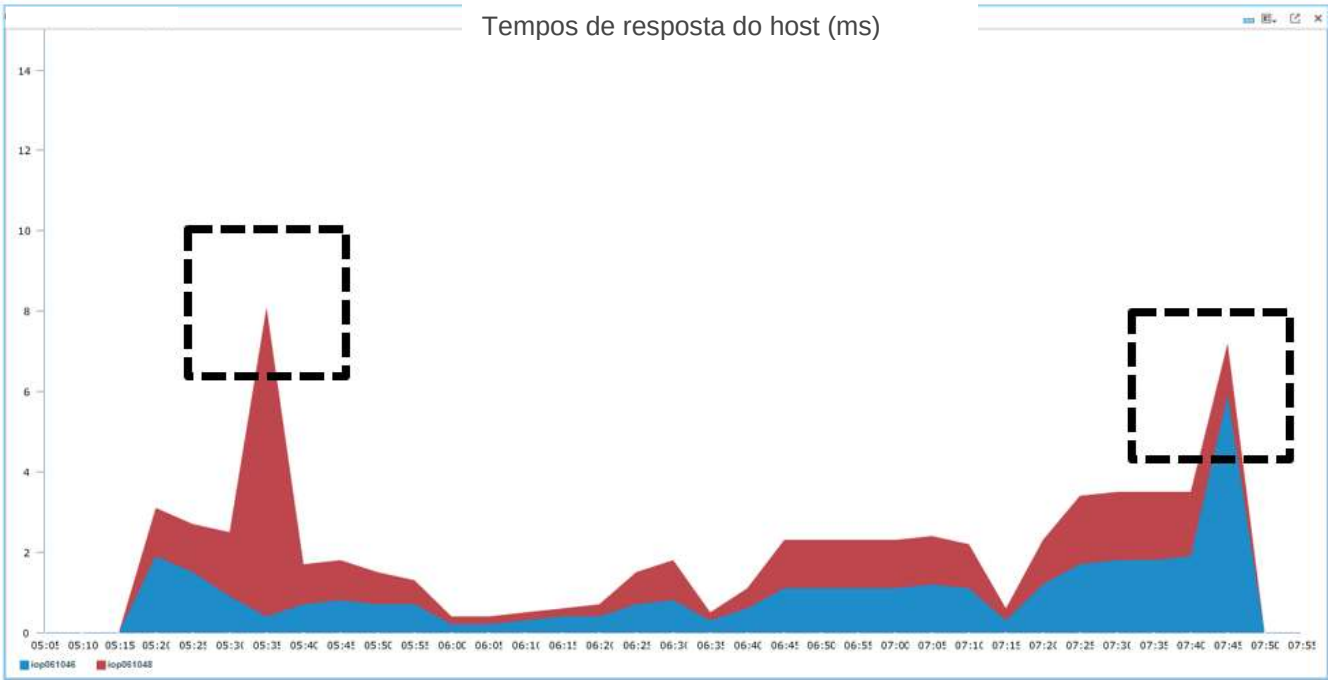


Figura 13 Tempos de resposta de leitura e gravação do host (ms)

CONCLUSÃO

Para recapitular todas as informações que sabemos até o momento neste estudo de caso:

- SAN do Connectrix:

1. Nossa SAN está relatando um grande número de créditos de buffer para buffer iguais a zero.
2. Estamos vendo a alta utilização de tráfego em nossa(s) porta(s) F.

- Dell EMC VMAX e PowerMax

1. Altos tempos de resposta durante a utilização completa do link

Conforme mencionado anteriormente, o congestionamento devido à disparidade da largura de banda é extremamente difícil de detectar e confirmar com o conjunto de ferramentas disponível atualmente. No entanto, com base nos alertas acima, podemos deduzir que o problema ocorre devido à disparidade da largura de banda e de leituras/gravações de blocks grandes. Isso é indicado pelos altos tempos de resposta durante a utilização completa do link.

Outra maneira de detectar esse problema é usar a Proporção de congestionamento. Hoje, calcularemos isso manualmente no ambiente (você também pode tentar criar o script dele), mas sabemos que, assim que a taxa de C for maior que 0,2, você enfrentará congestionamento por causa da contrapressão que está ocorrendo no ambiente SAN. A taxa C seria a primeira indicação de uma drenagem lenta.

5 Correção:

PREVENÇÃO

Para este estudo de caso específico (propagação do congestionamento devido à superatribuição), existem algumas opções que você pode implementar em seu ambiente para ajudar a evitar que esse problema ocorra.

Taxa de largura de banda

- Ao analisar a SAN, é recomendável identificar os dispositivos que estão sendo executados em velocidades mais baixas e, em seguida, entender o tipo de perfil de tráfego de aplicativos deles. Lembre-se de que, apenas porque você tem uma disparidade de largura de banda, NÃO significa necessariamente que há um problema.
- Analise o fabric de ponta a ponta para garantir que todos os dispositivos finais estejam em execução nas mesmas velocidades do link.
- Garanta uma grande quantidade de largura de banda em seus ISLs. Uma boa regra geral é que a largura de banda total do ISL deve ser igual ou maior do que a quantidade total de largura de banda de armazenamento no fabric sempre que possível.
- Você pode modernizar toda a SAN, garantindo que você faça upgrade completo de todos os componentes, conforme mostrado na Figura 14 abaixo. A vantagem dessa abordagem é que a superatribuição completa igual a zero é impraticável em ambientes maiores. Além disso, ela pode ser muito dispendiosa. Portanto, você deve se concentrar apenas no upgrade do host, comutador e conectividade de armazenamento específicos.

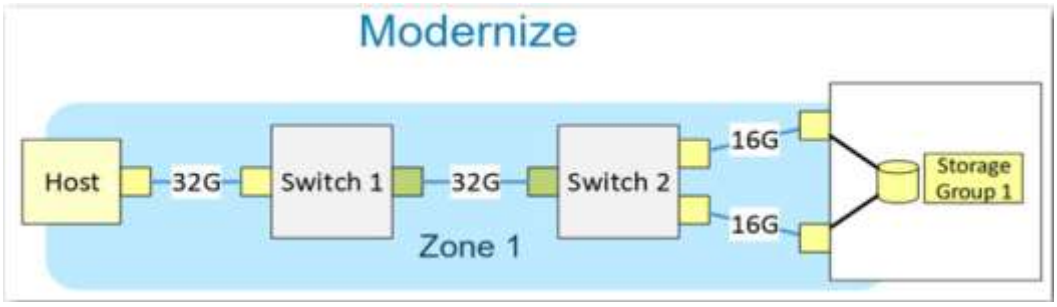


Figura 14 Modernização

Correção:

Outra maneira seria reposicionar a zona como mostrado na Figura 15.

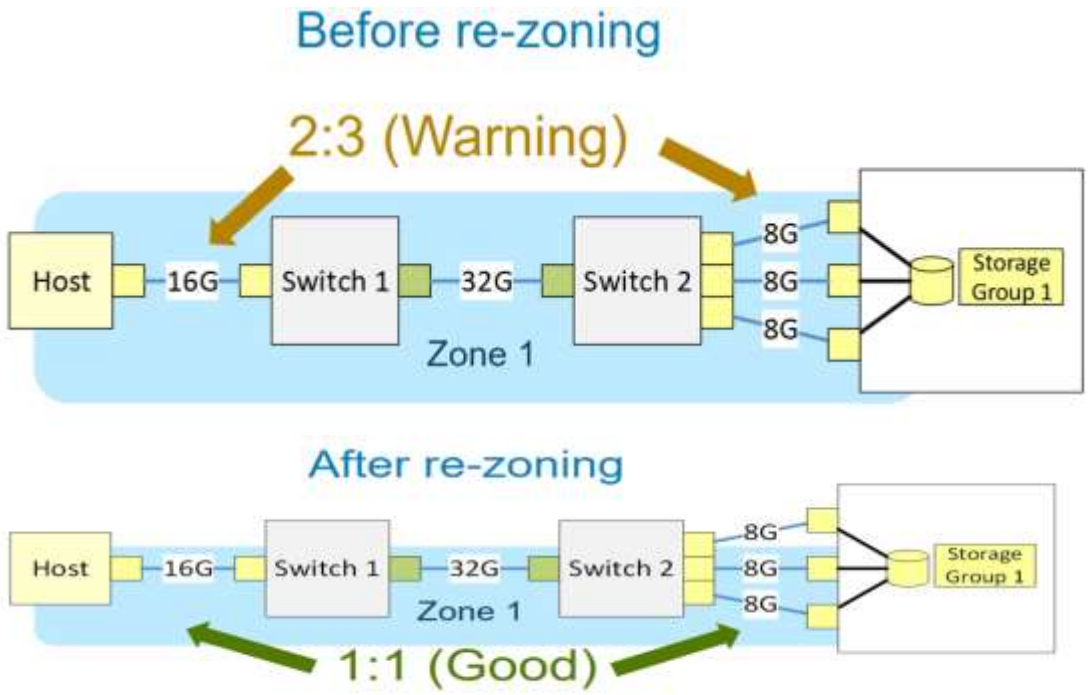


Figura 15 Antes e depois do reposicionamento de zonas

Implementar limites de largura de banda

As plataformas Dell EMC VMAX e Dell EMC Unity criam limites de largura de banda nos Storage Groups (VMAX) ou nos LUNs (Unity). No estudo de caso mencionado acima, em que houve a propagação do congestionamento devido à superatribuição, quando implementamos limites de largura de banda, vimos que o desempenho foi restaurado, conforme mostrado na Figura 16 abaixo. Isso pode ser feito diretamente por meio do Unisphere no Storage Group.

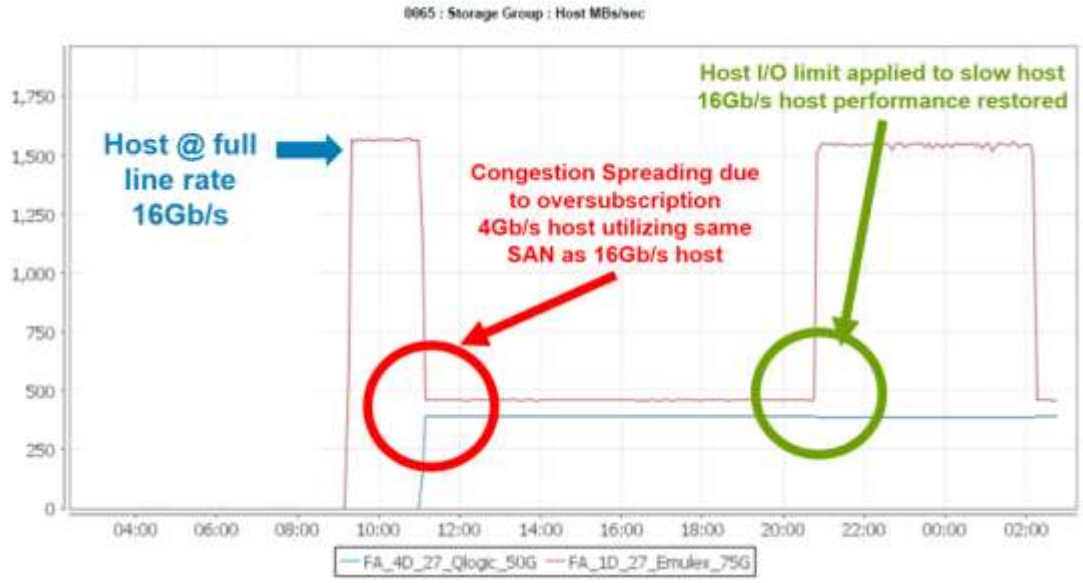


Figura 16 Limites de E/S de host aplicados

Correção:

É importante observar que os limites de E/S não funcionam bem com clusters. Vamos analisar como exemplo a Figura 17 abaixo. Quando o limite de host é aplicado a um host de 4 GB que está causando a contrapressão, o array começa a limitar o volume de dados que ele envia de volta para o 4 GB (com base no limite definido de E/S); portanto, você elimina o problema de contrapressão, e outros fluxos podem operar com transmissão de dados completa.

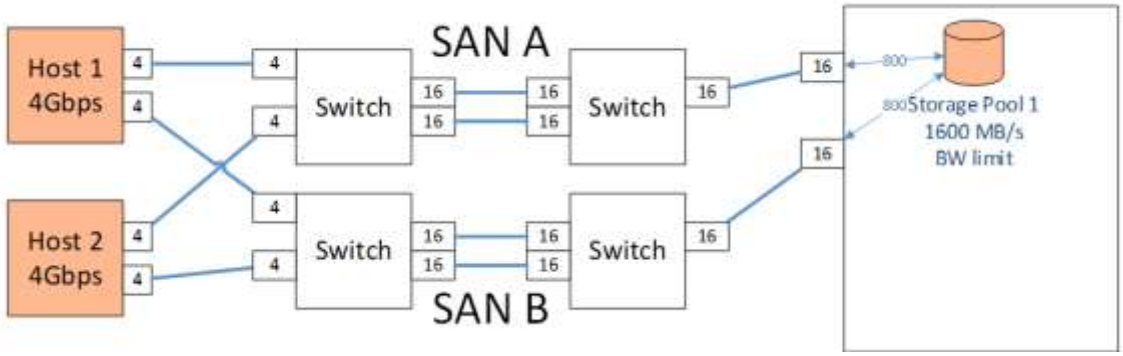


Figura 17 Limites de E/S com clusters

Neste exemplo, temos dois hosts em execução a 4 GB/s localizados em um cluster. Como eles estão em um cluster, os dois terão acesso ao volume por meio de cada fabric. Isso significa que precisamos definir um limite de E/S de largura de banda de 1.600 MB/s (800 MB/s para cada FA). No entanto, com essa abordagem, nada impede que um HBA único consuma todos os 800 MB/s.

- Isolamento

Outra maneira de evitar esse problema é isolar o tráfego mais lento do tráfego de alta velocidade e usar ISLs dedicados. Isso pode ser feito por meio da criação de fabrics virtuais (Brocade) ou VSANs (Cisco) como indicado na Figura 18, abaixo. A vantagem dessa abordagem é que você precisa de portas dedicadas, mas isso evita que seu tráfego mais lento afete o tráfego de maior velocidade. Viabilizar fabrics virtuais na Brocade resultaria em tempo de inatividade, já que o comutador inteiro precisaria ser reinicializado. Ao mover uma porta para um VSAN diferente na Cisco, apenas os dispositivos finais que estão sendo movidos serão afetados.

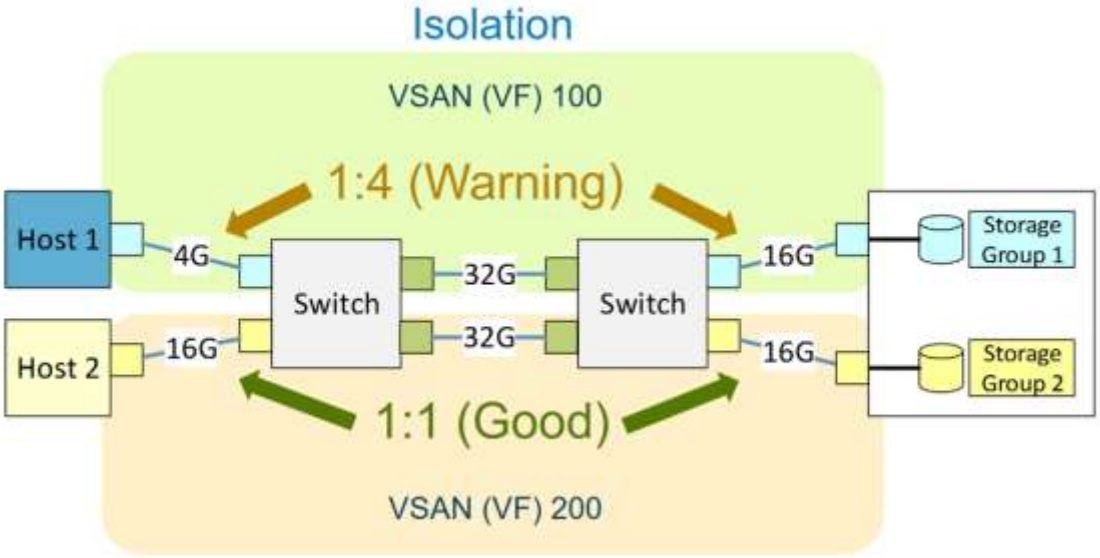


Figura 18 Isolamento

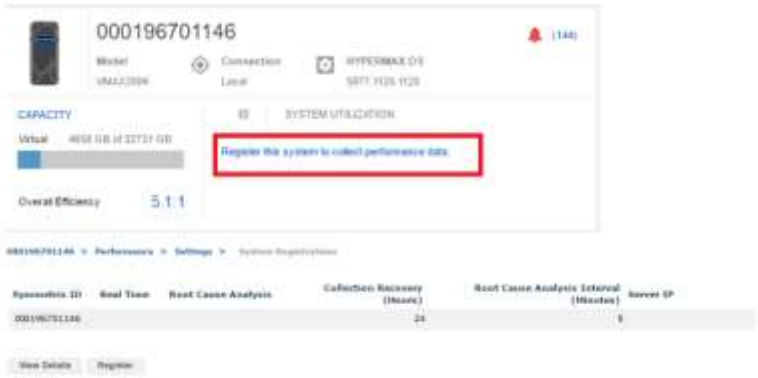
6 Apêndice

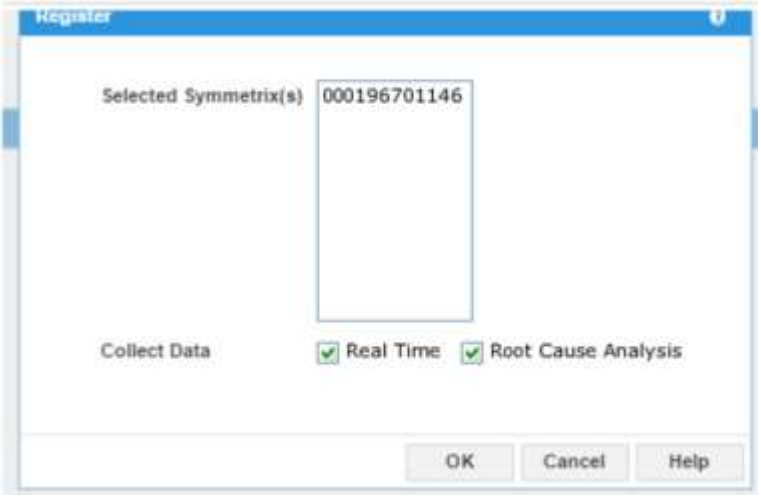
HABILITAR MONITORAMENTO DE DESEMPENHO

Esta seção apresenta etapas sobre como habilitar e analisar o monitoramento de desempenho de dados no Unisphere for VMAX.

1. Faça log-in na GUI do Unisphere.

2. Certifique-se de que o array esteja registrado para fins de coleta de dados de desempenho. Caso contrário, registre-o.




Register
Selected Symmetrix(s)
000196701146
Collect Data
Real Time
Root Cause Analysis
OK
Cancel
Help

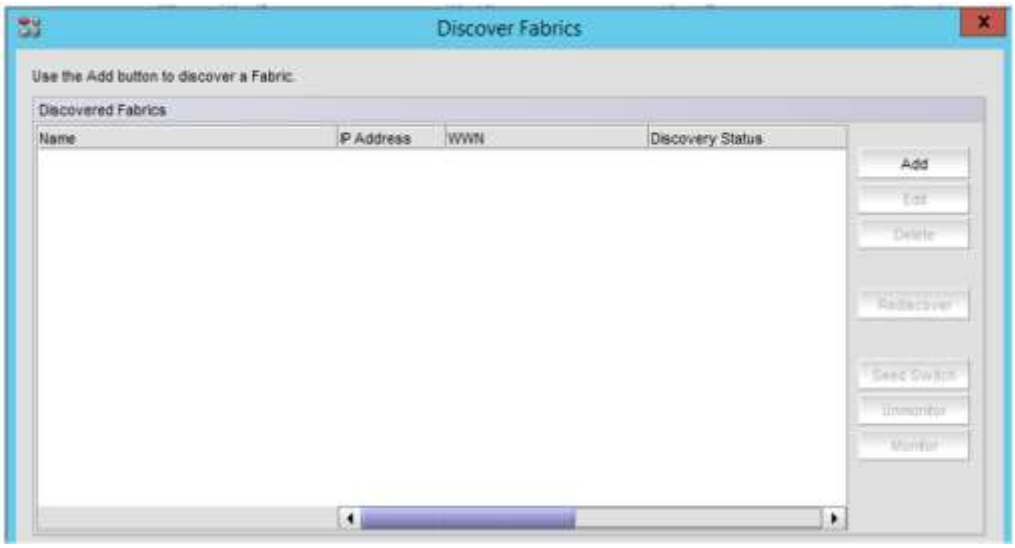
MONITORAMENTO DE PROPAGAÇÃO DE CONGESTIONAMENTO DO CONNECTRIX

6.1.1 Brocade

- Identifique o fabric

1 Faça log-in no servidor do CMCNE e clique em **Discover > Fabrics (SANnav?)**

2 Na nova janela, clique em **Add**.

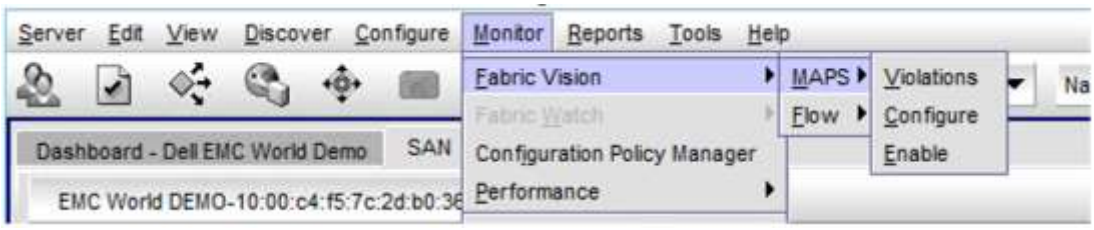


3 Preencha as informações necessárias para um dos comutadores do fabric. O CMCNE detectará automaticamente todos os comutadores nesse fabric, pressupondo que o nome de usuário e a senha sejam os mesmos para todos os comutadores do fabric.

4 Repita esta seção para todos os outros fabrics.

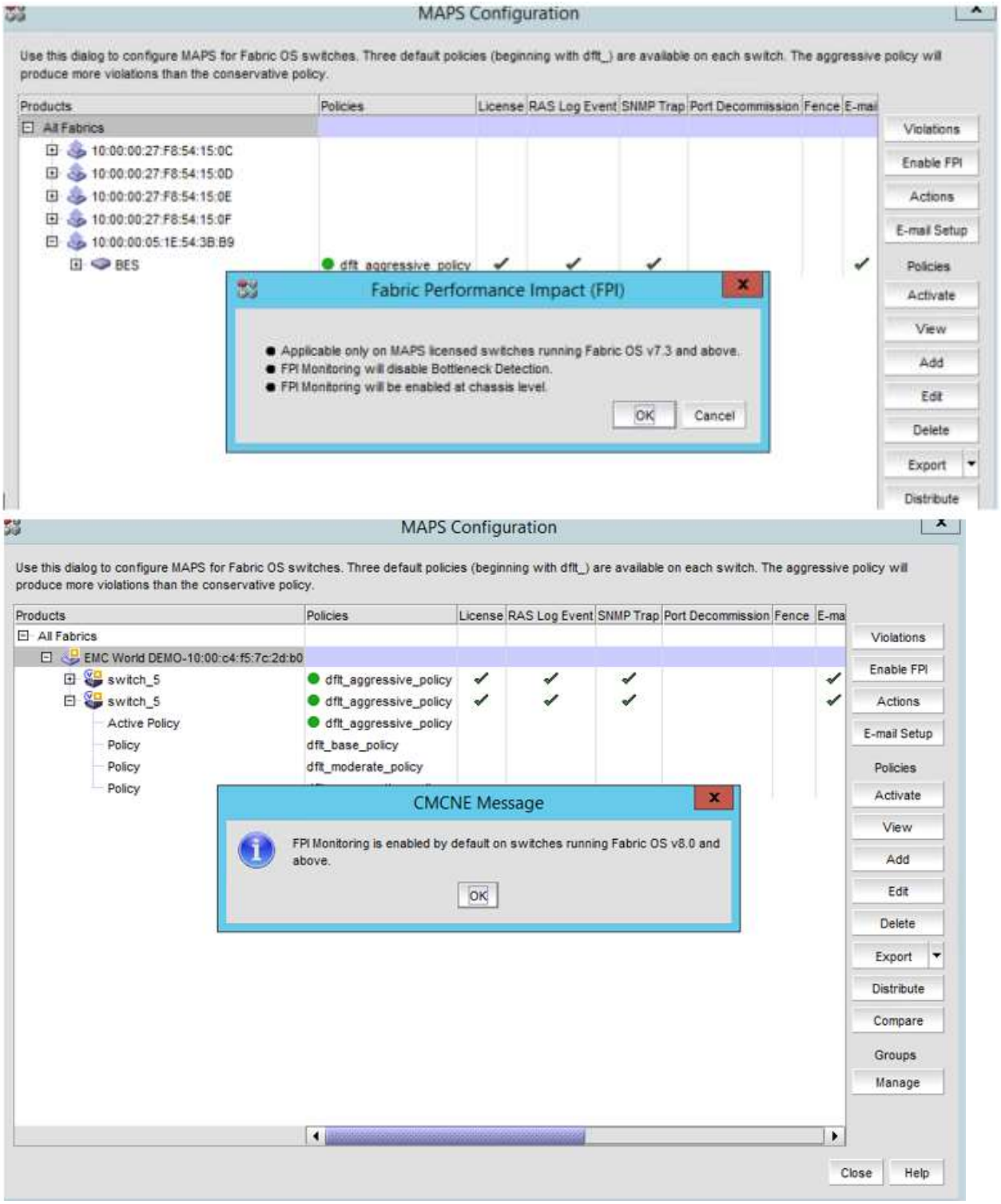
- Habilitar MAPS e FPI

1 Clique em Monitor > Fabric Vision > MAPS > Configure



2. Destaque o fabric e habilite o FPI.

Obs.: O FPI é habilitado por padrão nos comutadores que executam o FOS 8.0 e superior.

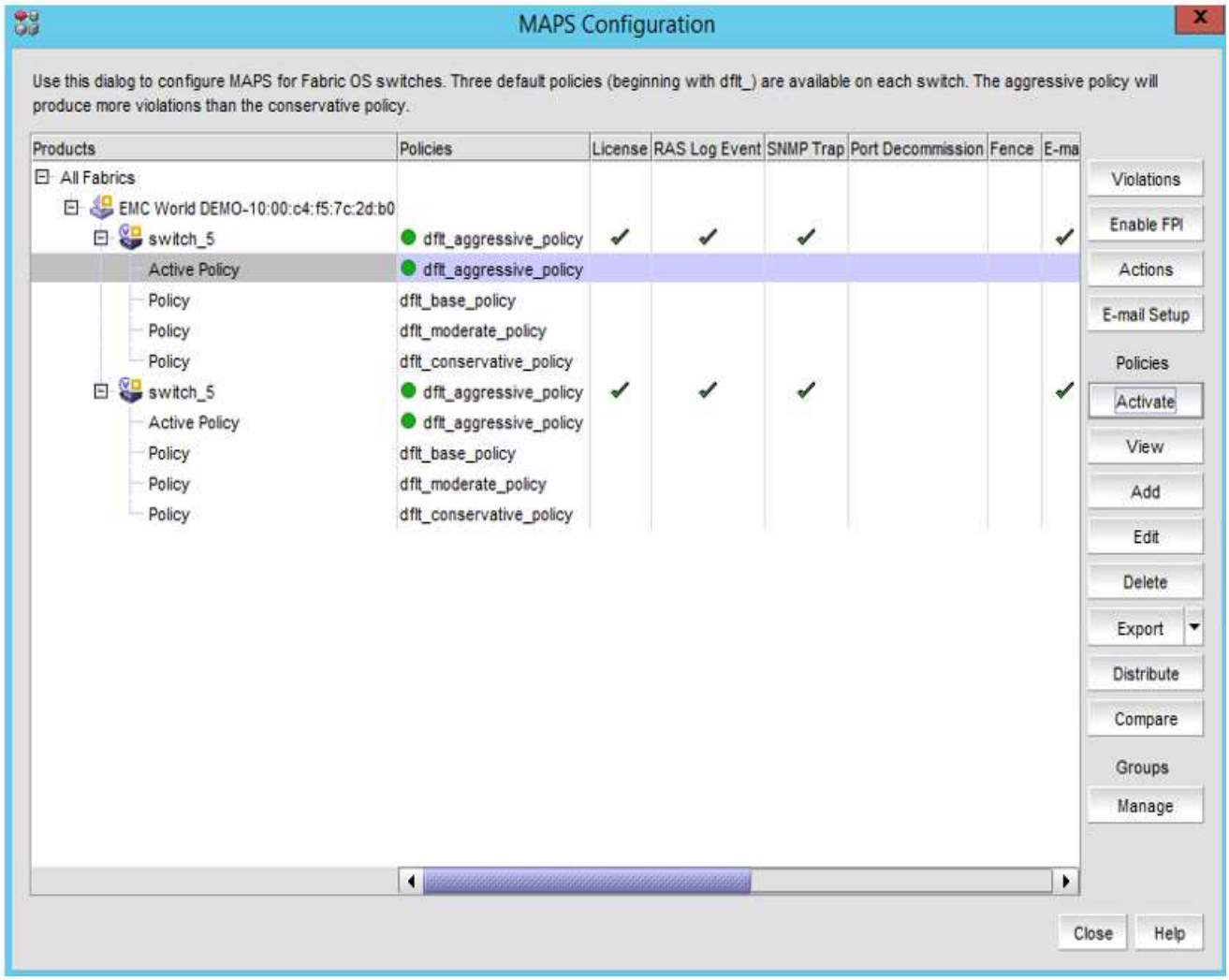


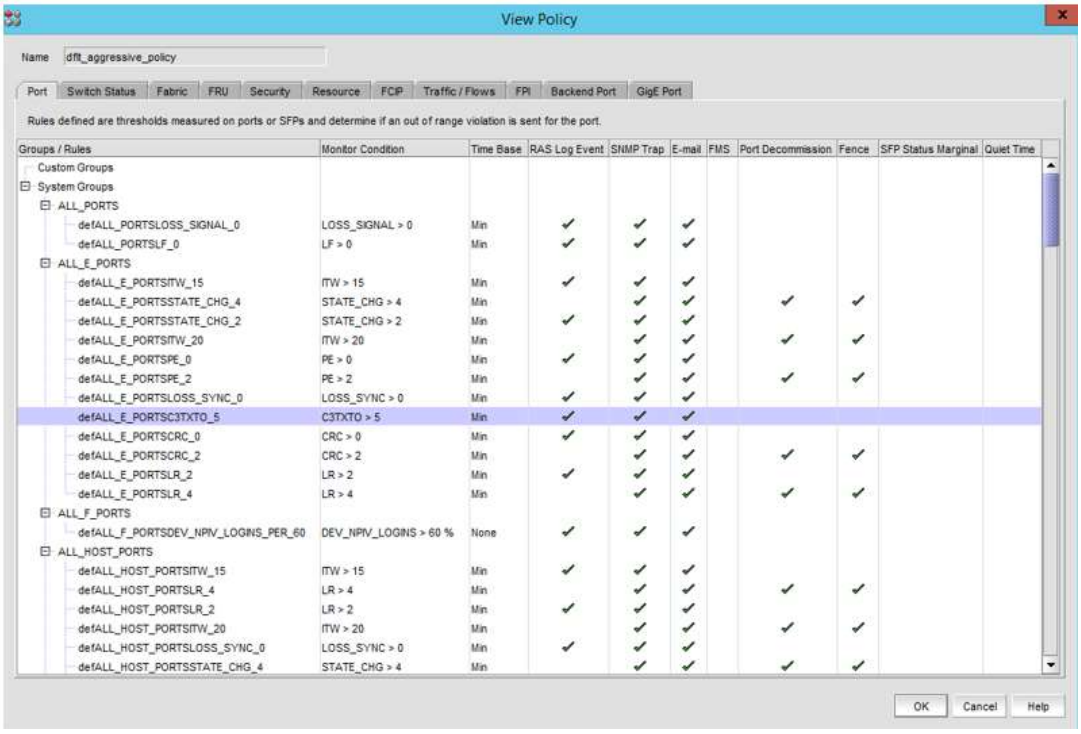
3. Nesse menu, você pode configurar cada comutador em seu fabric e definir a política do MAPS que deseja.

Obs.: O CMCNE oferece políticas predefinidas que você pode clonar e, em seguida, editar. Não é possível editar as políticas padrão. Consulte o guia de administração do MAPS para obter mais detalhes sobre cada política e as configurações.

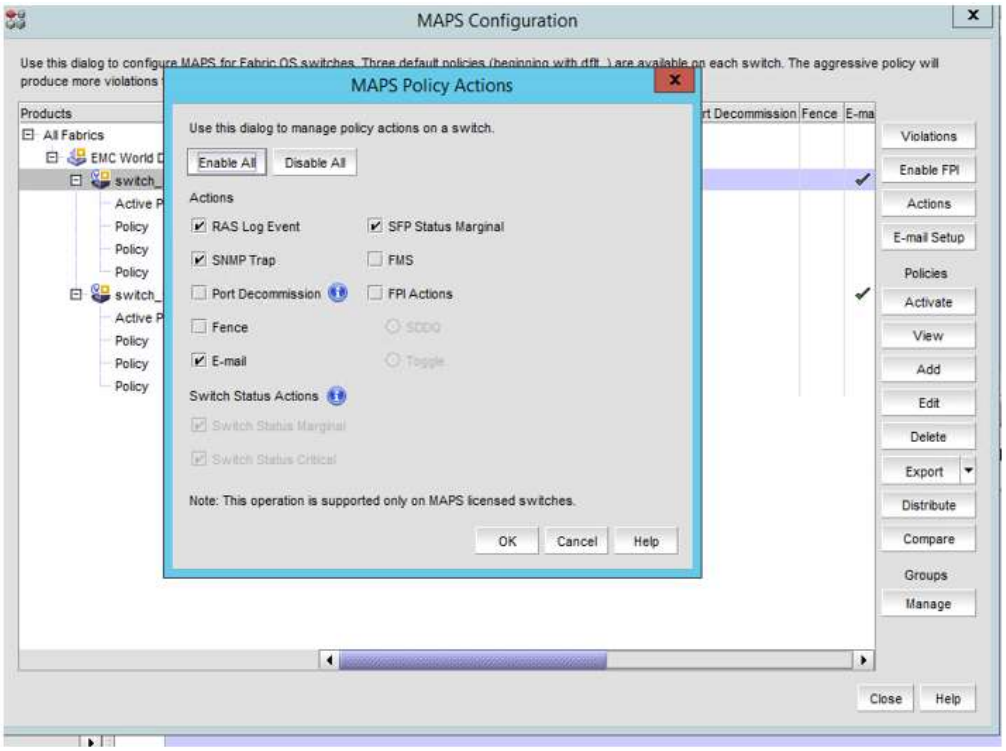
Nesse caso, a política agressiva padrão será ativada. Para isso, destaque **"dflt_aggressive_policy"** e clique em **activate**. Essa etapa deve ser repetida em TODOS os comutadores do fabric em que você deseja que a política seja habilitada. No momento, você não pode habilitá-la para todo o fabric.

Estamos ativando a política agressiva primeiro para ter uma ideia dos problemas no fabric imediatamente. Depois disso, você pode ajustar e usar as outras políticas se recebermos muitos alertas.
Se você clicar em **View**, poderá analisar os limites de cada evento.

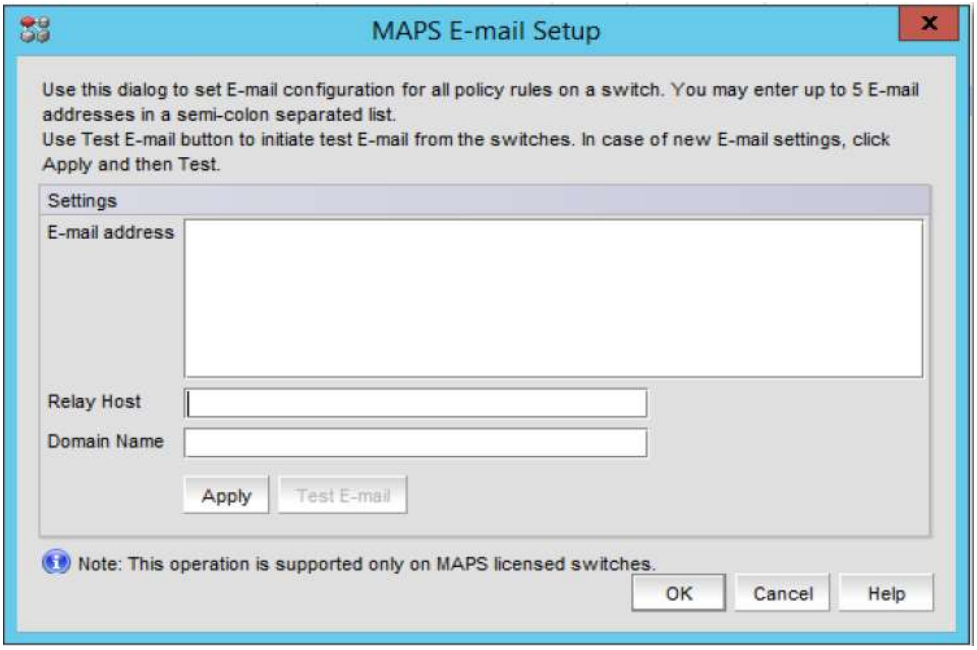




4. Destaque um comutador e clique em **Actions**. A partir daqui, você pode decidir sobre as ações que deseja realizar em caso de propagação do congestionamento. Para nosso estudo de caso específico de propagação de congestionamento devido à superatribuição, precisamos apenas garantir que o **evento de log do e-mail** e **do RAS** foram verificados.



5. Se você quiser receber alertas por e-mail, clique em **E-mail Setup** e preencha os campos adequados.

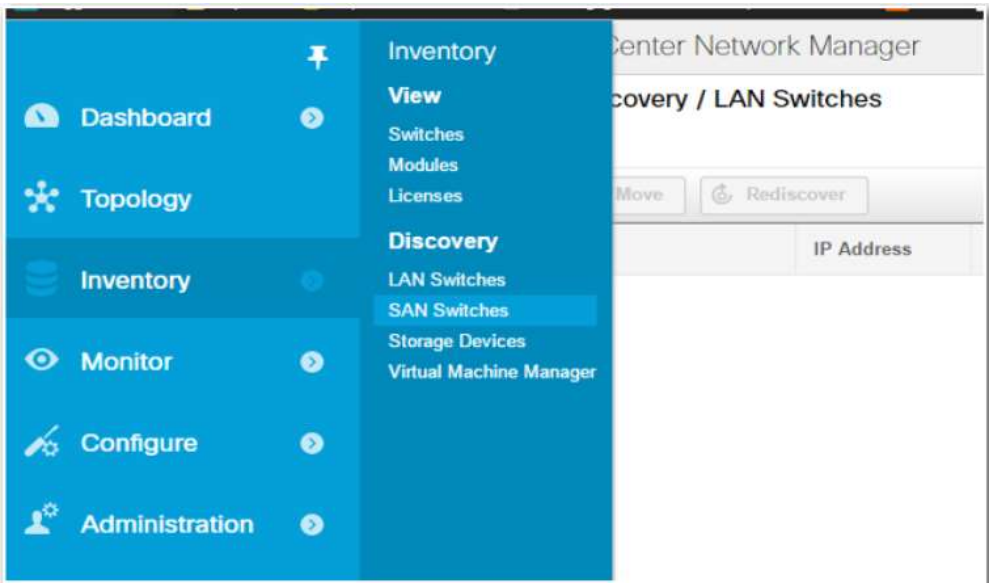


6. Certifique-se de que você tenha repetido essas etapas em TODOS os comutadores do fabric.

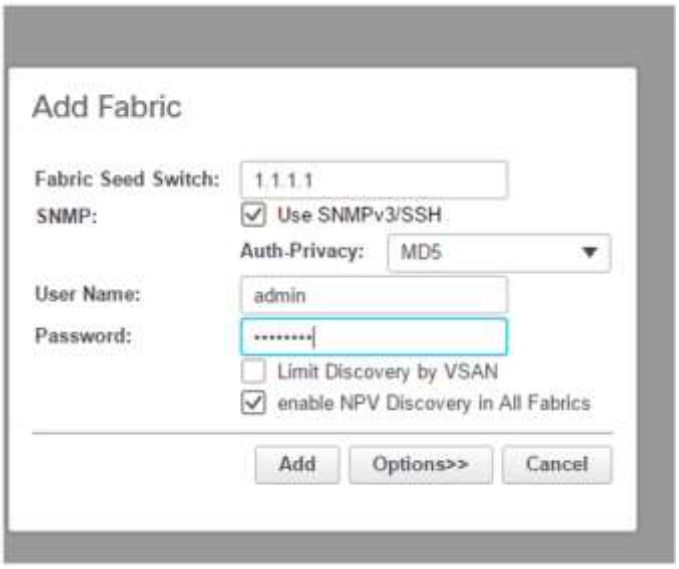
6.1.2 Cisco

- Identifique o fabric

1 Faça log-in no DCNM e clique em Inventory > Discovery > SAN Switches.



2. Na nova janela, clique no **sinal de adição (+**) e preencha as informações necessárias para um dos comutadores no fabric.

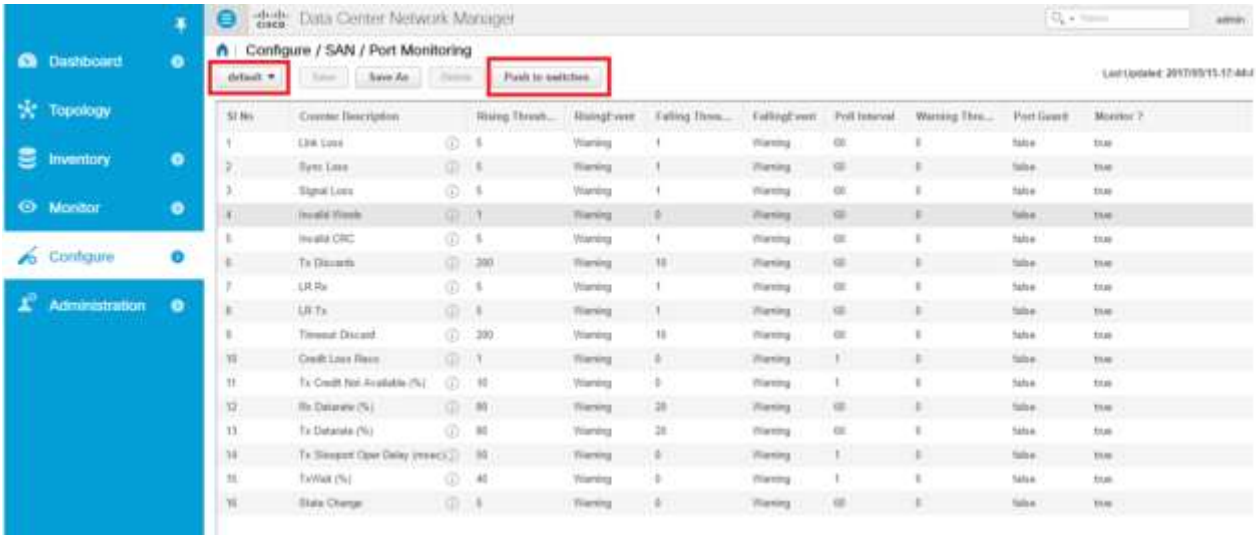


3. Repita esta seção para todos os outros fabrics.

- Ativar o monitoramento de porta da Cisco (PMON)

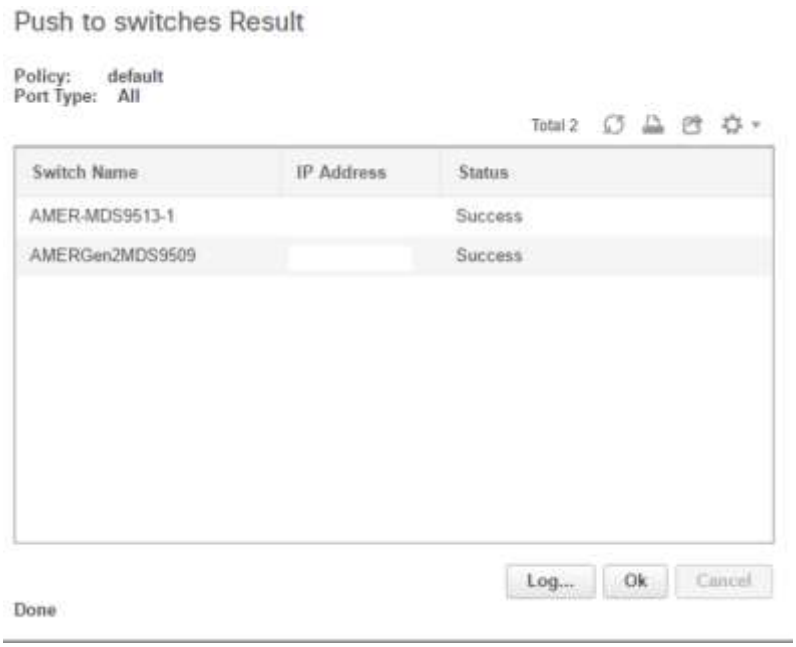
1 Clique em **Configure > SAN > Port Monitoring**.

2. Selecione o perfil padrão e clique **em Push to switches**.



3. Selecione todos os fabrics e clique em **Push**.

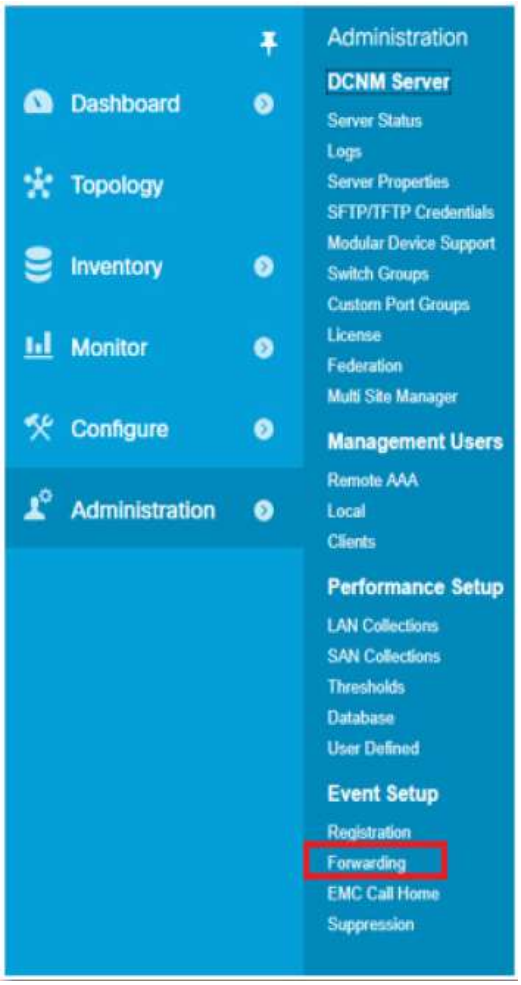
Obs.: Os endereços IP foram removidos propositalmente.



4. Com o Cisco MDS, você pode receber alertas por meio do SNMP ou Syslog. Consulte o seguinte guia de configuração para ambas as opções:

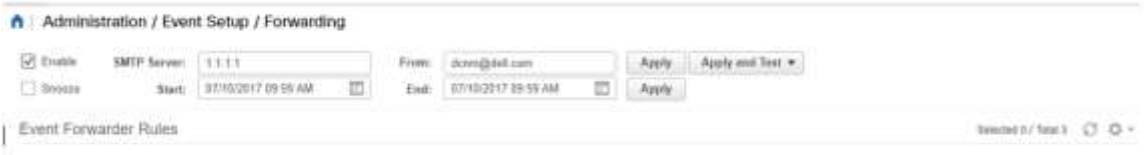
   http://www.cisco.com/c/en/us/support/storage-networking/mds-9000-nx-os-san-os-software/products-installation-and-configuration-guides-list.html

5. Para configurar o E-mail Home (opcional), clique em **Administration > Event Setup.**

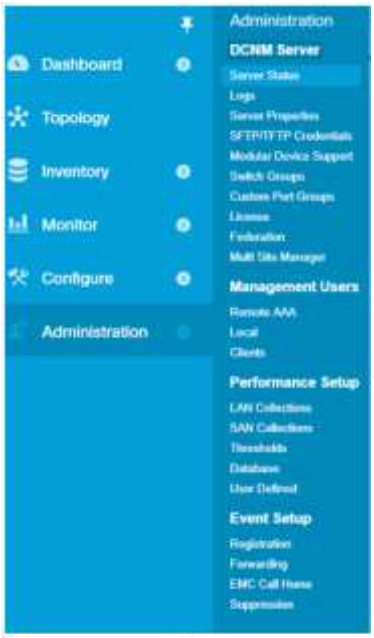


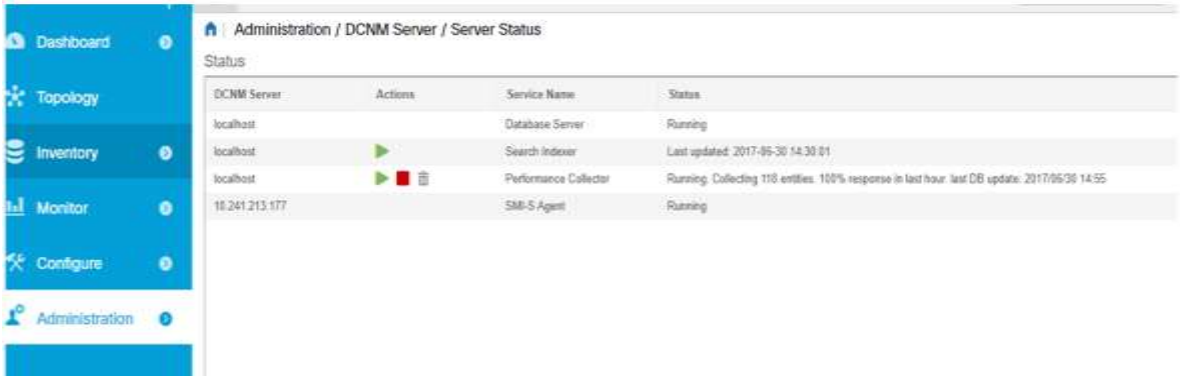
6. Clique no **sinal de adição (+),** informe o endereço de e-mail do destinatário e clique em **Add.**

7. Preencha as informações do servidor SMTP e o endereço de e-mail do remetente. Em seguida, clique em **Apply and Test** para confirmar que você recebeu o e-mail.

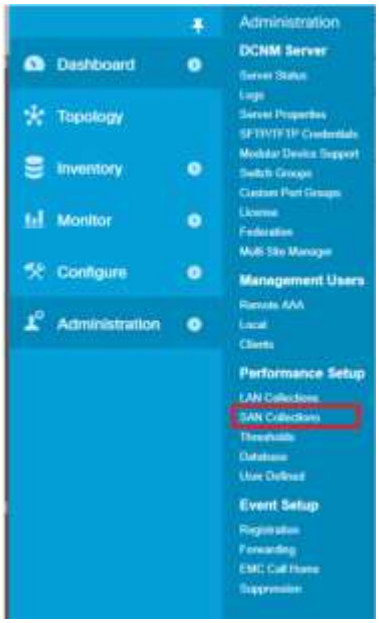


8. Verifique se o monitoramento de desempenho está em execução. Clique em **Administration > Server Status**. Certifique-se de que o **Performance Collector** esteja em execução. Caso contrário, aperte o botão **Play** para iniciá-lo.





9. Clique em **Administration > Performance Setup > SAN > Collections.**



10. Certifique-se de que os fabrics para os quais você deseja coletar as estatísticas de desempenho estejam verificados. O serviço do Performance Collector será reiniciado.

Referências

Guia de configuração do MAPS da Brocade:
http://www.brocade.com/content/html/en/configuration-guide/fos-80x-maps/GUID-426E1CD4-3763-419D-9D54-91F824F463EB-homepage.html

White paper sobre o dispositivo de drenagem lenta da Cisco:
http://www.cisco.com/c/dam/en/us/products/collateral/storage-networking/mds-9700-series-multilayer-directors/whitepaper-c11-737315.pdf

Referência geral sobre os recursos de limites de E/S de host do VMAX:
https://community.emc.com/thread/188068?start=0&tstart=0

Ferramenta de E/S Ezfio
https://github.com/earlephilhower/ezfio

Severidade da propagação de congestionamento

Embora as métricas de propagação de congestionamento sejam todas importantes, como a seção a seguir ilustra, a taxa na qual os eventos estão ocorrendo pode alterar drasticamente o impacto que cada evento pode ter em seu ambiente. O que complica ainda mais é o fato de que tanto a Brocade quanto a Cisco usam um esquema de categorização de severidade diferente. Como resultado, usaremos o seguinte esquema de categorização específico da Dell EMC e o mapearemos para cada um dos tipos de comutador, como mostrado abaixo:

6.1.3 Dell EMC

- Tipo 1:

  - Taxa de congestionamento maio ou igual a 0,2
  - Sem perda de quadros (descartes) ou redefinições de link

- Tipo 2:

  - Taxa de congestionamento maio ou igual a 0,2
  - Perda de quadros (descartes), mas sem redefinições de link

- Tipo 3:

  - Taxa de congestionamento maio ou igual a 0,2
  - Perda de quadros (descartes) e redefinições de link

6.1.4 Brocade

- Leve

  - Atraso de crédito reduzido
  - Baixa latência de fila (menos de 10 ms)
  - Sem perda de quadros (descartes) ou redefinições de link

- Moderado

  - Atraso de crédito médio
  - Latência de fila média (10 a 80 ms)
  - Perda de quadros (descartes), mas sem redefinições de link

- Severo

  - Grande atraso de crédito
  - Alta latência de fila (maior que 80 ms)
  - Perda de quadros (descartes) e algumas redefinições de link

6.1.5 Cisco

- Nível 1: Latência

  - Número reduzido de créditos restantes ou indisponibilidade de crédito com pouca duração
  - Sem Descartes, retransmissão ou redefinições de link

- Nível 2: Retransmissão

  - Duração mais longa da indisponibilidade de crédito
  - Os quadros são descartados (mas sem redefinição de link) devido ao tempo limite da queda de congestionamento ou da queda sem crédito* que leva à retransmissão.

- Nível 3 Atraso extremo

  - Duração prolongada da indisponibilidade de crédito (1 segundo para porta F, 1,5 segundo para a porta E)
  - Redefinições de link ou oscilações de porta

Referência cruzada de terminologia de propagação de congestionamento

As métricas e as severidades podem ser combinadas e usadas para ajudar a identificar os diferentes tipos de eventos de propagação de congestionamento. Como na seção anterior, há uma seção separada para Brocade e Cisco, mas como elas usam o termo superatribuição, esta seção começará com uma visão geral sobre esse assunto.

6.1.6 Superatribuição

A superatribuição é simplesmente uma condição em que “a possível demanda em um sistema excede a capacidade dele de atendê-la”. Um exemplo com o qual muitos estão familiarizados é o sistema de estrada. Se todo mundo repentinamente decidisse dirigir seus carros ao mesmo tempo (como durante um evento de evacuação devido a furacão), o tráfego ficaria paralisado.

No caso de uma FC SAN, é relevante considerar a superatribuição em termos de uma taxa de largura de banda (BW). Por exemplo, conforme mostrado na Figura 3, a proporção de largura de banda entre o Host 1 (4 Gbps) e o Armazenamento 1 (16 Gbps) é 1:4. Portanto, podemos dizer que o Host 1 é o 4:1 com superatribuição. Compare isso com a proporção de largura de banda entre o Host 2 (16 Gbps) e o Armazenamento 2 (16 Gbps), que é 1:1. Considere a possibilidade de que os hosts e o armazenamento que eles acessam utilizam um ISL de 32 Gbps e de que você possa ver que não há superatribuição entre o Host 2 e o Armazenamento 2. Nesse caso, dizemos que o Host 2 e o Armazenamento 2 não estão com superatribuição.

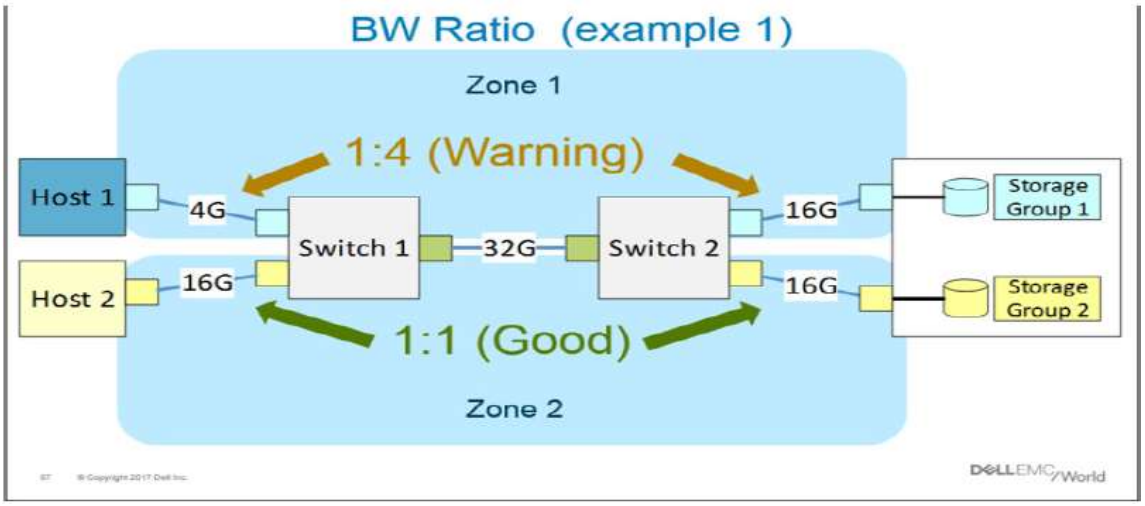


Figura 19 Taxa de largura de banda - Exemplo 1

É importante observar que, ao calcular a superatribuição, como mostrado na Figura 19, a taxa de largura de banda é calculada pela adição da largura de banda das interfaces que estão sendo consideradas. A princípio, você pode achar que temos um HBA de 16 Gbps acessando o armazenamento de 8 Gbps, mas como há, na verdade, três interfaces de armazenamento, temos um HBA de 16 Gbps acessando 24 Gbps de armazenamento. Como resultado, o host é de 3:2 com superatribuição.

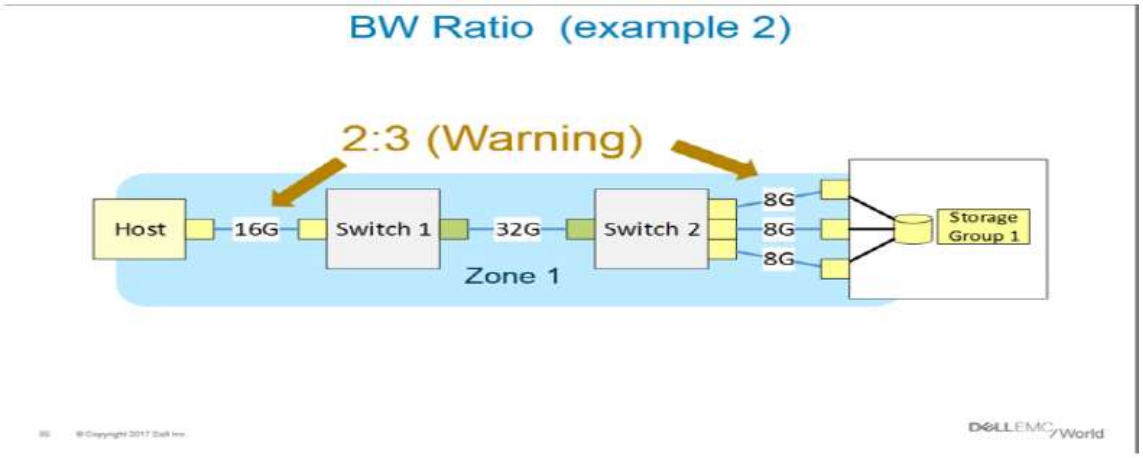


Figura 20 Taxa de largura de banda – Exemplo 2

Nos dois exemplos anteriores, a largura de banda do ISL era sempre maior ou igual ao valor de largura de banda que os dispositivos finais podiam dar suporte. Geralmente, esse não é o caso. Conforme mostrado na Figura 20, o host é, na verdade, de 3:4 com superatribuição, mas como o ISL é apenas de 16 Gbps, há uma superatribuição entre o dispositivo final e os ISLs que serão usados, e você pode dizer que os ISLs são de 3:2 com superatribuição.

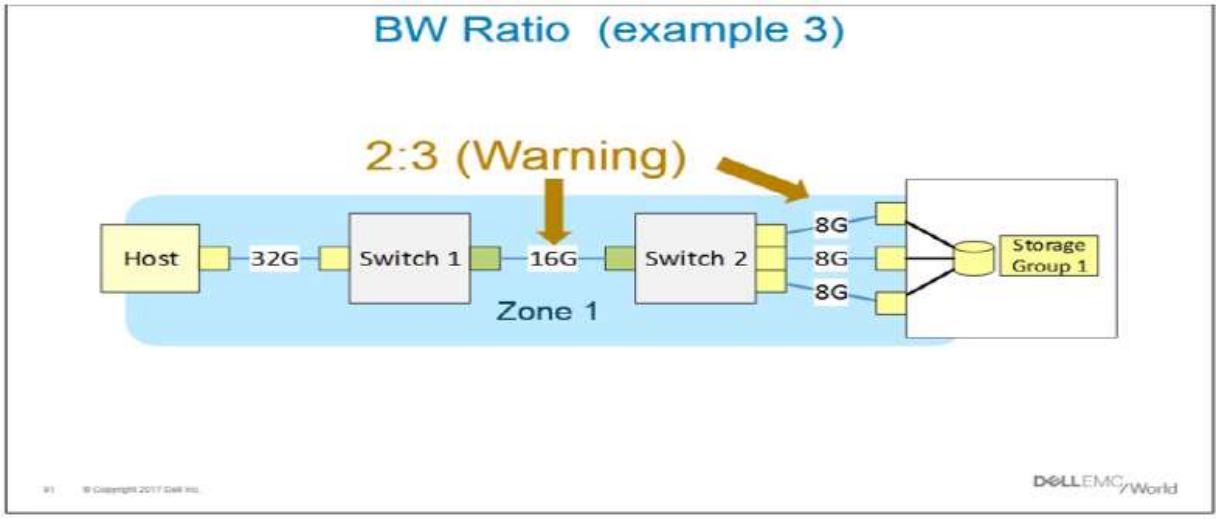


Figura 21 Taxa de largura de banda – Exemplo 3

6.1.7 Brocade

A Brocade define três classes diferentes de eventos de propagação de congestionamento:

- Superatribuição

Conforme definido na seção anterior (acima).

- Dispositivo com comportamento inadequado

Um dispositivo final ou um ISL que não está liberando o crédito a uma taxa rápida o suficiente para sustentar a transmissão de dados. Por exemplo, se um dispositivo final tiver negociado uma velocidade de link de 16 Gbps e não puder devolver o crédito a uma taxa que o permita receber 16 Gbps de dados, você poderá ter um dispositivo com comportamento inadequado. Esses tipos de dispositivos também são chamados de “drenagens lentas”. É importante destacar que um dispositivo pode ter comportamento inadequado por vários motivos, inclusive por um problema de driver ou, no caso de um ISL, porque a porta está enfrentando os efeitos da propagação do congestionamento.

- Crédito perdido

Um cenário de crédito perdido significa que, por algum motivo (por exemplo, erros de bits ocasionais), um ou ambos os dispositivos de um determinado link acreditam que eles têm menos crédito de transmissão do que realmente têm. Uma causa dessa situação seria um erro de bit para corromper um R_RDY. Se isso ocorrer com frequência suficiente, o desempenho começará a reduzir ao longo do tempo e degradará lentamente a capacidade da SAN de transportar dados. Esse problema é explorado com mais detalhes em KB 464245 (Erros de bit e seu impacto).

Referência cruzada de terminologia da propagação de congestionamento da Brocade

Para a Brocade, reunir todas as peças resulta na seguinte Referência cruzada de terminologia de propagação de congestionamento específico da empresa.

Causa	Leve	Moderado	Severo
Superatribuição[1]	1. Grande largura de banda na porta do dispositivo 2. Baixa latência de crédito na porta ISL 3. Sem perda de quadros ou redefinições de link	1. Grande largura de banda na porta do dispositivo 2. Latência de crédito média na porta ISL 3. Entre 10 m a 80 ms de latência de fila na porta ISL 4. Sem perda de quadros ou redefinições de link	1. Grande largura de banda na porta do dispositivo 2. Grande latência de crédito na porta ISL 3. Latência de fila maior que 80 ms na porta ISL 4. Perda de quadros na porta upstream (ISL) (indica latência de fila de 220 ms a 500 ms) 5. Sem redefinições de link.
Dispositivo com comportamento inadequado	1. Baixa latência de crédito na porta do dispositivo e na porta ISL upstream 2. Menos de 10 ms de latência de fila na porta do dispositivo e na porta ILS upstream 3. Sem perda de quadros ou redefinições de link	1. Média latência de crédito na porta ISL e na porta ISL upstream 2. Entre 10 ms a 80 ms de latência de fila na porta do dispositivo e na porta ILS upstream 3. Sem perda de quadros ou redefinições de link	1. Grande latência de crédito na porta do dispositivo e na porta ISL upstream 2. Latência de fila maior que 80 ms na porta do dispositivo e na porta ILS upstream 3. Perda de quadros na porta do dispositivo ou upstream (ISL) (indica latência de fila de 220 ms a 500 ms) 4. Redefinição de link em uma porta ISL (indica paralisação de crédito por mais de 2 s)
Crédito perdido[2]	1. Baixa latência de crédito na porta 2. Menos de 10 ms de latência de fila na porta ou upstream da porta 3. Sem perda de quadros ou redefinições de link	1. Latência de crédito média na porta 2. Entre 10 ms a 80 ms de latência de fila na porta ou link de upstream da porta 3. Sem perda de quadros ou redefinições de link	1. Grande latência de crédito na porta 2. Latência de fila maior que 80 ms na porta ou upstream da porta 3. Perda de quadros na porta ou upstream da porta (indica paralisação de crédito por 220 ms a 500 ms) 4. Redefinição de link na porta ou upstream da porta (indica paralisação de crédito por mais de 2 s)

[1]A ocorrência de congestionamento grave devido a superatribuição é rara ou extremamente rara.
[2]As causas da perda de crédito geralmente são erros de transmissão, como ITW e CRC, ou outros problemas relacionados a sinal.

6.1.8 Cisco

A Cisco define duas classes diferentes de eventos de propagação de congestionamento:

- Superatribuição

  Conforme definido acima.

- Escassez de crédito

Um dispositivo final ou um ISL que não está liberando o crédito a uma taxa rápida o suficiente para sustentar a transmissão de dados. Por exemplo, se um dispositivo final tiver negociado uma velocidade de link de 16 Gbps e não puder devolver o crédito a uma taxa que o permita receber 16 Gbps de dados, você poderá ter um dispositivo com comportamento inadequado. Esses tipos de dispositivos também são chamados de “drenagens lentas”. É importante destacar que um dispositivo pode ter comportamento inadequado por vários motivos, inclusive por um problema de driver ou, no caso de um ISL, porque a porta está enfrentando os efeitos da propagação do congestionamento.

- Referência cruzada de terminologia de propagação de congestionamento da Cisco

Para a Cisco, reunir todas as peças resulta na seguinte Referência cruzada de terminologia de propagação de congestionamento específico da empresa.

Tipo de congestionamento	Nível - 1: Latency	Nível - 2: Retransmissão	Nível - 3: Atraso extremo
Superatribuição	1. Alta utilização de link na porta do dispositivo final 2. Sem escassez de crédito B2B na porta do dispositivo final 3. Propagação de congestionamento em direção aos ISLs 4. Sem perda de quadros ou redefinições de link	A ocorrência de retransmissão ou atraso extremo devido a superatribuição é rara ou extremamente rara	
Escassez de crédito	1. Baixa utilização de link na porta do dispositivo final 2. Número reduzido de créditos restantes ou indisponibilidade de crédito com pouca duração 3. Propagação de congestionamento em direção a ISLs 4. Sem Descartes, retransmissão ou redefinições de link	1. Baixa utilização de link na porta do dispositivo final 2. Duração mais longa da indisponibilidade de crédito 3. Propagação de congestionamento em direção a ISLs 4. Os quadros são descartados (mas sem redefinição de link) devido ao tempo limite da queda de congestionamento ou da queda sem crédito* que leva à retransmissão	1. Nenhum quadro é transmitido para o dispositivo final 2. Duração prolongada da indisponibilidade de crédito (1 segundo para a porta F, 1,5 para a porta E) 3. Congestionamento severo em direção a ISLs 4. Redefinições de link ou oscilações de porta
*Configuração padrão: tempo limite da queda de congestionamento – 500 ms, tempo limite da queda sem crédito – desativado Opção configurável: tempo limite da queda de congestionamento – 100 ms a 500 ms, tempo limite da queda sem crédito – 1 ms a 500 ms *Configuração recomendada: tempo limite da queda de congestionamento – 200 ms, tempo limite da queda sem crédito – 50 ms			