
cisco Live!
Let's go
#CiscoLive

The bridge to possible

# SAN Analytics & SAN Insights

## Real-time & Always-on NVMe and SCSI Visibility at Scale

Paresh Gupta
Technical Leader, Technical Marketing Engineering, Cisco
@reach2paresh

BRKDCN-3645

# Cisco Webex App

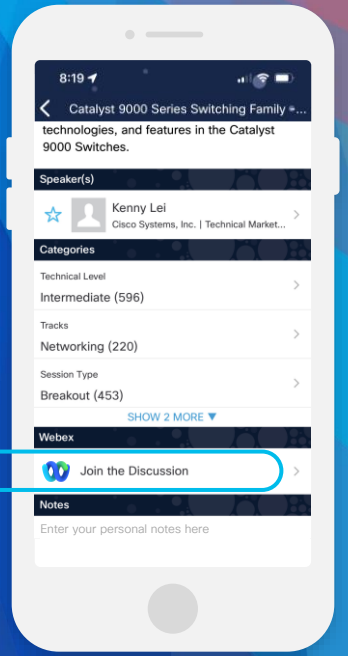
## Questions?

Use Cisco Webex App to chat
with the speaker after the session

## How

1. Find this session in the Cisco Live Mobile App
2. Click “Join the Discussion”
3. Install the Webex App or go directly to the Webex space
4. Enter messages/questions in the Webex space

## Webex spaces will be moderated by the speaker until June 9, 2023.

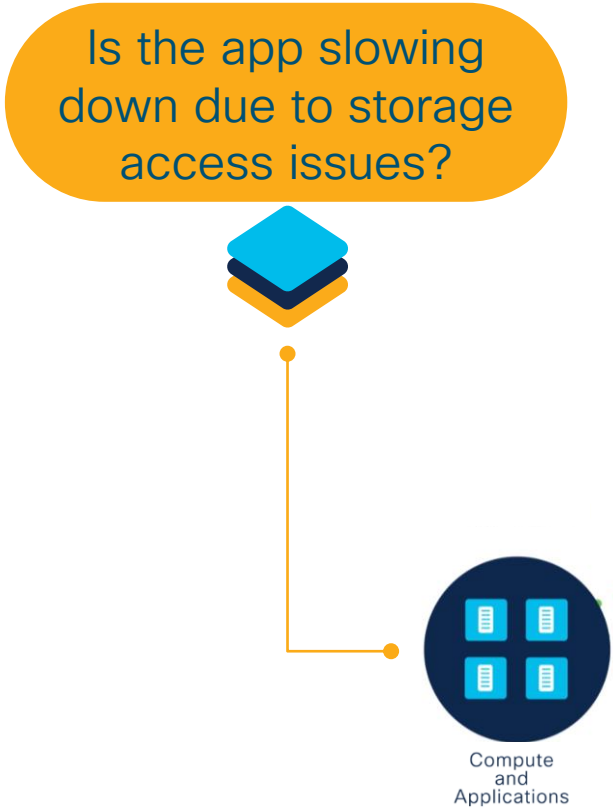



https://ciscolive.ciscoevents.com/ciscolivebot/#BRKDCN-3645

# Agenda

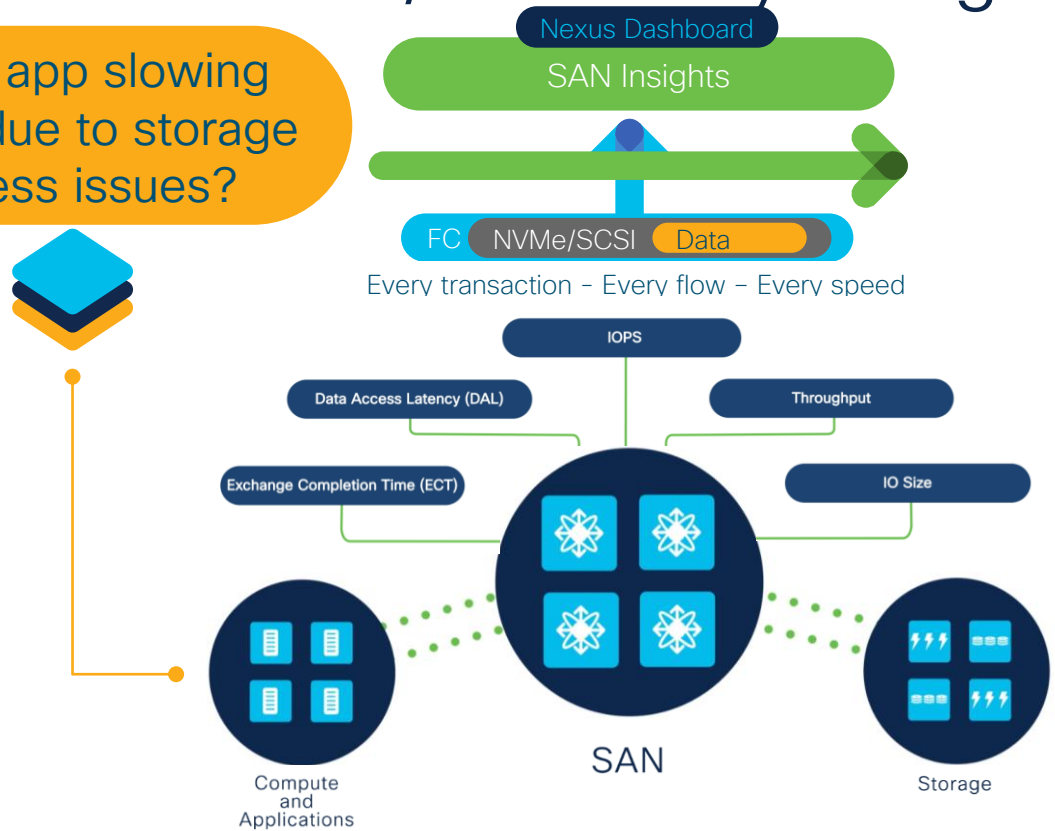
- Solution Overview
- SAN Analytics Architecture
- SAN Analytics Deployment
- I/O Flow Metrics
- Use-cases and case studies
- Summary

# NVMe and SCSI I/O Visibility Using SAN Analytics



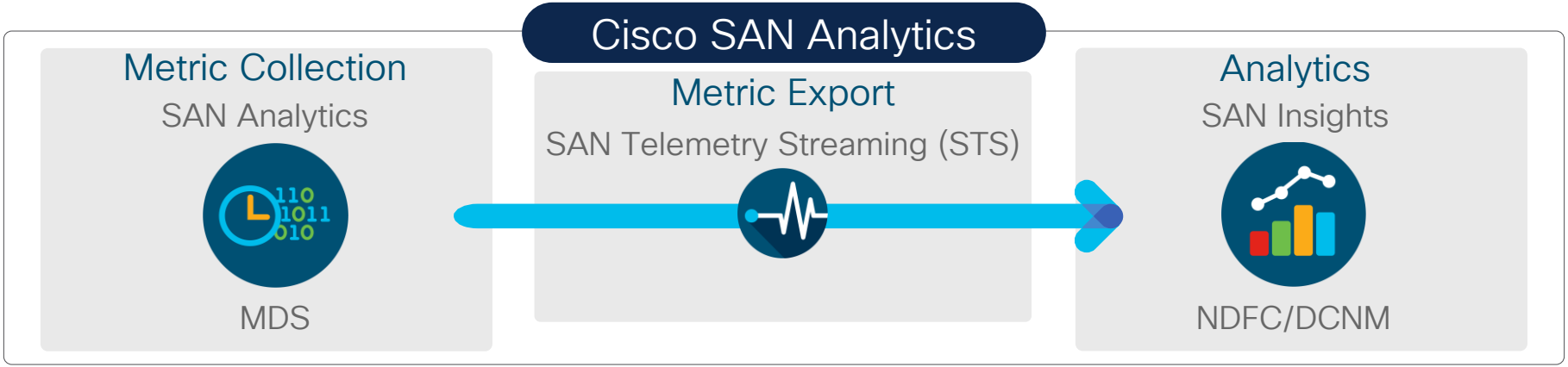
# NVMe and SCSI I/O Visibility Using SAN Analytics

Is the app slowing down due to storage access issues?




# Solution Components

‘Cisco SAN Analytics’ is the umbrella name for the overall solution

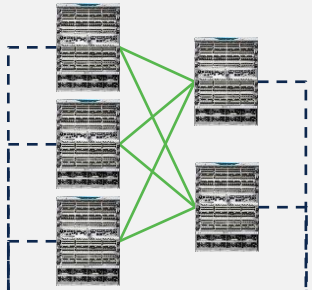



‘SAN Analytics’ is also the name of the feature to enable flow metric collection on MDS switches (NX-OS command: `feature analytics`)

‘SAN Telemetry Streaming’ is an efficient mechanism to export metrics from MDS switches (NX-OS command: `feature telemetry`)

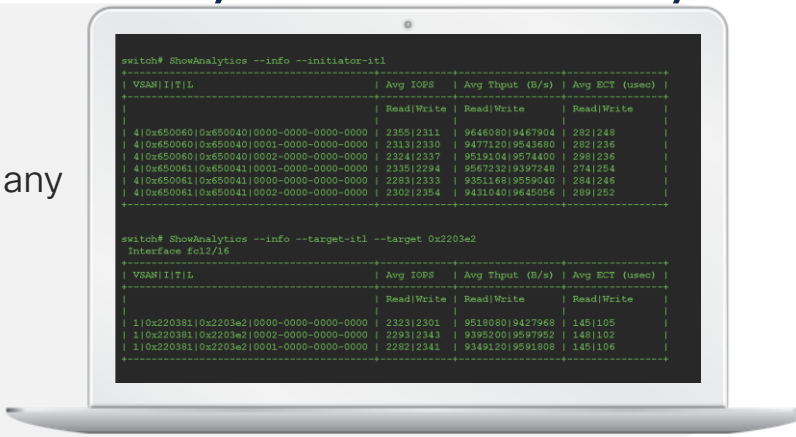
‘SAN Insights’ is an analytics and visualization engine within NDFC/DCNM

# Cisco SAN Analytics Scale for Always-on Visibility




## MDS 9700

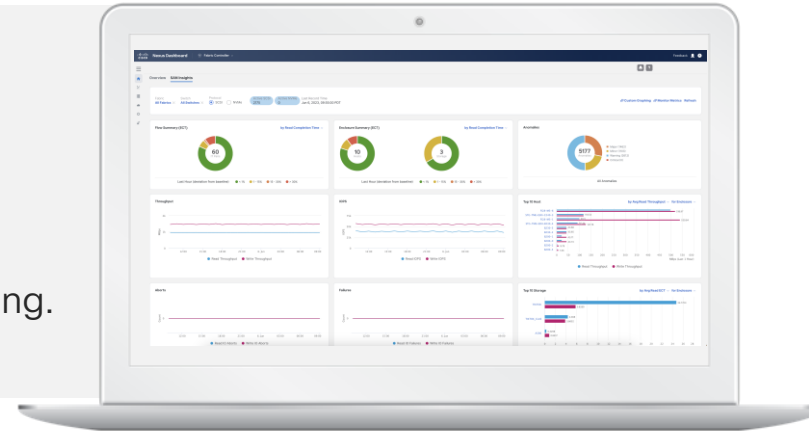
- 100K flows monitored per MDS 9700
- Real-time (microsecond) visibility into any flow using on-switch CLI
- 30-second export interval to NDFC





## NDFC SAN Insights

- 1 Million flows received from multiple switches
- Ready-made use cases
- Anomaly detection
- Automatic baseline, deviations, trending.
- End-to-end correlation



# Cisco 32G SAN Analytics – Architecture

## Traffic Inspection

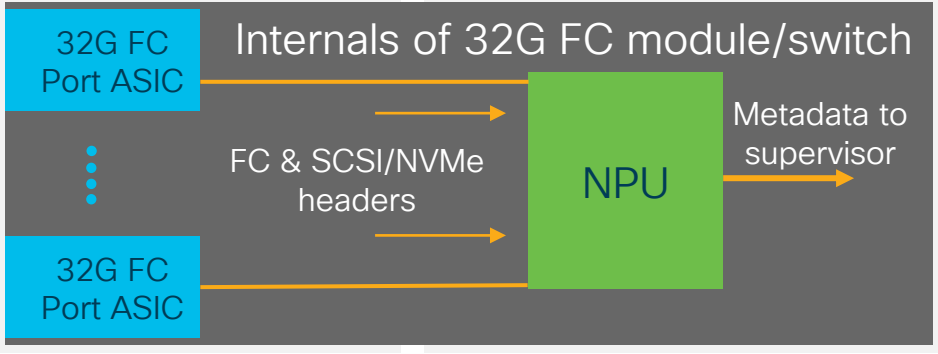
- Inbuilt tap in 32G FC port-ASIC
- Traffic inspection capability on all ports
- Zero impact to switching functionality
- Inspects only FC & SCSI/NVMe headers, not data

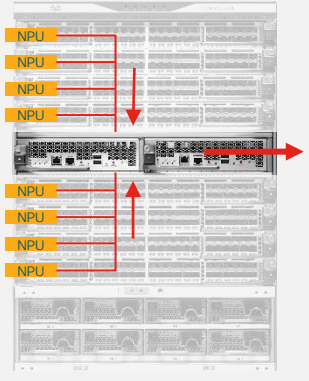
## Metric Calculation

- Network Processing Unit (NPU) on 32G FC products
- Receives headers of specific frames from port-ASIC
- Extracts metrics from headers
- Stores metrics in multiple views

## Metric Export

- SAN Telemetry Streaming (STS) exports flow metrics to external receivers
- Extremely efficient mechanism
- Works using existing mgmt. port
- On-switch CLI and remote RESTful access available




External Receivers: DCNM SAN Insights Or Virtual Instruments or any other 3rd party app

2022

# Cisco SAN Analytics

using

Cisco MDS 64G Switches

2017

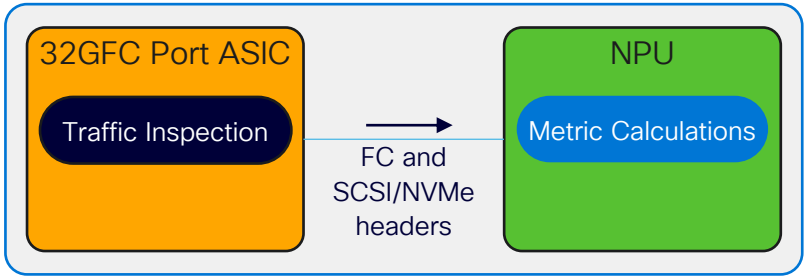
# Cisco SAN Analytics

using

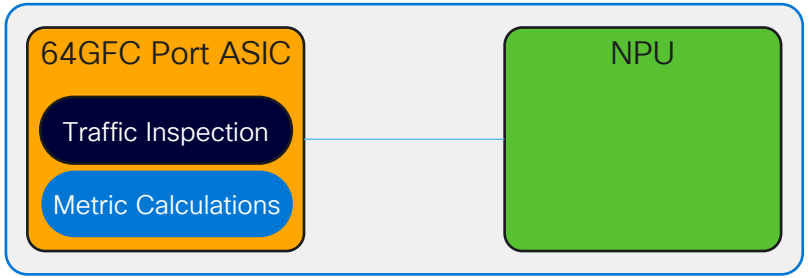
Cisco MDS 32G Switches

# Cisco SAN Analytics Architecture

## Using Cisco MDS 32G switches




## Using Cisco MDS 64G switches



# Cisco SAN Analytics

Using
Cisco MDS 64G switches

## Analytics for Billions of IOPS

Traffic inspection and metric calculation in ASIC

## Software Programmability

On-board Network Processing Unit

## Additional flow metrics

Host Response Latency, First Burst,
Optimized Read

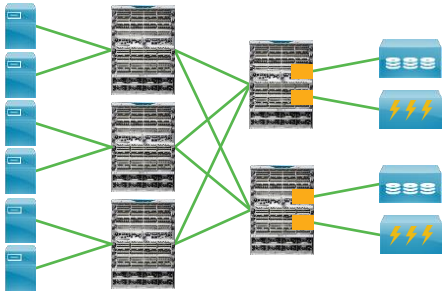
## Investment Protection

1GbE streaming port on the 64GFC module

Analytics enabled port(s)

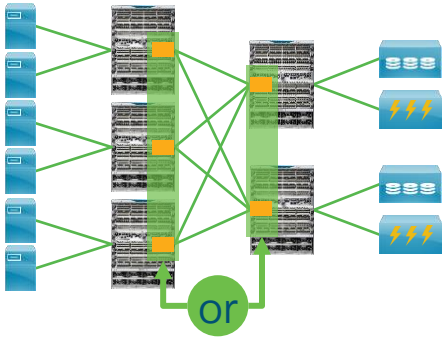
# Deployment Models

## Storage Ports



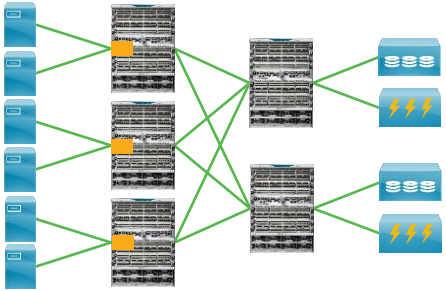
Closest to storage

## ISL Ports




High capacity 64G ISLs

## Host Ports

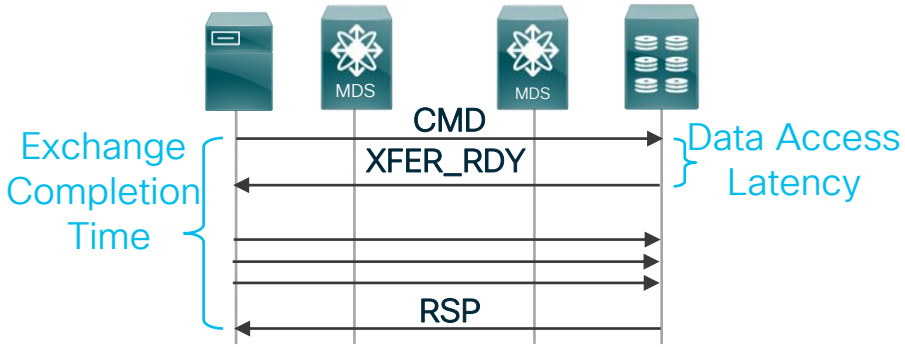


Closest to apps

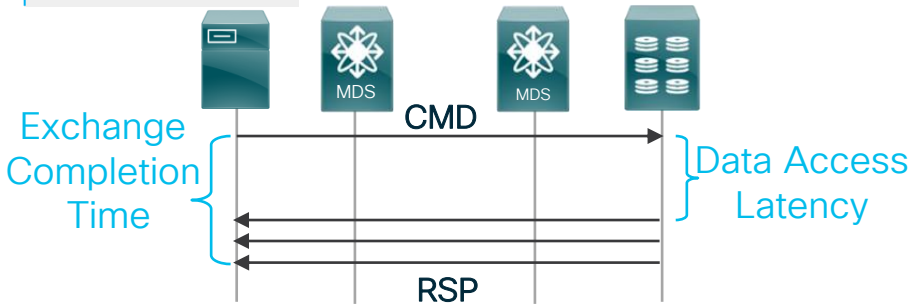
- Inspection of traffic at least once in the end-to-end data path is enough
- Avoid double inspection of traffic
- Design for uniform utilization of the NPU
- Most customers are enabling analytics on storage ports

# 80+ Metrics per SCSI or NVMe Flow

Write CMD




Read CMD




- Exchange Completion time
- Data Access Latency
- Outstanding IO
- IOPS
- I/O Size
- Failed exchanges, IO retransmissions
- Error conditions (Aborts, Rejects, etc.)
- Check conditions, queue-full, etc.

Measured at Initiator-Target-LUN/Namespace (ITL or ITN) level

# Metric Export

## On-switch CLI

**Flexible** SQL-like CLIs

- Pulls raw data
- Output in JSON format (key-value, similar to dictionaries)
- Basic trending & correlation

**Programmable** ShowAnalytics

- Output in nice-to-read tabular format, similar to any other NX-OS show command output

## NX-API

- Capability to query switch remotely
- Response in JSON format
- Response depends on the SQL-like CLI input
- Works best to track specific metrics
- Extremely flexible

## Streaming Telemetry

- MDS streams out the metrics regularly
- Data transport and encoding in industry leading open formats
- Optimized performance for continuous data export at high frequency

SAN Insights

3rd party apps

# Metric Export

## On-switch CLI

**Flexible** SQL-like CLIs

- Use for low-level troubleshooting in real-time
- Required detailed understanding of metrics, their format and units

**Programmable** ShowAnalytics

- Use for low-level troubleshooting in real-time
- Common use-cases – already available
- Special use-cases – Write your own

## NX-API

- Use for occasional access of selective metrics only. Example: a custom script which is invoked occasionally.
- Requires a simple http(s) call to the switch (NX-API)
- No need to setup a gRPC receiver
- Performance may be an issue at scale. Use streaming telemetry with large number of flows

SAN Insights

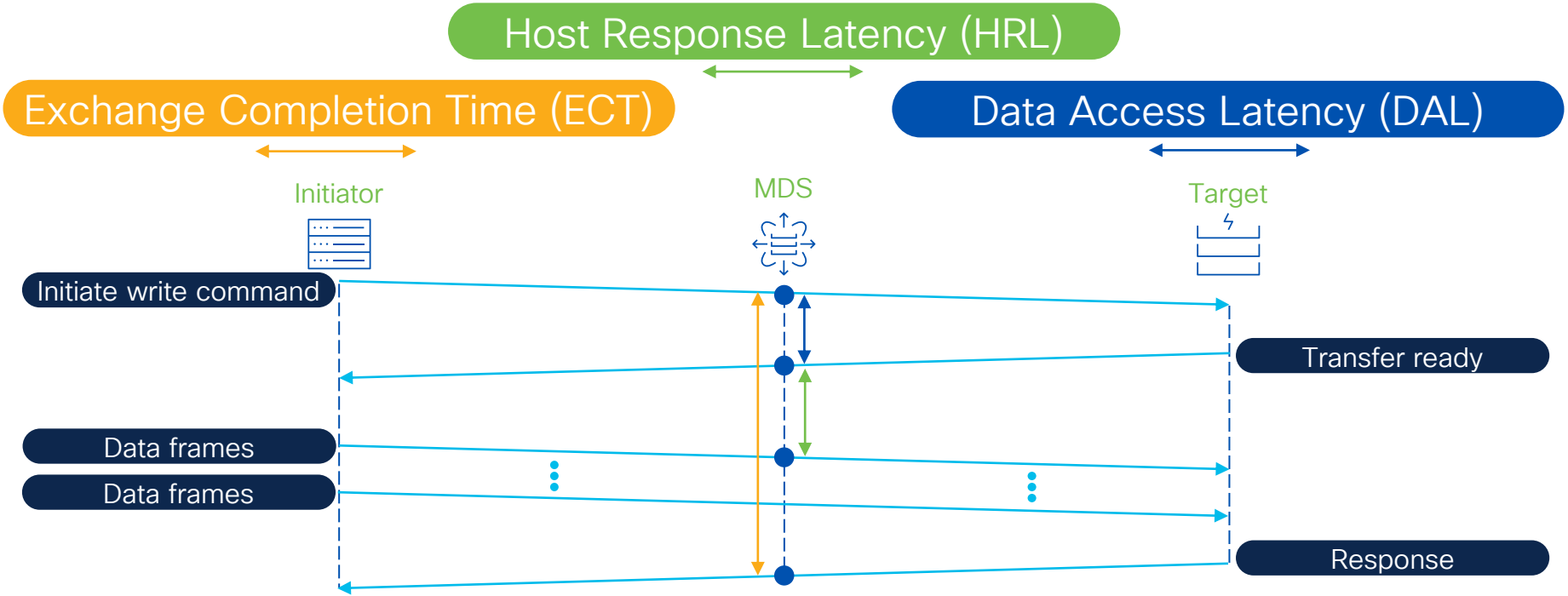
3rd party apps

## Streaming Telemetry

- Use for continuous export of metrics. Example: A continuous monitoring app
- Requires a special receiver to understand gRPC.

# Cisco SAN Analytics Flow Metrics

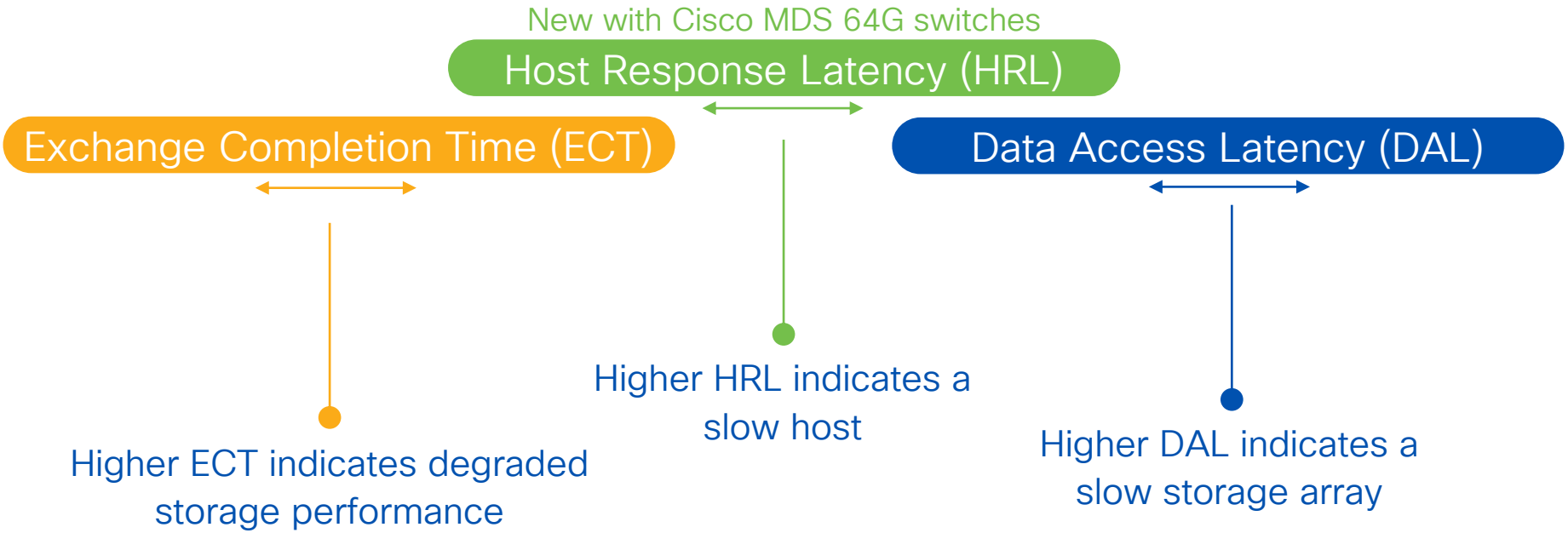
## Write I/O Operation




HRL is available only on 64G MDS switches. ECT and DAL are available on 32G MDS as well.

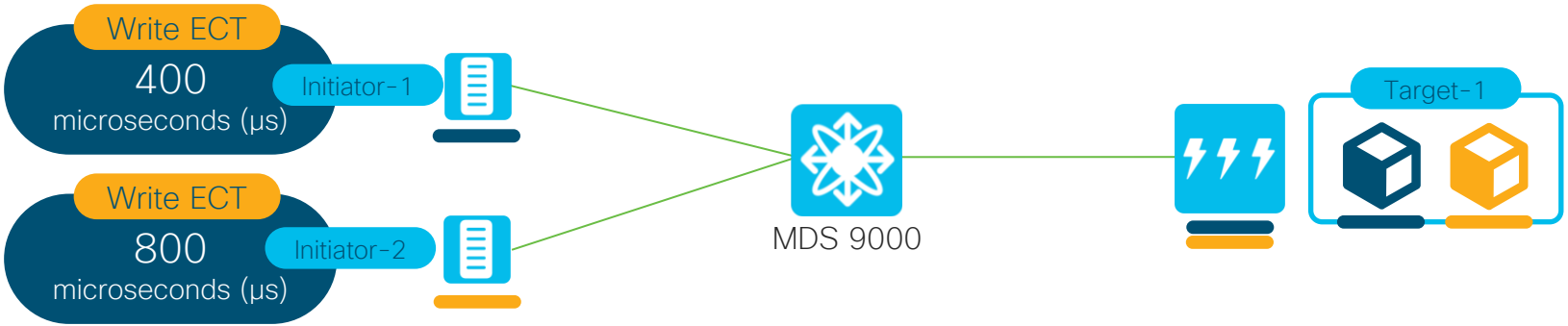
# Cisco SAN Analytics Flow Metrics

## Write I/O Operation

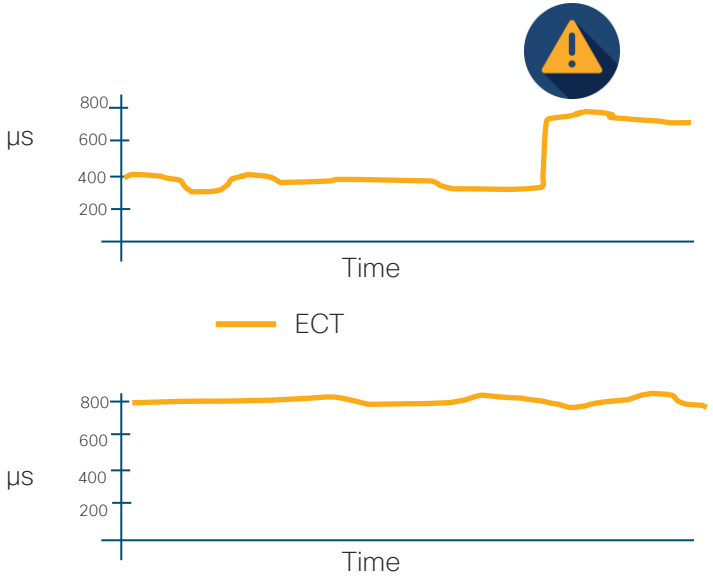


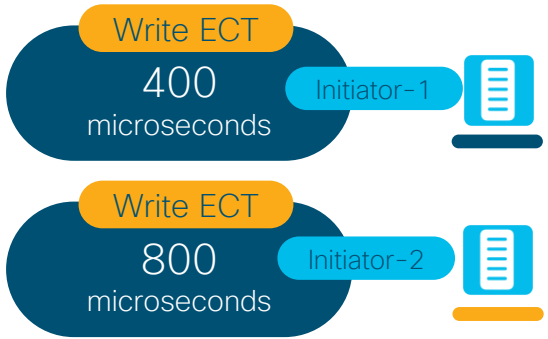
# Pin-pointing Storage I/O Performance Issues

Which server is performing better ?



# Pin-pointing Storage I/O Performance Issues



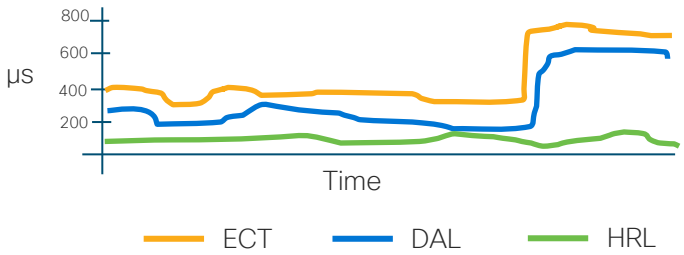




Which server is performing better ?

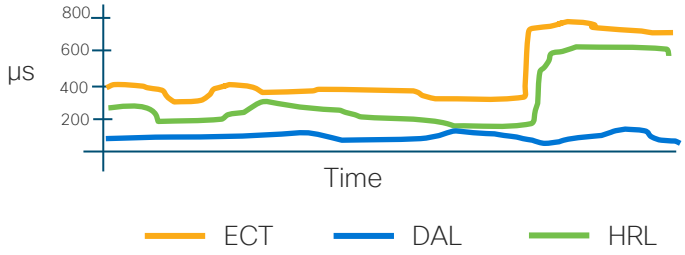
# Pin-pointing Storage I/O Performance Issues

## Where is the bottleneck ?

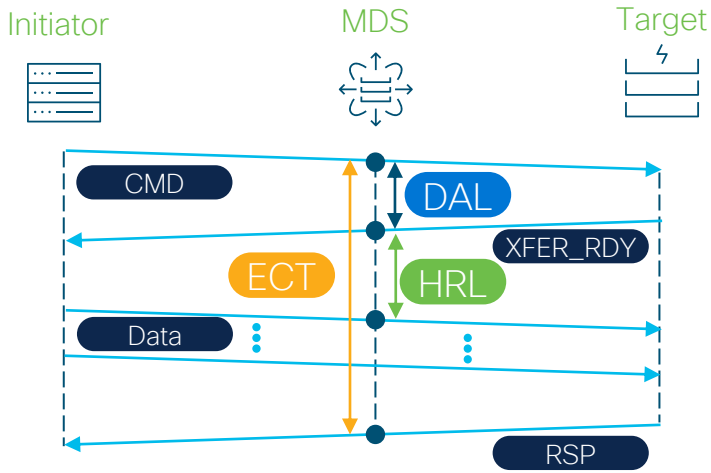
Storage performance is degraded due to delay caused by storage array




Storage performance is degraded due to delay caused by host



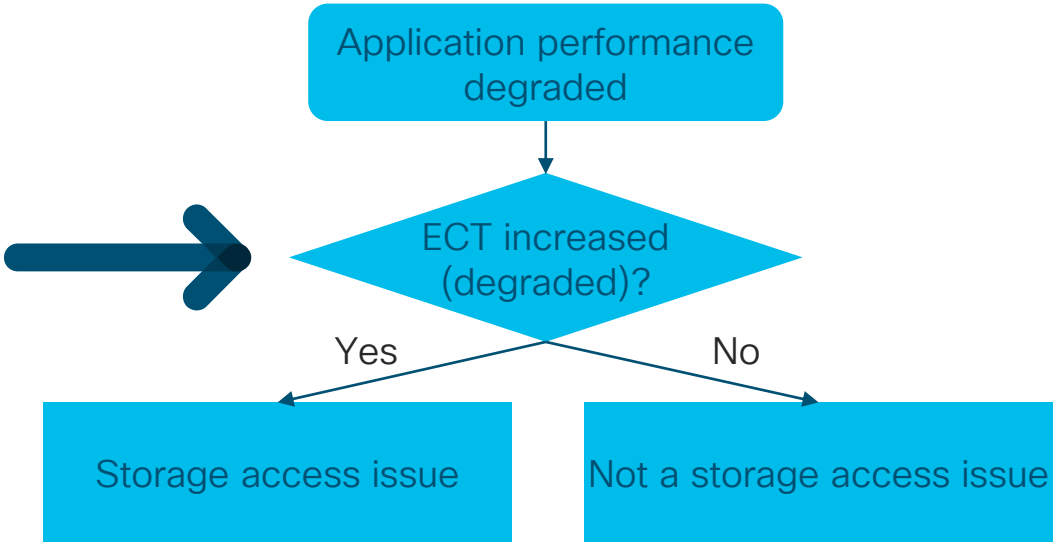

Write transaction



# Using ECT, DAL, and HRL for pin-pointing the delays

## 1st level pin-pointing

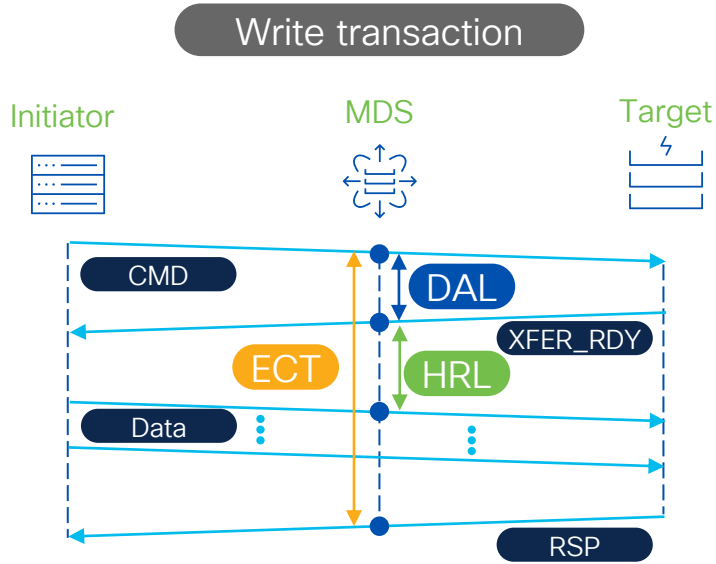
- Increase in ECT may directly lead to application slowdown and is the first level of pin-pointing towards storage access issue



# Using ECT, DAL, and HRL for pin-pointing the delays

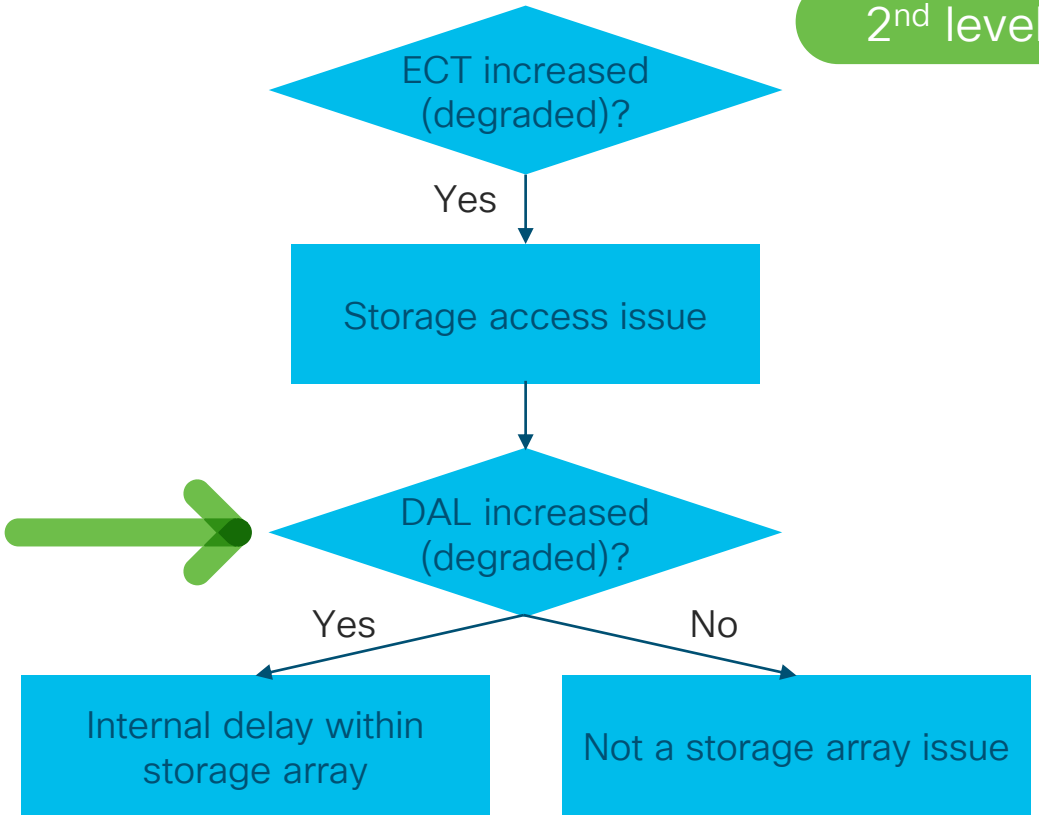
## 2nd level pin-pointing

- ECT may increase (degrade) due to
  - Internal delay within storage array
  - Delay in the fabric (SAN Congestion)
  - Internal delay within host

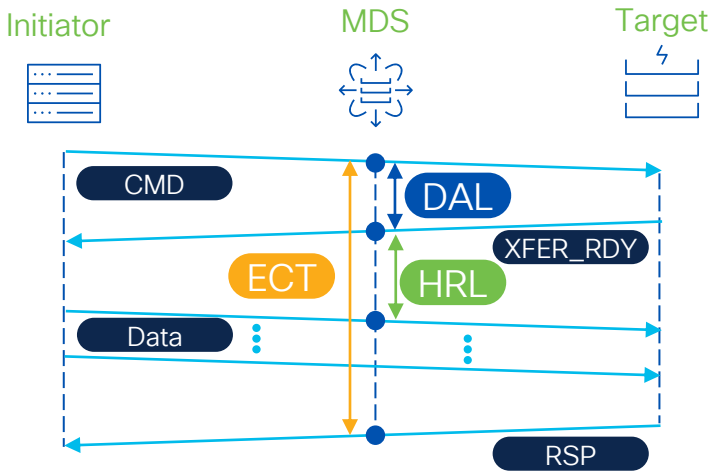


# Using ECT, DAL, and HRL for pin-pointing the delays

2nd level pin-pointing

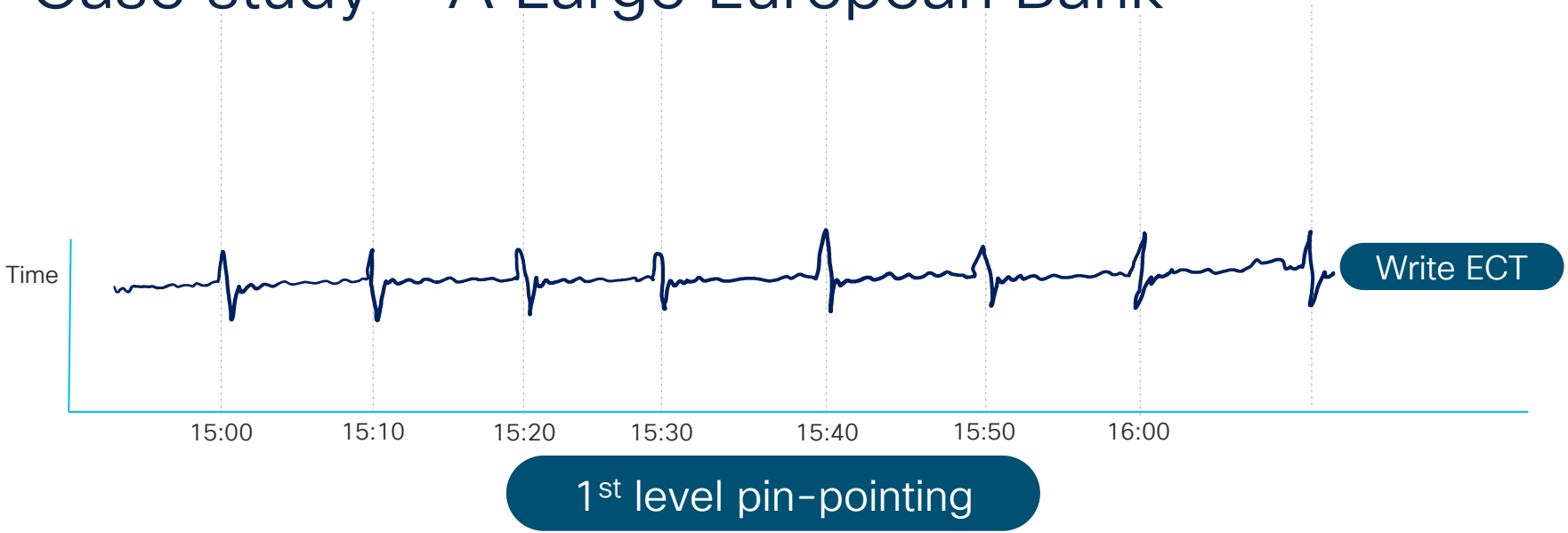



Write transaction



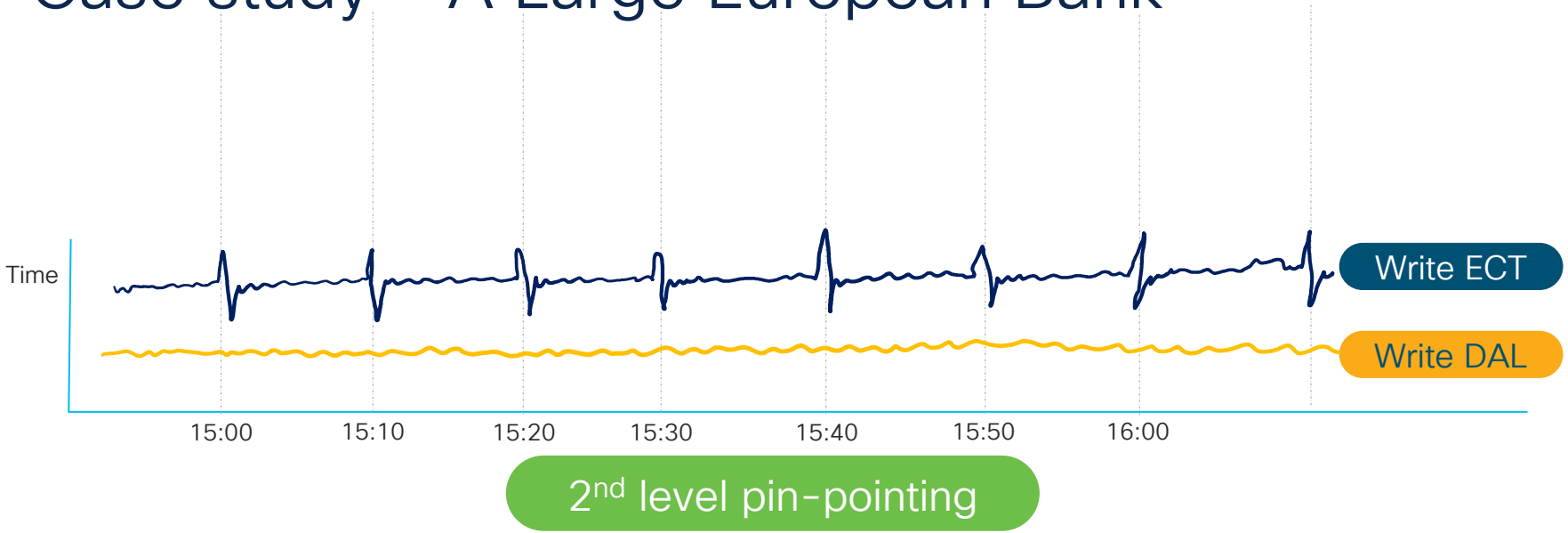

- Delay in the fabric (SAN Congestion)
- Internal delay within host (HRL)

# Case study – A Large European Bank



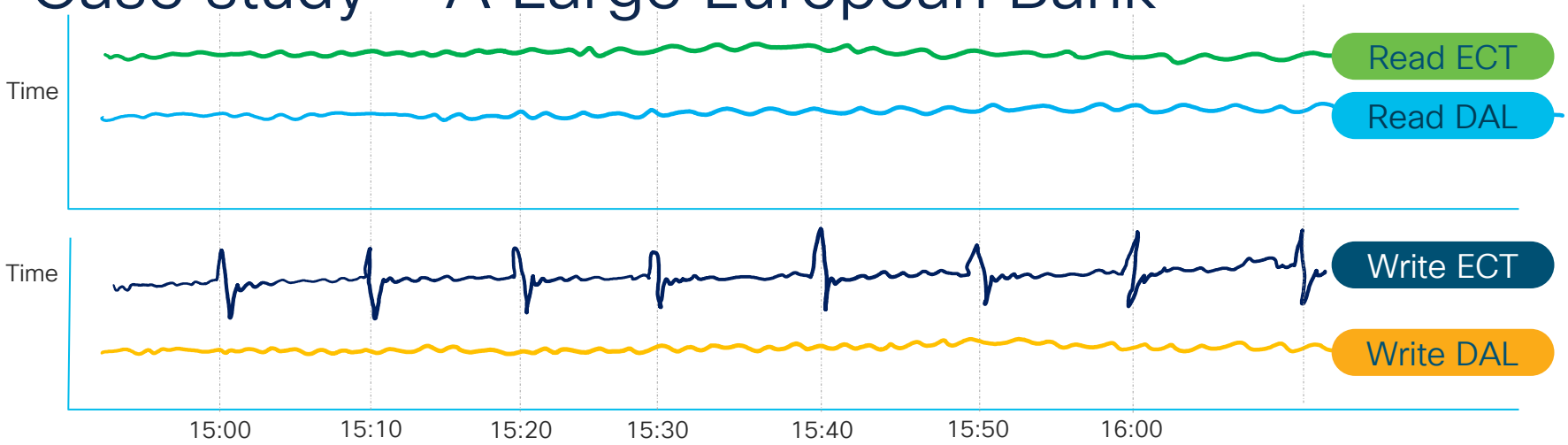

- Write ECT spikes followed by dips
  - May be the cause of application performance issues
- Frequency – every 10 minutes

# Case study – A Large European Bank




- Write ECT spike followed by dip. Frequency – every 10 minutes
- DAL is stable (no change)
  - Not a storage array issue

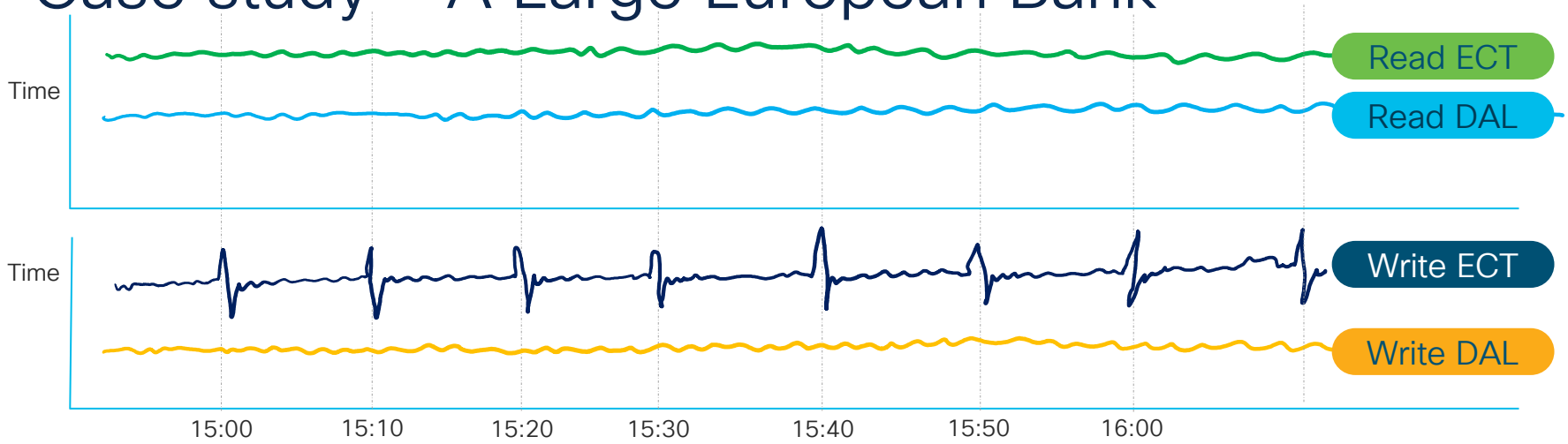
# Case study – A Large European Bank




3rd level pin-pointing

- Write ECT spike followed by dip. Frequency – every 10 minutes
- DAL is stable (no change). Not a storage array issue
- No changes in Read ECT and DAL. No fabric congestion observed.
  - No indication of fabric delay. Indication of delay within host.

# Case study – A Large European Bank

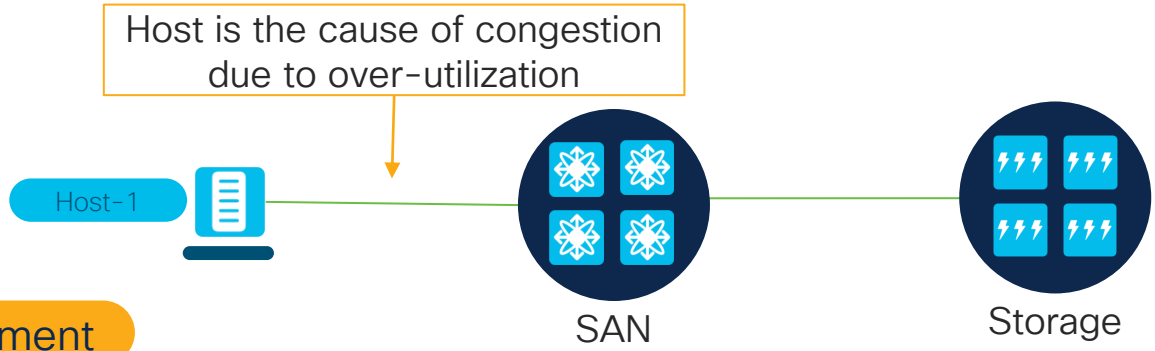



3rd level pin-pointing

- Write ECT spike followed by dip. Frequency – every 10 minutes
- DAL is stable (no change). Not a storage array issue
- No changes in Read ECT and DAL. Not a fabric issue
- Delay within host ➔ Resulted in detection of an unpatched Oracle app on host

# Culprit VM - Congestion due to Over-utilization

## Case Study – Many customers use this approach today

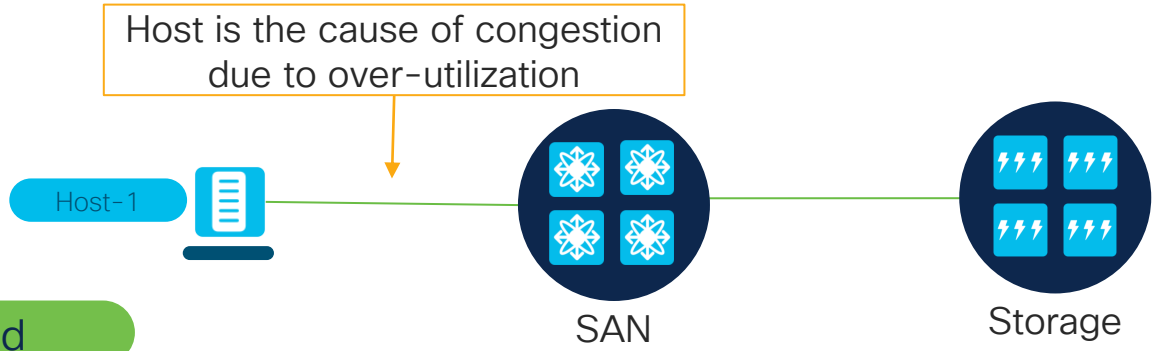



**Problem statement**

- Host-1 is virtualized. It uses LUN/Namespace/volumes from many storage arrays that are connected via SAN.
- Host-1 is the cause of congestion due to over-utilization
- Goal – Which VM and volumes are the top contributors to high link utilization?

# Culprit VM - Congestion due to Over-utilization

## Case Study – Many customers use this approach today

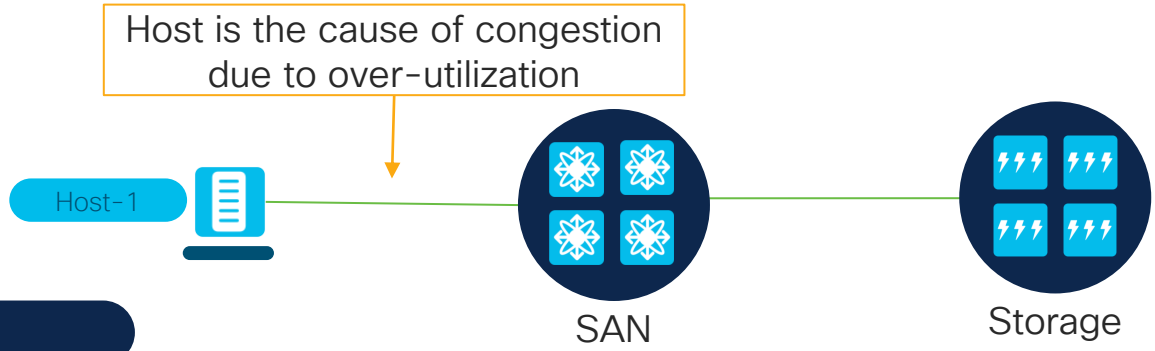



**Background**

- Network link utilization depends on I/O throughput
- Find I/O throughput using SAN Analytics at flow granularity
  - VM-I-T-L: If VE ID is supported (VE = Virtual Entity (Container or VM))
  - ITL: If VE ID is not supported (Most deployments)
- Traffic towards Hosts (Initiators) is mostly read I/O throughput, whereas traffic towards storage arrays (targets) is mostly write I/O throughput

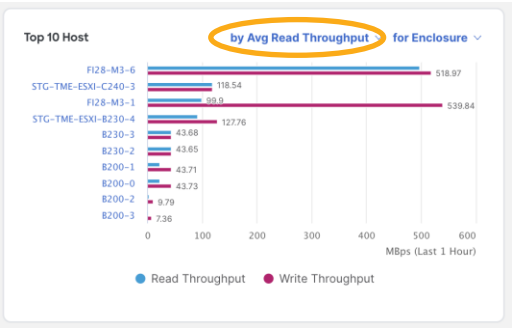
# Culprit VM - Congestion due to Over-utilization

## Case Study – Many customers use this approach today




**Solution**

- Use SAN Analytics to find
  - Storage arrays and storage ports that send most traffic to Host-1
  - LUN/Namespace/volume that send most traffic to Host-1
- Then, use vCenter to find the VM that's using that LUN/Namespace/volume
- Next steps: Move the VM to another host or add more HBA to Host-1 or increase the speed of Host-1 link, etc.




Graph Table

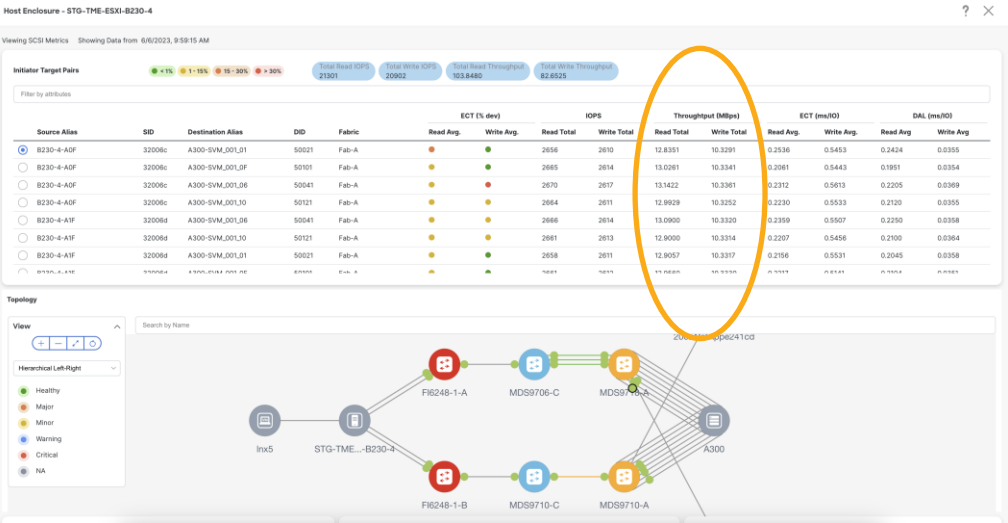
Filter by attributes

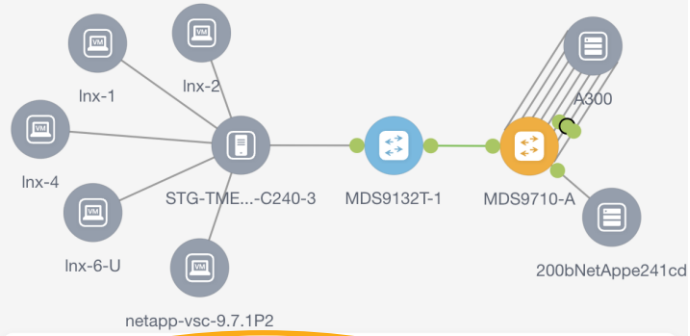
Download

| Initiator Enc | Initiator | Target Enc | Target | LUN | Switch IP Address | Port | Timestamp | Read Throughput (MB/s) | Write Throughput (MB/s) |
|---|---|---|---|---|---|---|---|---|---|
| STG-TME-ESXI-B230-4 | B230-4-A1F | A300 | A300-SVM_001_0F | 0018-0000-0000-0000 | 172.22.163.20 | fc1/35 | 2023-06-6 10:00:00 | 3.2164 | 3.2198 |
| STG-TME-ESXI-B230-4 | B230-4-A1F | A300 | A300-SVM_001_01 | 001c-0000-0000-0000 | 172.22.163.20 | fc1/33 | 2023-06-6 10:00:00 | 3.215 | 3.2337 |
| STG-TME-ESXI-B230-4 | B230-4-A0F | A300 | A300-SVM_001_0F | 001d-0000-0000-0000 | 172.22.163.20 | fc1/35 | 2023-06-6 09:55:00 | 0.0767 | 0.0005 |
| STG-TME-ESXI-B230-4 | B230-4-A1F | A300 | A300-SVM_001_0F | 001d-0000-0000-0000 | 172.22.163.20 | fc1/35 | 2023-06-6 10:00:00 | 0.0766 | 0.0005 |
| STG-TME-ESXI-B230-4 | B230-4-A1F | A300 | A300-SVM_001_06 | 0006-0000-0000-0000 | 172.22.163.20 | fc1/34 | 2023-06-6 09:55:00 | 0.0752 | 0.0005 |
| STG-TME-ESXI-B230-4 | B230-4-A1F | A300 | A300-SVM_001_06 | 0004-0000-0000-0000 | 172.22.163.20 | fc1/34 | 2023-06-6 09:55:00 | 0.0745 | 0.0005 |

100 Rows

Page 1 of 1 1-100 of 100

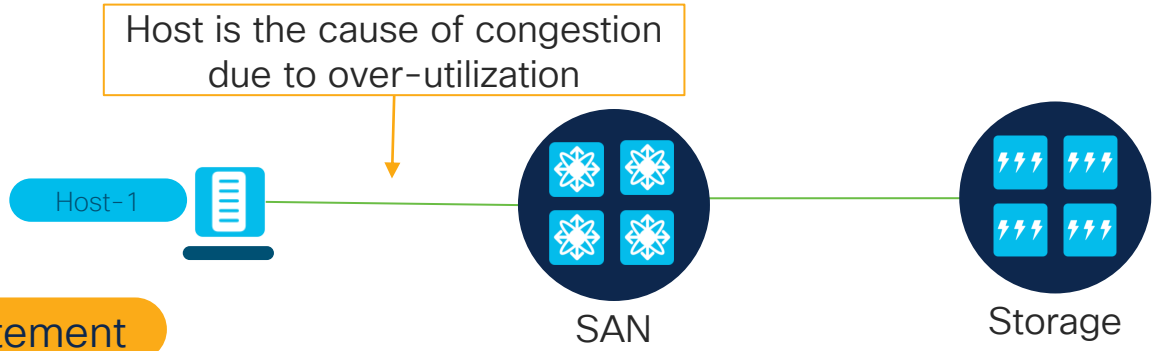



| VM Name | VM IP |
|---|---|
| lnx-2 | 172.22.163.205 |
| lnx-1 | 172.22.163.217 |

# MPIO issues - Congestion due to Over-utilization

## Case Study - A university in the mid-west

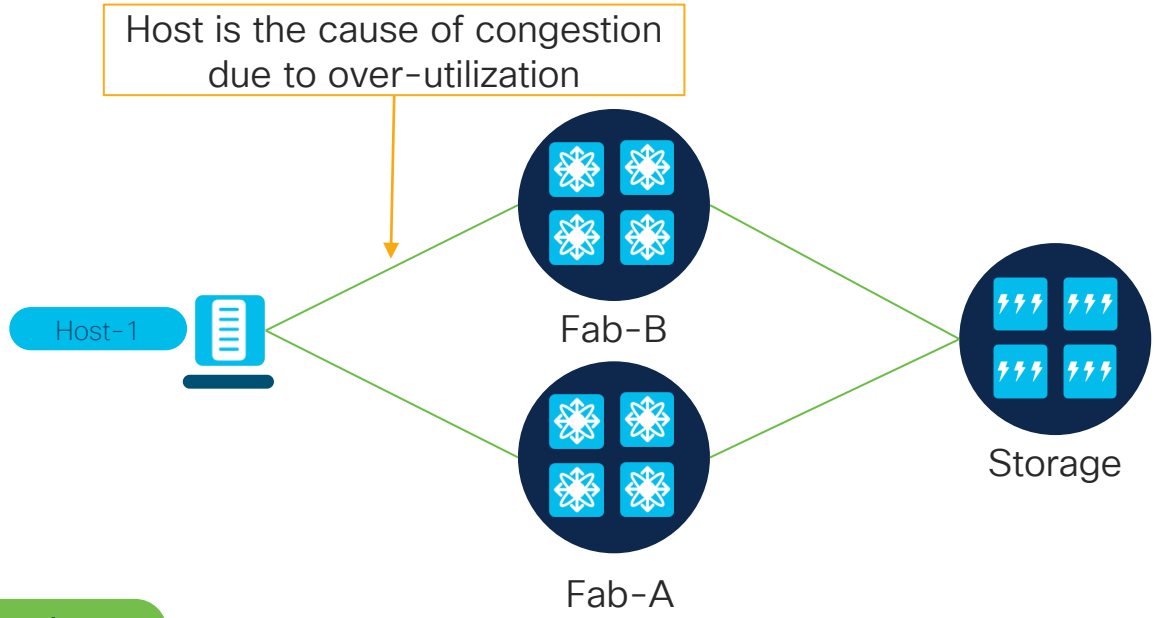



**Problem statement**

- Host-1 is the cause of congestion due to over-utilization
- Goal - Find the root cause and solve the problem

# MPIO issues - Congestion due to Over-utilization

## Case Study - A university in the mid west

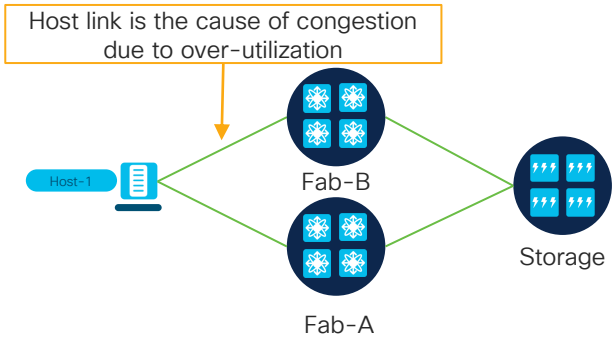


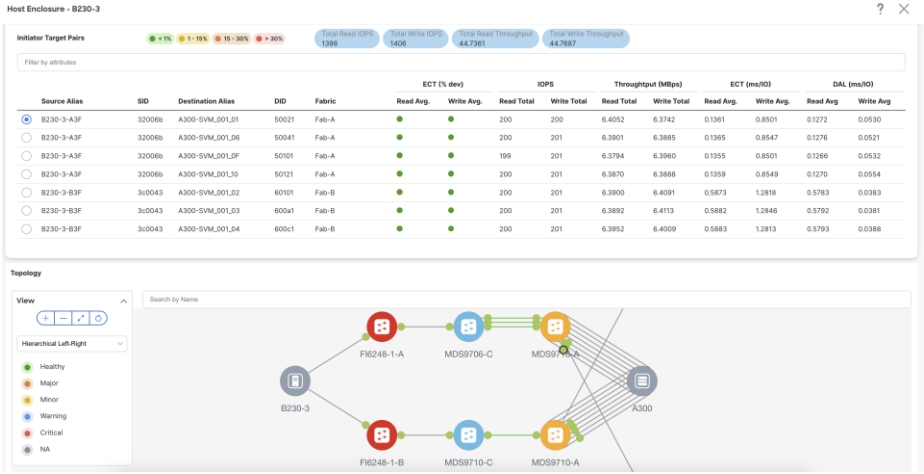

**Background**

- Hosts connect to the storage arrays via two redundant SAN (Fab-B and Fab-A)

# MPIO issues – Congestion due to Over-utilization

## Case Study – A university in the mid west




## Solution

- Use SAN Insights to find I/O throughput per path
- I/O throughput on Fab-B is much higher than Fab-A – Indicates incorrect MPIO config
- Solution – After changing MPIO config, I/O throughput on Fab-A and Fab-B is uniform, no over-utilization of a single link

# Predicting SAN Congestion

## Case Study – A trading company selectively upgraded SAN in phases using predictive capabilities of SAN Insights

**Problem statement**

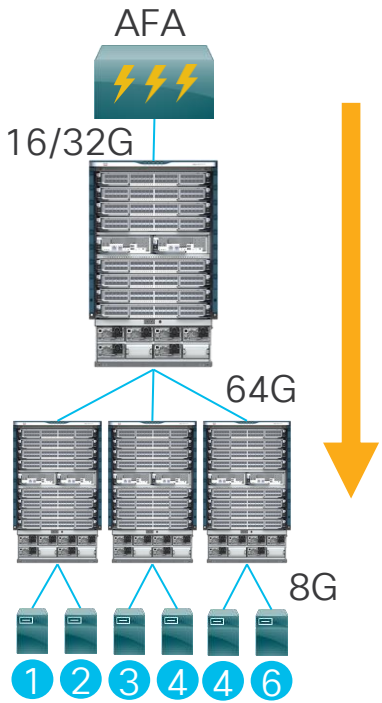
- Large SAN environment with thousands of ports per fabric. Many fabrics.
- Upgraded to all-flash storage, without upgrading the host speed at the same time
- Observed increased occurrences of congestion due to over-utilization of host links
- Aware that the ultimate solution was to upgrade end-to-end connectivity
- But resources weren't enough for an overnight upgrade
- Goal – Among thousands, which server to upgrade first?

# Predicting SAN Congestion

## Case Study – A trading company selectively upgraded SAN in phases using predictive capabilities of SAN Insights

Background

- In this example
  - 1 storage port connected at 32GFC speed
  - 6 hosts connected at 8GFC speed
- Question: Which host-links are more likely to get over-utilized? One, All, Few None?
- Host-links with larger I/O size are more likely to get over-utilized
- Did you know? – A host with 1% egress link utilization can cause 100% ingress link utilization?
  - Depends how large the I/O size is
  - SAN Analytics shows I/O size at I, T, IT, ITL flow granularity



# Predicting SAN Congestion

## Case Study – A trading company selectively upgraded SAN in phases using predictive capabilities of SAN Insights

### Solution

- The trading company enabled SAN Analytics on storage ports
- Collected the peak and average read and write I/O size for all hosts
  - Peaks are important.
- Made a sorted list and started upgrading the hosts first that have larger read I/O size
- The data collected by SAN Analytics gave them predictive insights for an informed upgrade plan.
  - Without SAN Analytics they would have run into many more congestion issues

# Summary of Case Studies of SAN Analytics

- A European bank detected an unpatched application server
- Many organizations are able to pin-point the root cause of congestion to a VM
- A mid-west university was able to solve congestion because of MPIO mis-config
- A trading company predicted SAN congestion and planned an upgrade to not get affected

# Summary of Case Studies of SAN Analytics

- A European bank detected an unpatched application server
- Many organizations are able to pin-point the root cause of congestion to a VM
- A mid-west university was able to solve congestion because of MPIO mis-config
- A trading company predicted SAN congestion and planned an upgrade to not get affected

## Cisco SAN Analytics

Always-on visibility at Scale

100K flows monitored by MDS 9700 at microsecond granularity

1 Million flows monitored by NDFC SAN Insights

# Fill out your session surveys!

Attendees who fill out a minimum of four session surveys and the overall event survey will get **Cisco Live-branded socks** (while supplies last)!

Attendees will also earn 100 points in the **Cisco Live Challenge** for every survey completed.

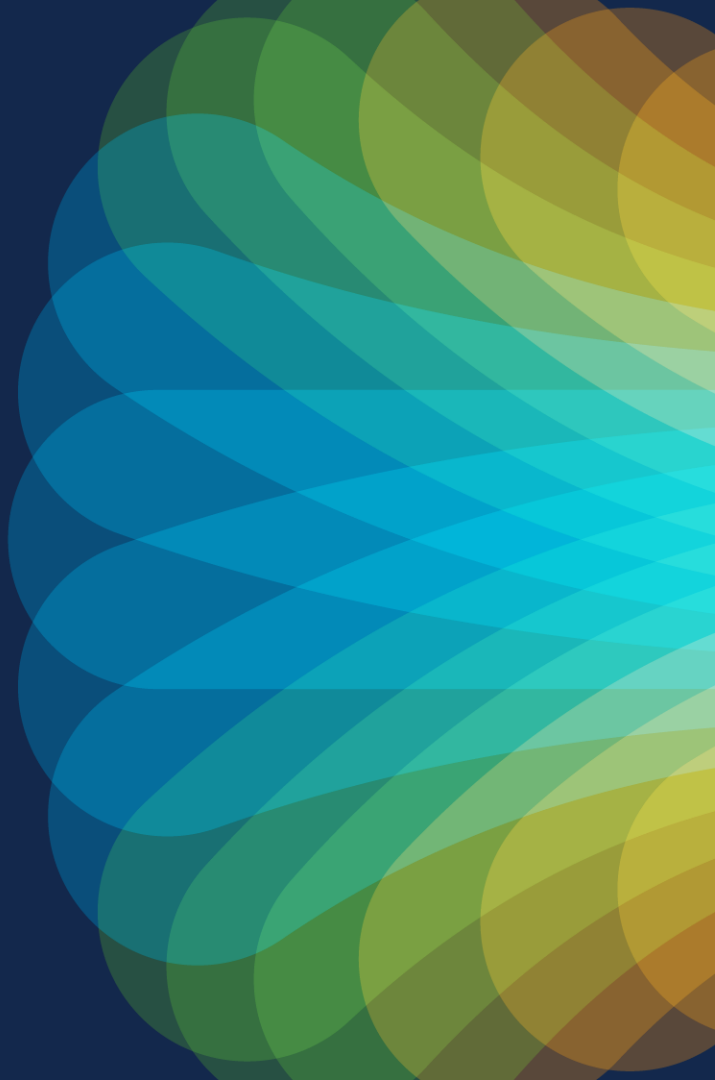

**These points** help you get on the leaderboard and increase your chances of winning daily and grand prizes

# Continue your education

- Visit the Cisco Showcase for related demos
- Book your one-on-one Meet the Engineer meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

The bridge to possible

# Thank you



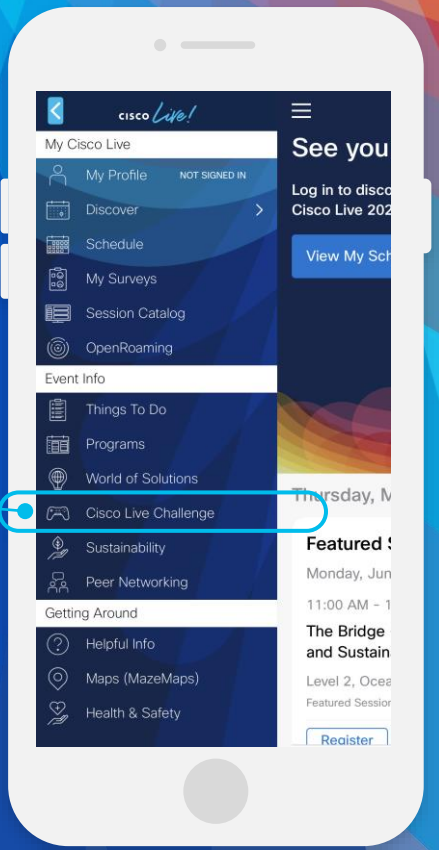
# Cisco Live
# **Challenge**

## Gamify your Cisco Live experience!

**Get points for attending this session!**

## How:

1. Open the Cisco Events App.
2. Click on 'Cisco Live Challenge' in the side menu.
3. Click on View Your Badges at the top.
4. Click the + at the bottom of the screen and scan the QR code:




cisco Live!
Let's go
#CiscoLive