nvm EXPRESS®

Discovery Automation for NVMe® IP-Based SANs

Sponsored by NVM Express organization, the owner of NVMe specifications

Speakers

Erik Smith
Distinguished Engineer

DELL Technologies

Curtis Ballard
Distinguished Technologist

Hewlett Packard
Enterprise

Agenda

NVMe-oF™ technology overview and discovery

Discovery types – Direct versus Centralized

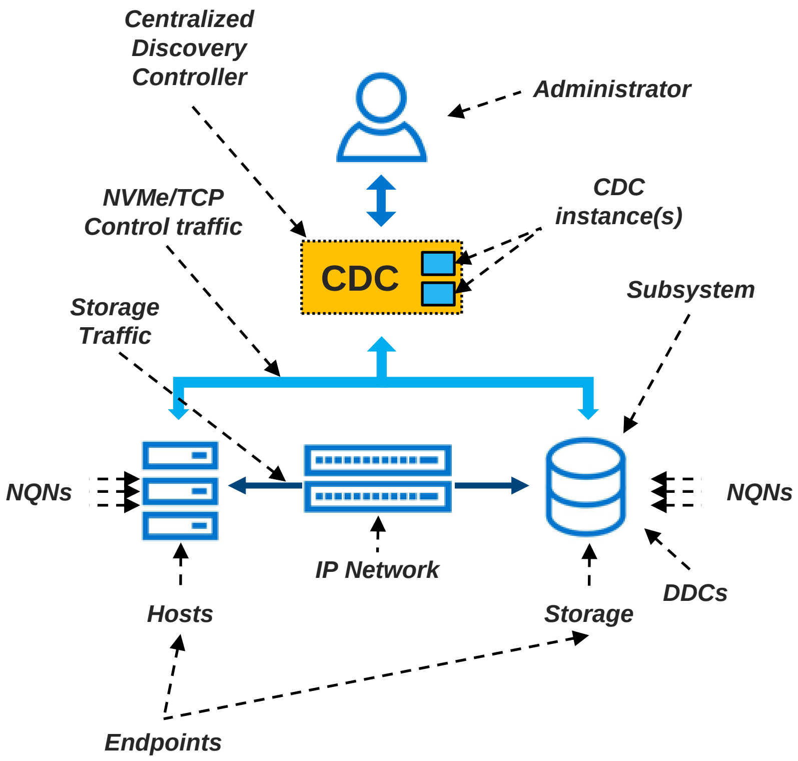
Centralized Discovery details

Centralized Discovery walk-through

NVMe-oF™ Discovery Built Around Discovery Controllers

- A Discovery controller is a single location that reports all known NVM subsystem interfaces
- Simplifies administration – A single Discovery controller IP **can** provide information about subsystem interfaces for multiple subsystems (arrays)
- The concept of a "referral" allows a Discovery controller to point to other Discovery controllers
- Common implementation today: every storage subsystem contains a discovery controller that only describes interfaces on that subsystem
- Until recently, there was no standardized method for Hosts, Subsystems or Discovery controllers to register information with a single Discovery controller (The Centralized Discovery section will cover this)

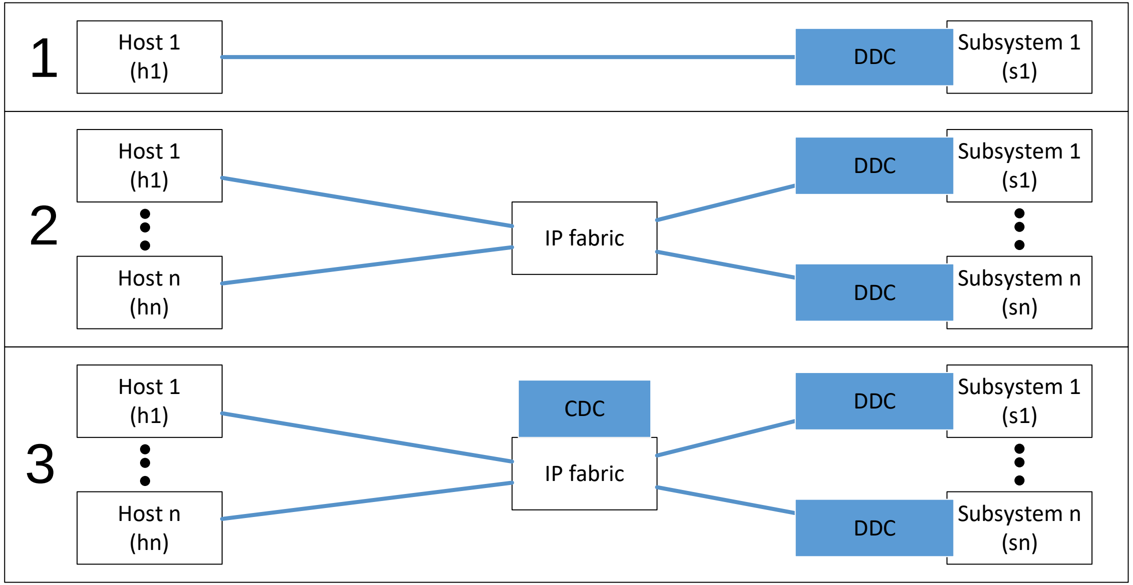
NVMe® IP-based SAN Terminology



- **Endpoints**: On hosts and storage systems
  - Identified by NVMe Qualified Name (NQN) and IP Address
- **IP Network:**
  - Most modern switches (e.g., 25GbE capable and above) will work.
- **Subsystem**: Storage array, analogous to SCSI target
  - Identified by NVMe Qualified Names (NQN). NQN has a similar function to FQN in FC, and IQN in iSCSI.
- **CDC**: Centralized Discovery Controller Instances
  - Each CDC instance provides a Discovery controller for Endpoints that are taking part in a particular NVMe IP-based SAN instance.
- **DDC**: Direct Discovery Controllers
  - An NVMe Discovery controller that resides on Subsystems
  - Hosts could connect directly to storage via the DDC, but would lose the advantages of Centralization

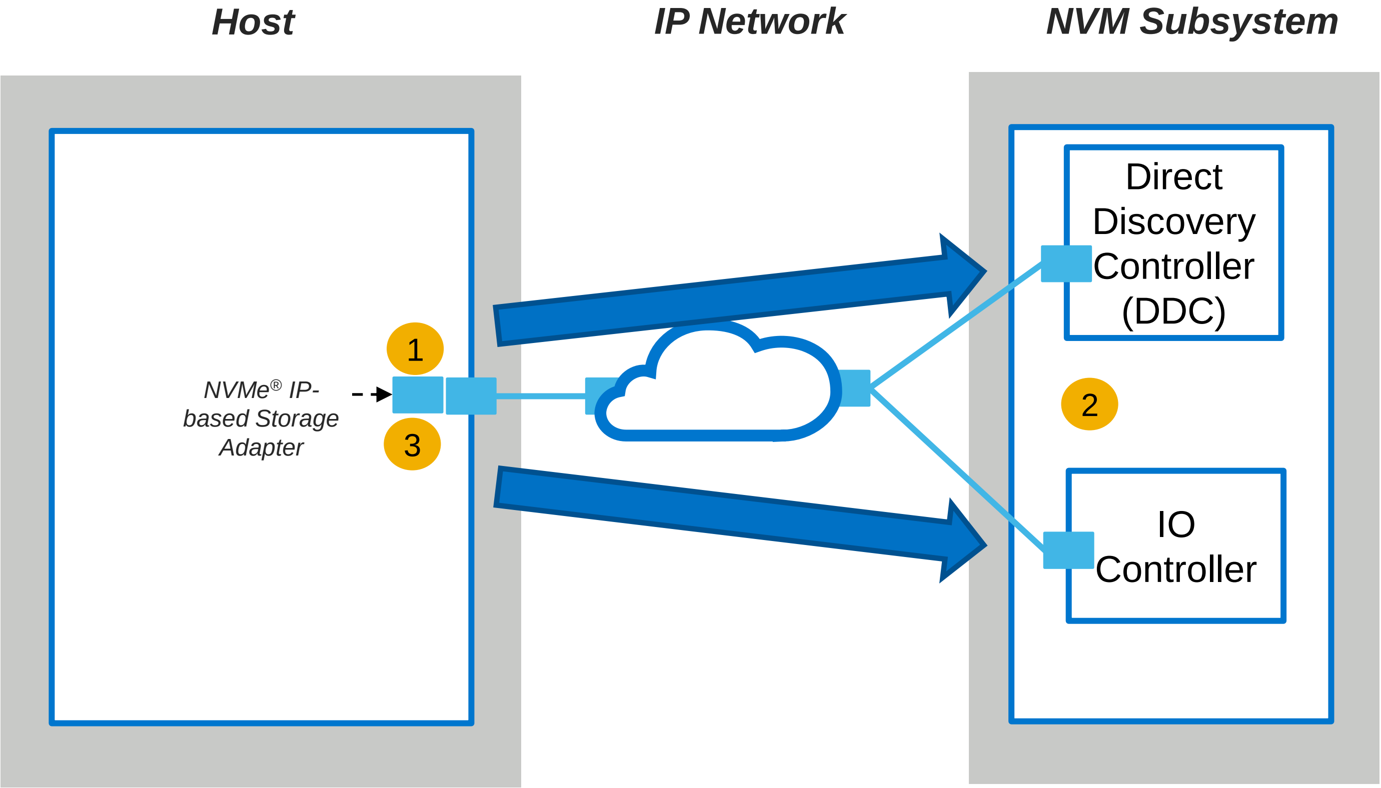
Deployment Types that Support Automated Discovery

1. **Physically connected**: Host and Storage subsystem are connected by a cable
2. **Direct Discovery**: Multiple Hosts and subsystems without a CDC in the network
3. **Centralized Discovery**: Multiple Hosts and subsystems with a CDC in the network



CDC (Centralized Discovery Controller) – A Discovery controller that supports registration and zoning. Typically runs stand-alone (as a VM) or embedded on a switch in the fabric.
DDC (Direct Discovery Controller) – A Discovery controller that is not a CDC. Typically associated with a storage subsystem

Configuration Steps with Direct Discovery (Existing)



Host
IP Network
NVM Subsystem
NVMe® IP-based Storage Adapter
1
3
Direct Discovery Controller (DDC)
2
IO Controller

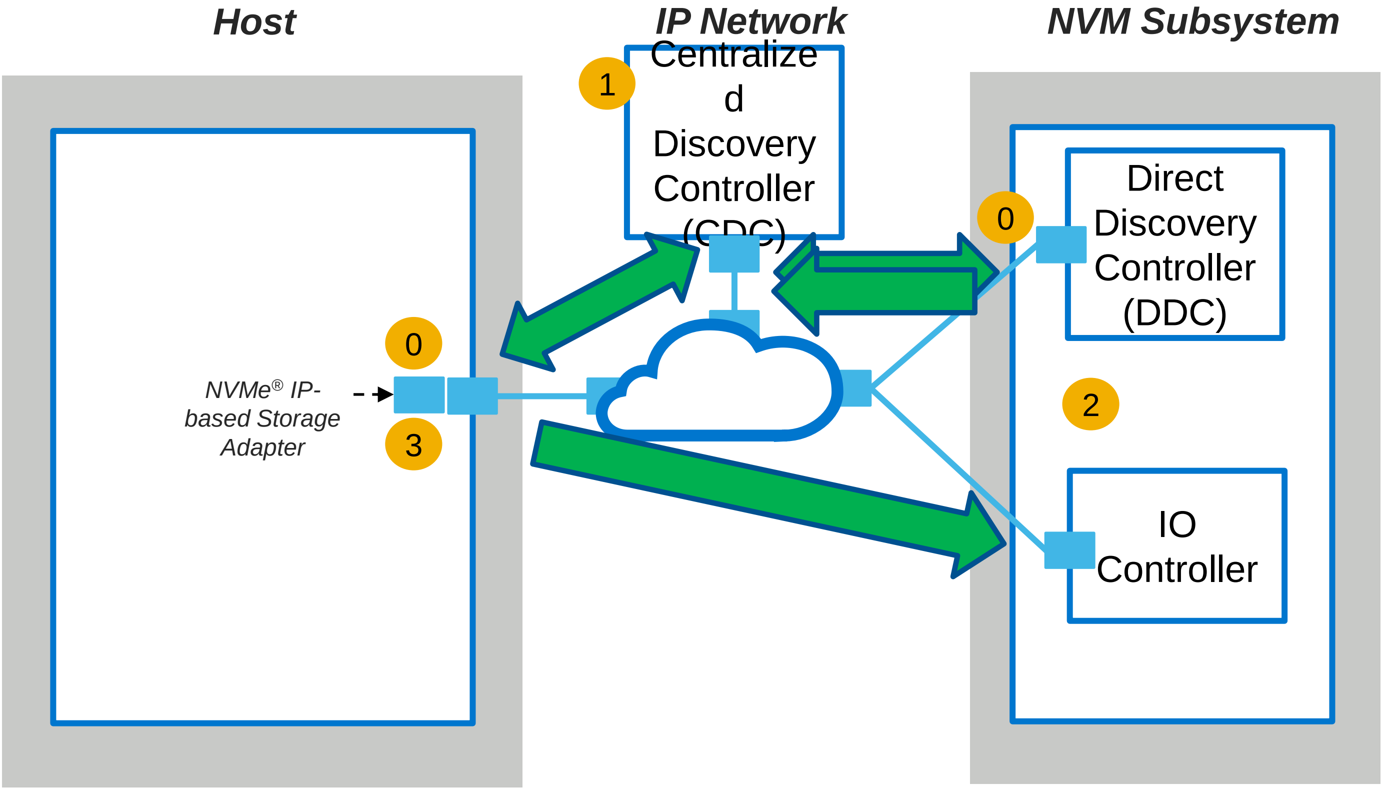
1 Host sends connect to Discovery Controller at IP Address supplied by admin

2 Storage admin provisions Namespaces (Storage) to the Host NQN

3 Host Admin uses nvme connect-all to Discover and connect to IO Controllers on that subsystem.

4 Repeat 1-3 on all Hosts for each subsystem

Configuration Steps with Centralized Discovery (New)



Host
IP Network
NVM Subsystem
1
Centralized Discovery Controller (CDC)
0
NVMe® IP-based Storage Adapter
3
0
Direct Discovery Controller (DDC)
2
IO Controller

0 Host and subsystems automatically discover the CDC, connect to it and Register Discovery info

1 Zoning performed on CDC (optional)

2 Storage admin provisions namespaces to the Host NQN. Storage may send zoning info to CDC

3 After zoning, Host receives AEN, uses get log page, and connects to each IO Controller

4 Repeat 1-2 for each Hosts on each subsystem

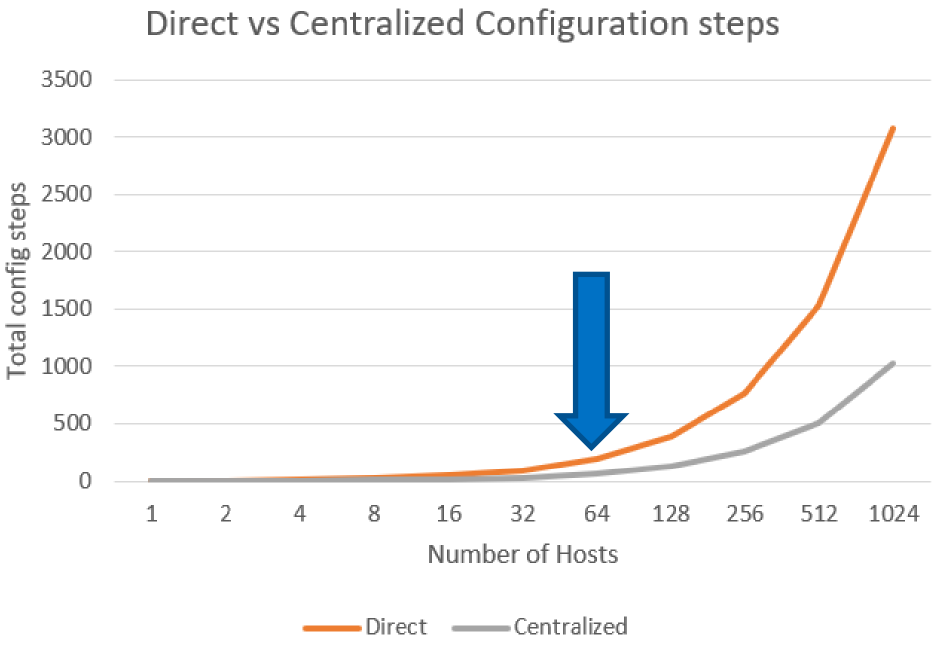
Direct vs Centralized Discovery at Scale

Direct Discovery config steps

1. **Host**: Determine subsystem Discovery controller IP -> connect
2. **Storage**: Provision storage
3. **Host**: Discover / connect all

Centralized Discovery config steps

1. **Host**: N/A
2. **CDC**: Configure Zoning (optional)
3. **Storage**: Provision storage



What the chart doesn’t show

1. **Direct becomes impractical @ >64 hosts**
2. **Direct requires interaction with each host every time a storage subsystem is added or removed.**
3. **Direct may lead to extended discovery time if many subsystem interfaces are present.**

Additional Points about Discovery Automation

Discovery Automation does not depend entirely upon a Centralized Discovery Controller (CDC).

Smaller scale environments can make use of mDNS (as described in TP-8009) to automatically discover NVMe® Discovery Controllers.

This approach does not allow for Centralized Control, and this means:

- Access control at the network is much more complicated/impractical
- Hosts will not be notified when a new storage subsystem is added to the environment

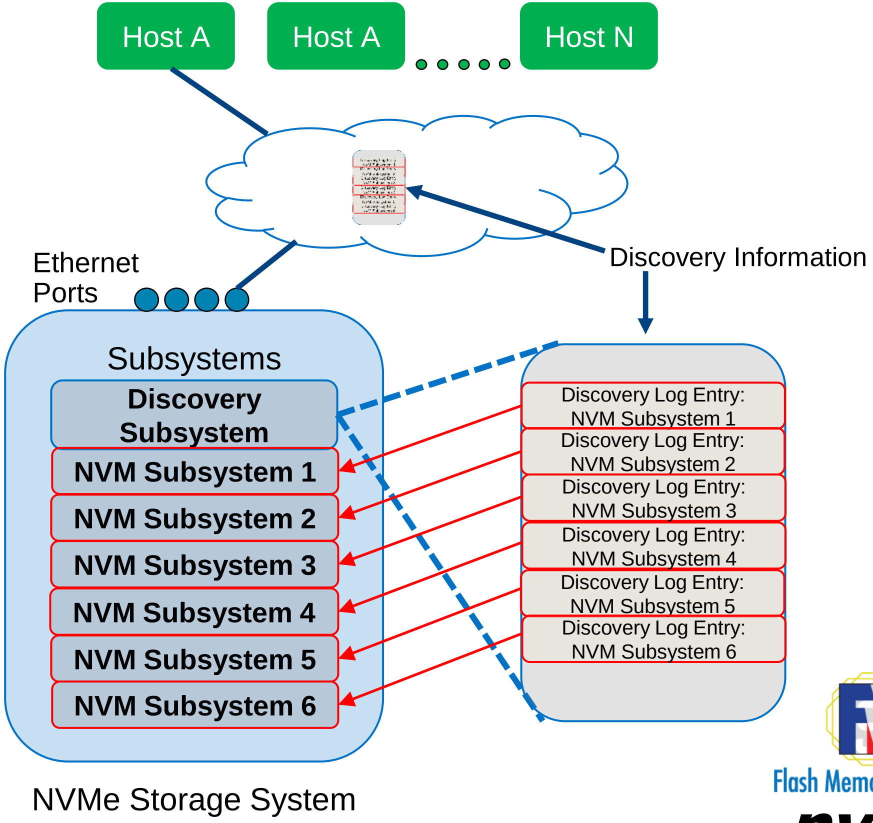
mDNS can become excessively chatty in larger configurations

- Especially when there are more than 1000 ports in a single broadcast domain

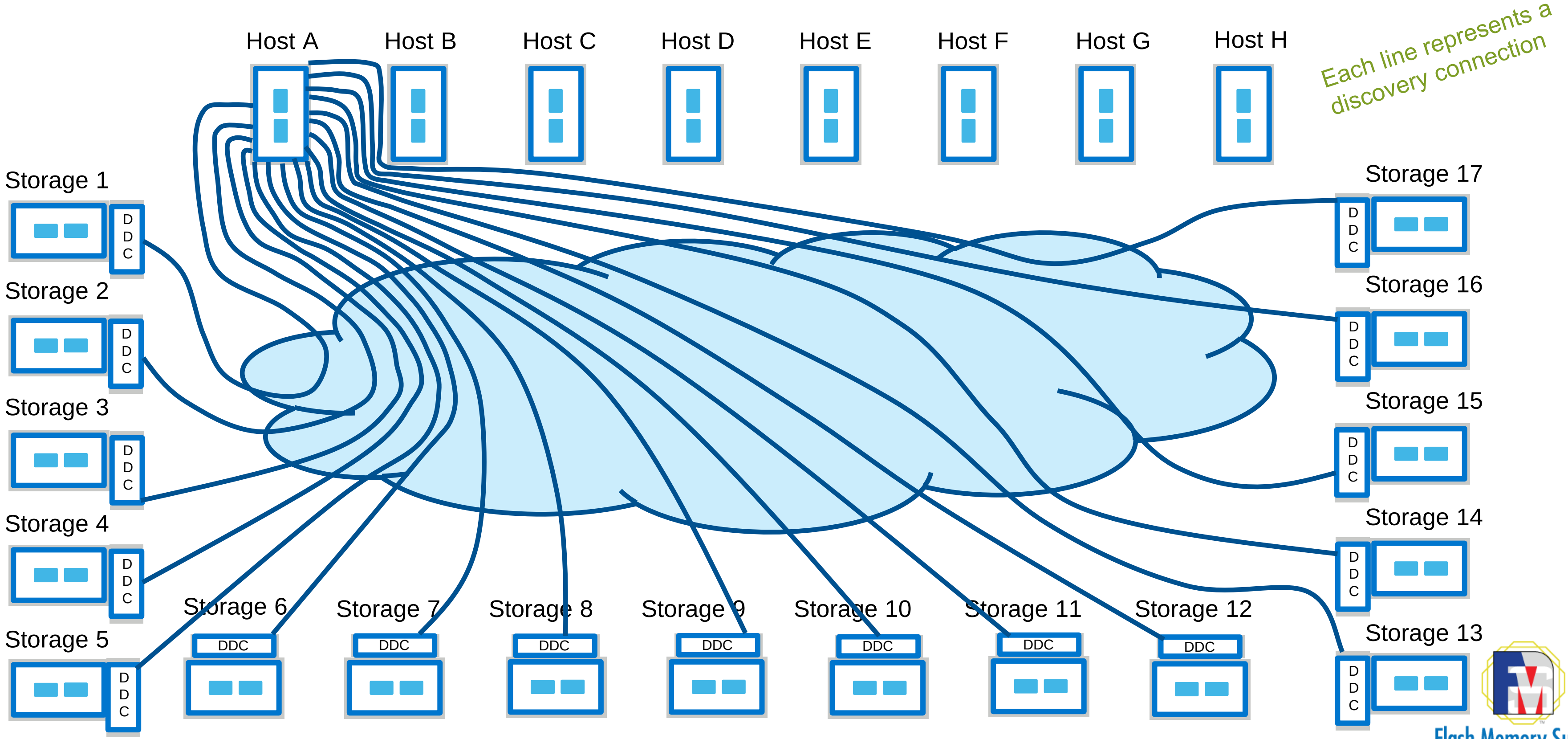
Centralized Discovery

Discovery Information

- Administrator configures path to a Discovery Subsystem
- Host connects to Discovery Controller in the Discovery Subsystem
- Discovery information is reported in discovery log page entries


Host A
Host A
Host N
Ethernet Ports
Discovery Information
Subsystems
Discovery Subsystem
NVM Subsystem 1
NVM Subsystem 2
NVM Subsystem 3
NVM Subsystem 4
NVM Subsystem 5
NVM Subsystem 6
Discovery Log Entry: NVM Subsystem 1
Discovery Log Entry: NVM Subsystem 2
Discovery Log Entry: NVM Subsystem 3
Discovery Log Entry: NVM Subsystem 4
Discovery Log Entry: NVM Subsystem 5
Discovery Log Entry: NVM Subsystem 6
NVMe Storage System

The Scaling Problem



Each line represents a discovery connection

Each “Storage” represents a storage system presenting a Discovery Controller for one or more NVM Subsystems

The Solution: Centralized Discovery Controllers

NVM Express® standard model for cooperating Discovery Controllers

- Single Fabric entity that aggregates NVMe® discovery information from cooperating Discovery Controllers
- Standard API for sharing discovery information between a Centralized Discovery Controller (CDC), Hosts, and NVM Express storage system Discovery Controllers (Direct Discovery Controllers, DDCs)
- Single location for storage systems to register discovery information
- Single location for Hosts to query discovery information

Additional new functionality for both CDCs and DDCs

- Mechanism for Hosts to register Host information into Discovery Controllers
- Mechanism for sharing connectivity rules, “Fabric Zoning”, information

Discovery Changes for Hosts

Clean evolution of existing Host discovery

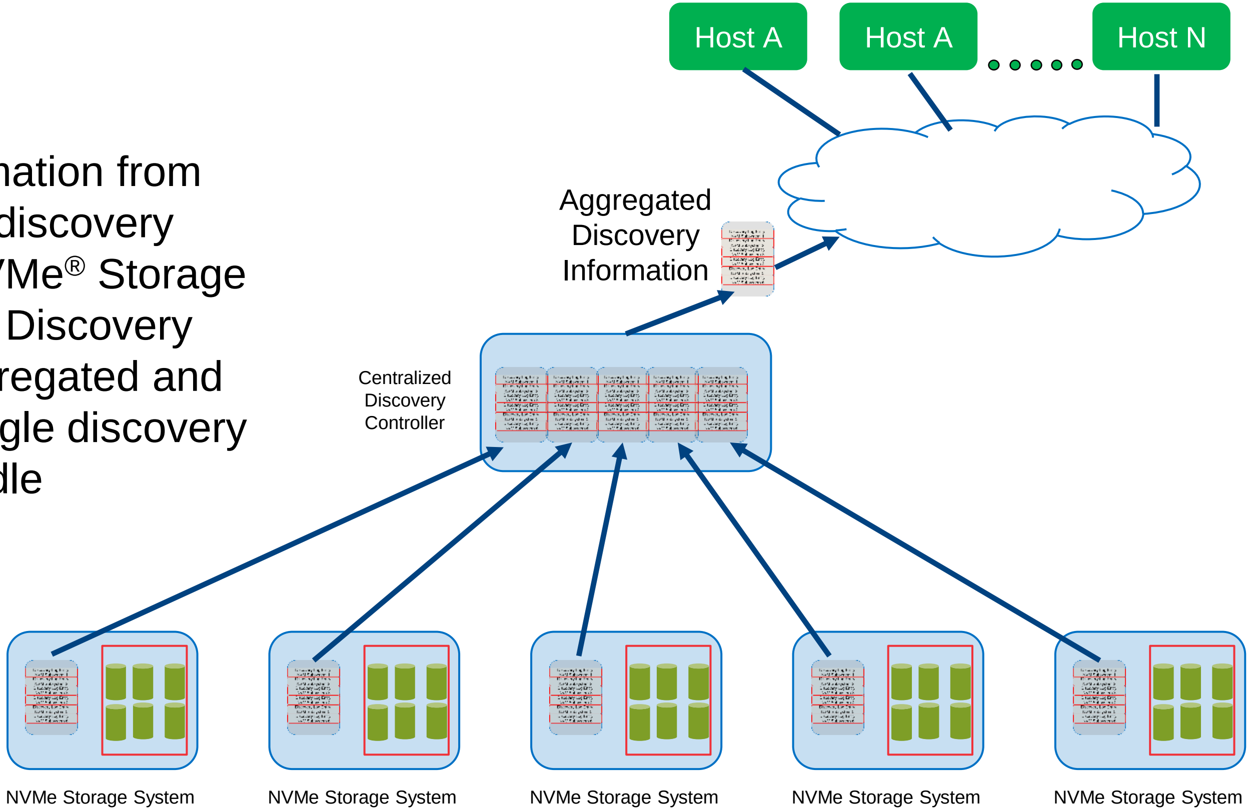
CDC reports available NVM Subsystems

- Same format Discovery log pages as today
- Same Host specific accessible NVM Subsystems filtering as today is allowed

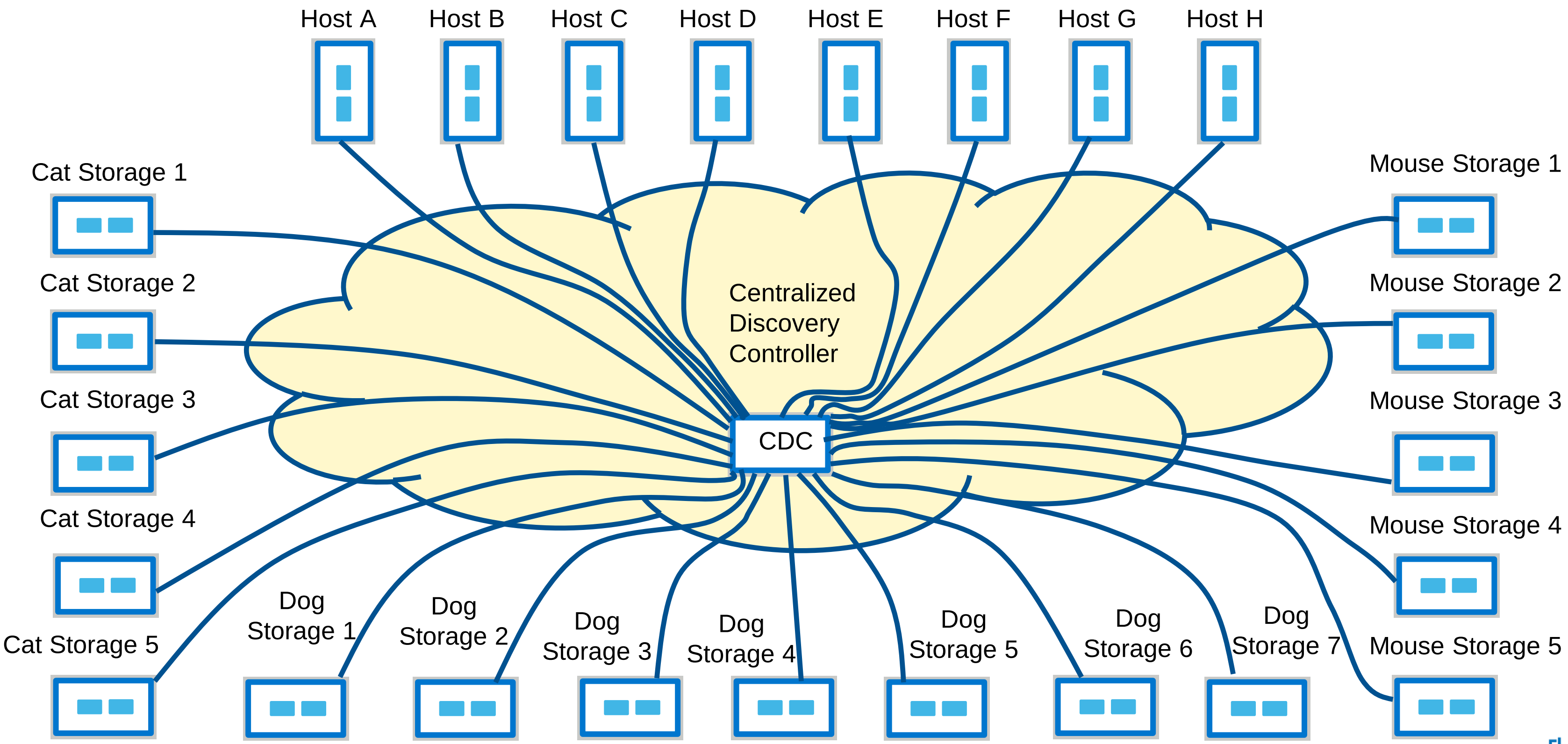
Only completely new functionality is Host registering with the storage fabric

Same Discovery Information: Centralized

Discovery Information from each of several discovery controllers in NVMe® Storage Systems (Direct Discovery Controllers) aggregated and reported in a single discovery information bundle



Playing Nicely Together



Each “Storage” represents a cooperating storage system sharing a Centralized Discovery Controller

New Capability: Host Registrations

Push Registrations

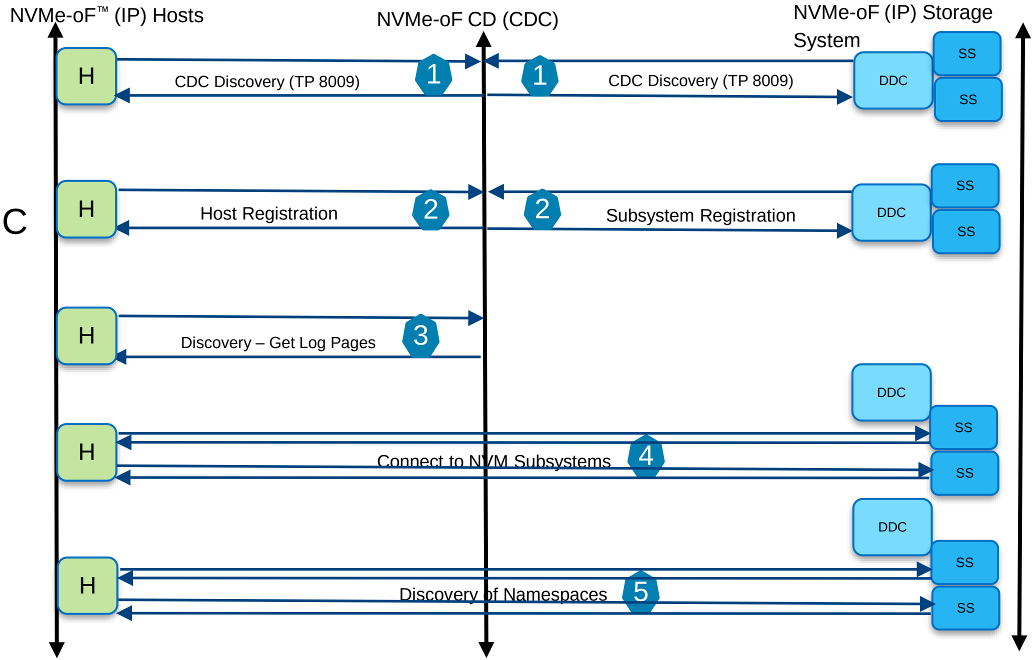
- Proactive registration from a Host or a Storage System into a CDC
- Only registration model defined for Hosts
- For Direct Discovery Controllers (DDCs) this requires host functionality and ability to send commands to CDC
- NVMe-oF™ connections established with CDC
- Hosts and CDCs use same registration commands with slightly different data formats

Pull Registrations

- Only for Direct Discovery Controllers (DDCs) in storage systems
- Storage systems discover CDC and request Pull Registration
- CDC uses existing Get Log Page commands to read existing Discovery Log Pages
- DDC that supports CDC discovery reports full list of all available NVM Subsystems to CDC

Example Simple Centralized Discovery Sequence

1. Hosts and storage systems discover the CDC
2. Hosts and storage systems register with the CDC
3. Hosts read discovery information from CDC (accessible NVM subsystems)
4. Hosts connect to NVM Subsystems
5. Hosts discover namespaces
6. Go!



How Does the CDC Filter Responses by Host?

Today’s storage systems often implement access controls and the Discovery Controllers only report information about “accessible NVM Subsystems”

The CDC has to get the full list of all “available” NVM Subsystems

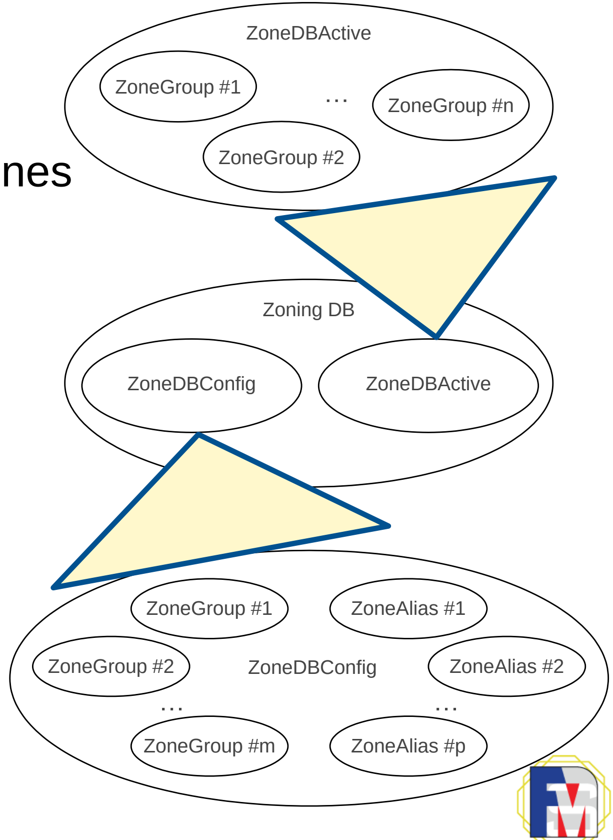
How does the CDC know which NVM Subsystems are “accessible” by which Hosts?

Answer: Fabric Zoning

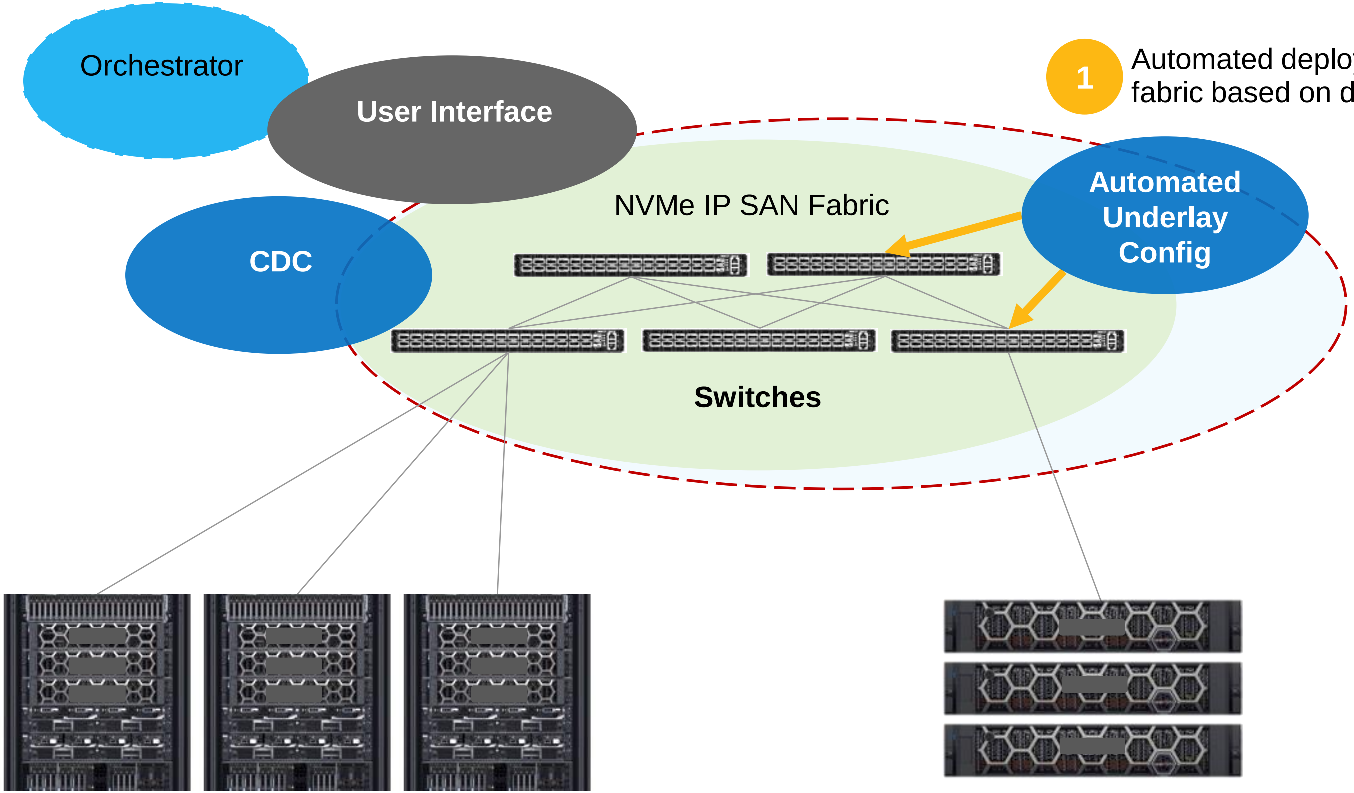
Fabric Zoning Quick Intro

- Zoning database in CDC stores configured and active Zones
  - Configured Zones is list of
    - All Fabric ZonesGroups; and
    - All Fabric ZoneAliases (a related set of Zone members)
  - Active Zones is list of ZoneGroups that are being enforced
- ZoneGroups contain Zone members
  - Hosts, NVM Subsystems, ZoneAliases
  - Member identification NQN, NQN/IP tuple, NQN/PortID tuple, etc.
- Admin commands defined for CDC and NVM Subsystem to share Fabric Zone information

- Multiple active ZoneGroups allowed



NVMe® IP SAN operations

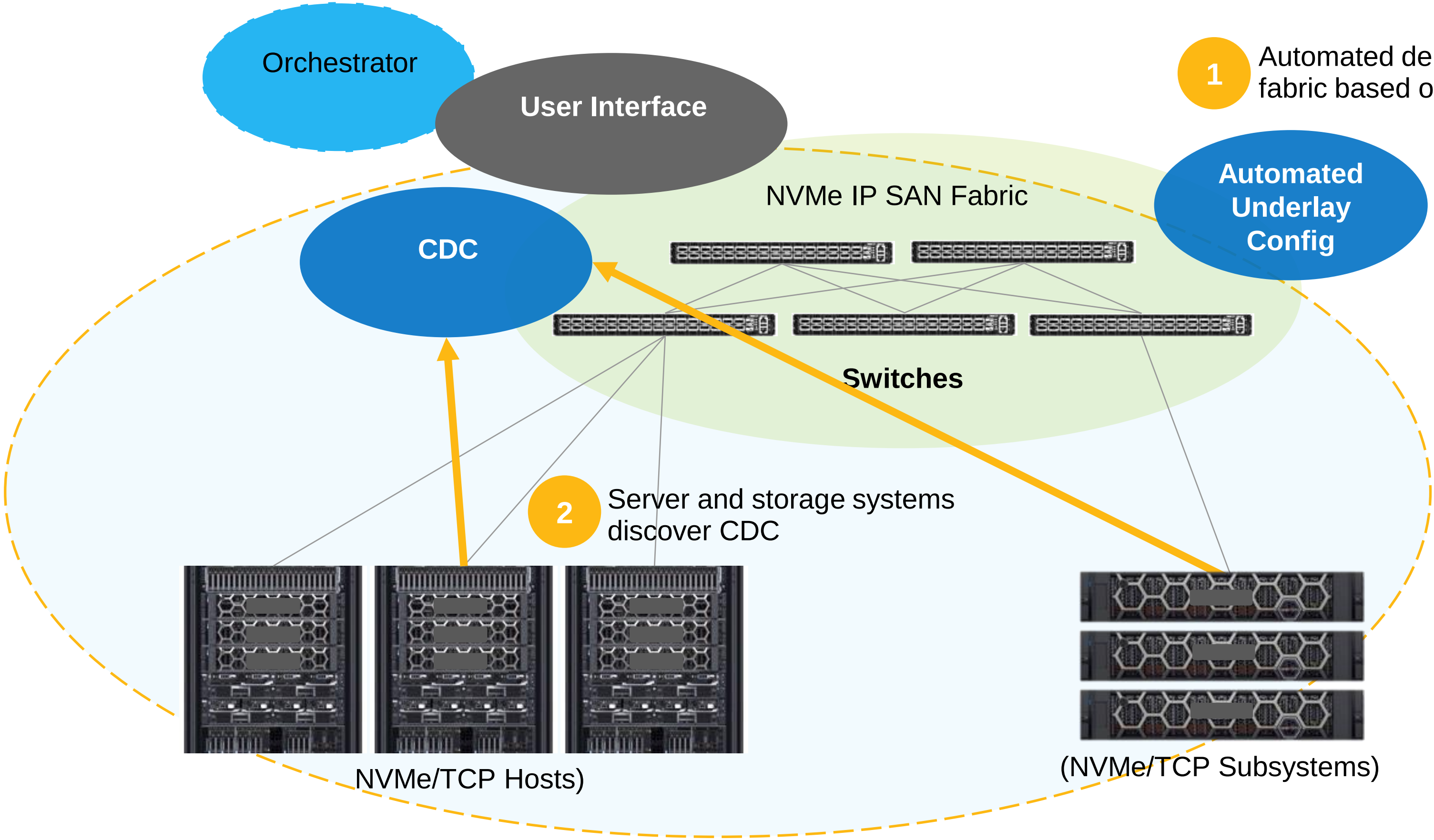


(NVMe/TCP Hosts)

(NVMe/TCP Subsystems)

1 Automated deployment of the fabric based on defined policies

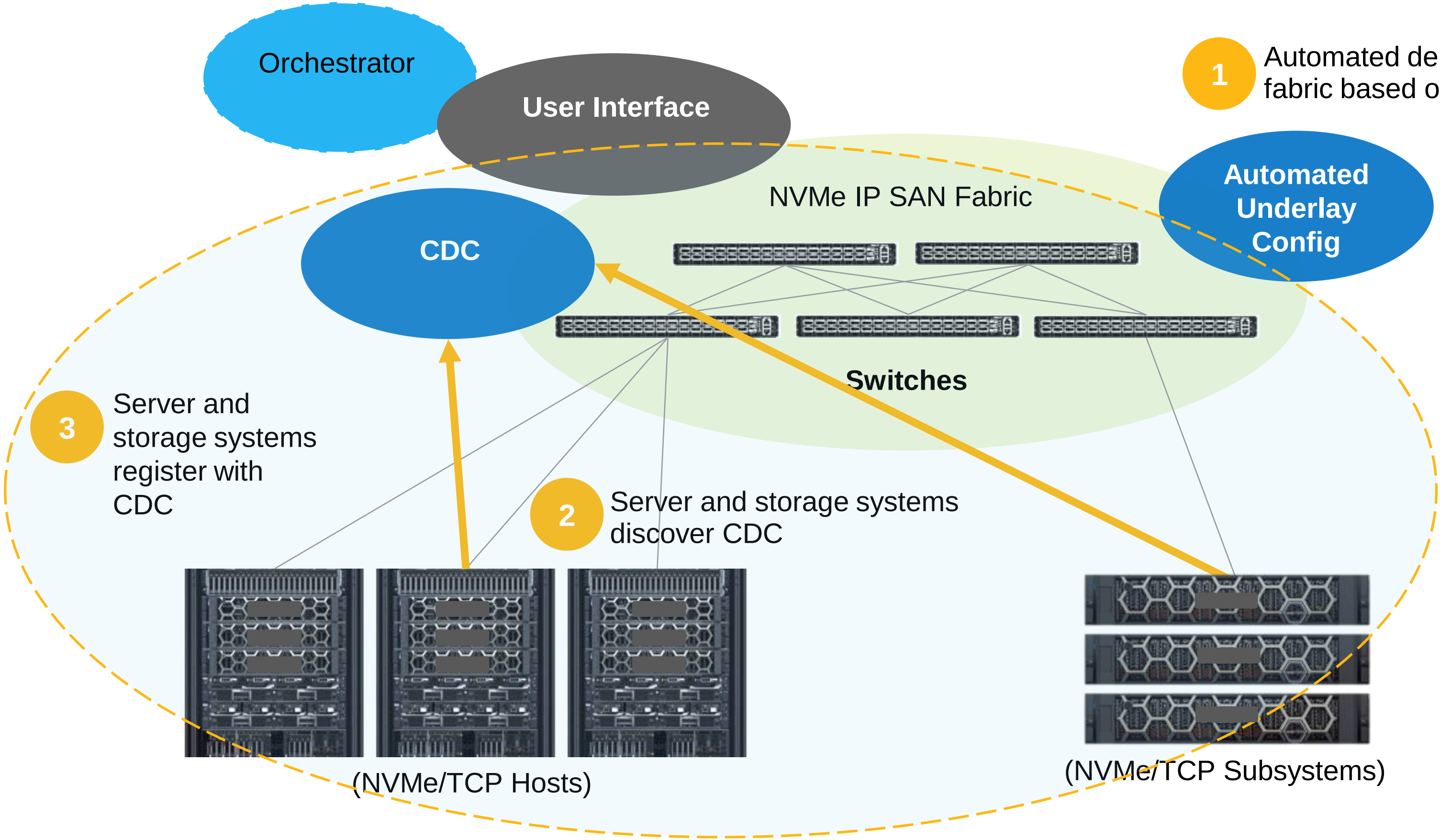
NVMe® IP SAN operations



1 Automated deployment of the fabric based on defined policies

2 Server and storage systems discover CDC

NVMe® IP SAN operations

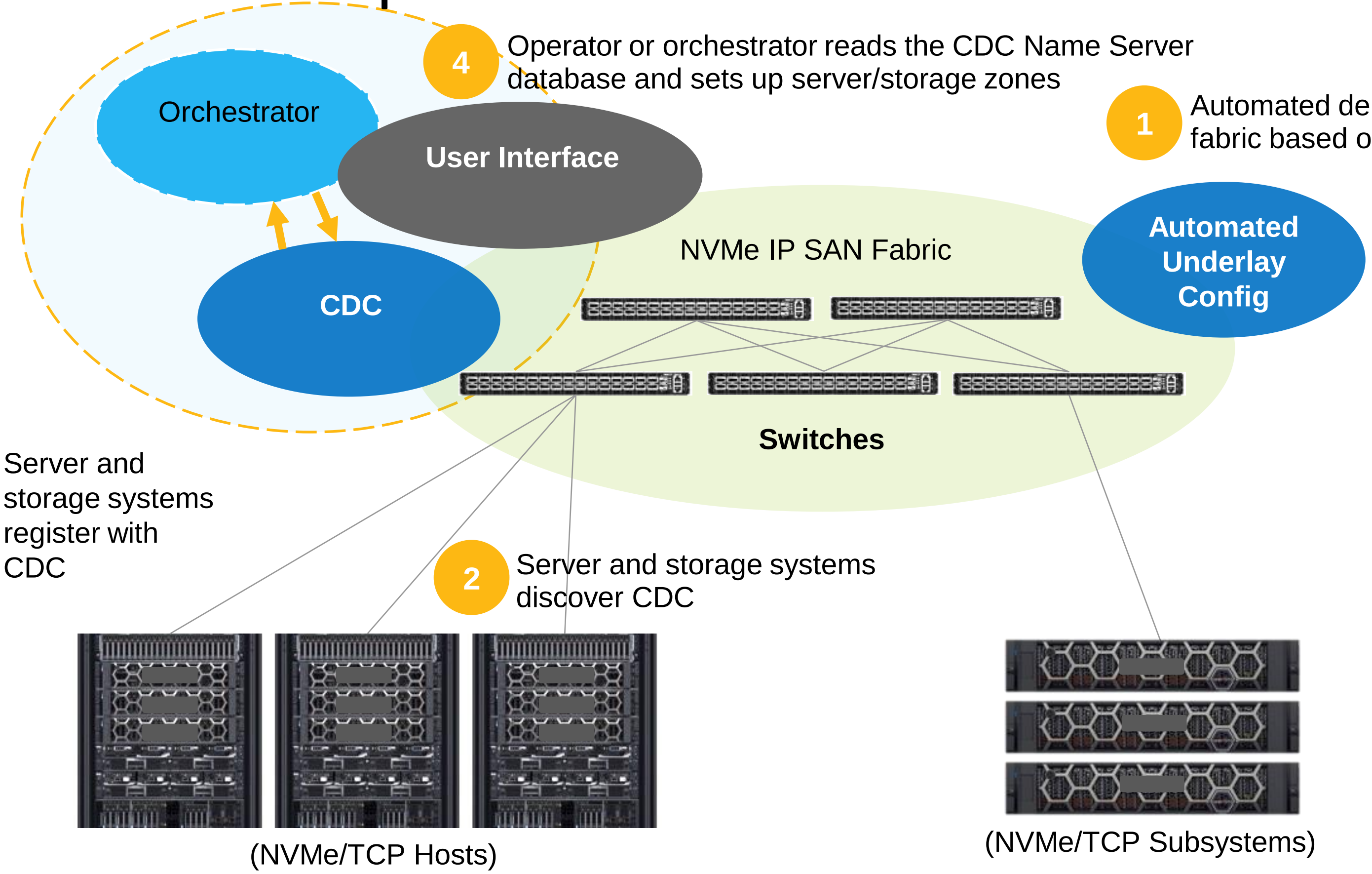


1 Automated deployment of the fabric based on defined policies

2 Server and storage systems discover CDC

3 Server and storage systems register with CDC

NVMe® IP SAN operations



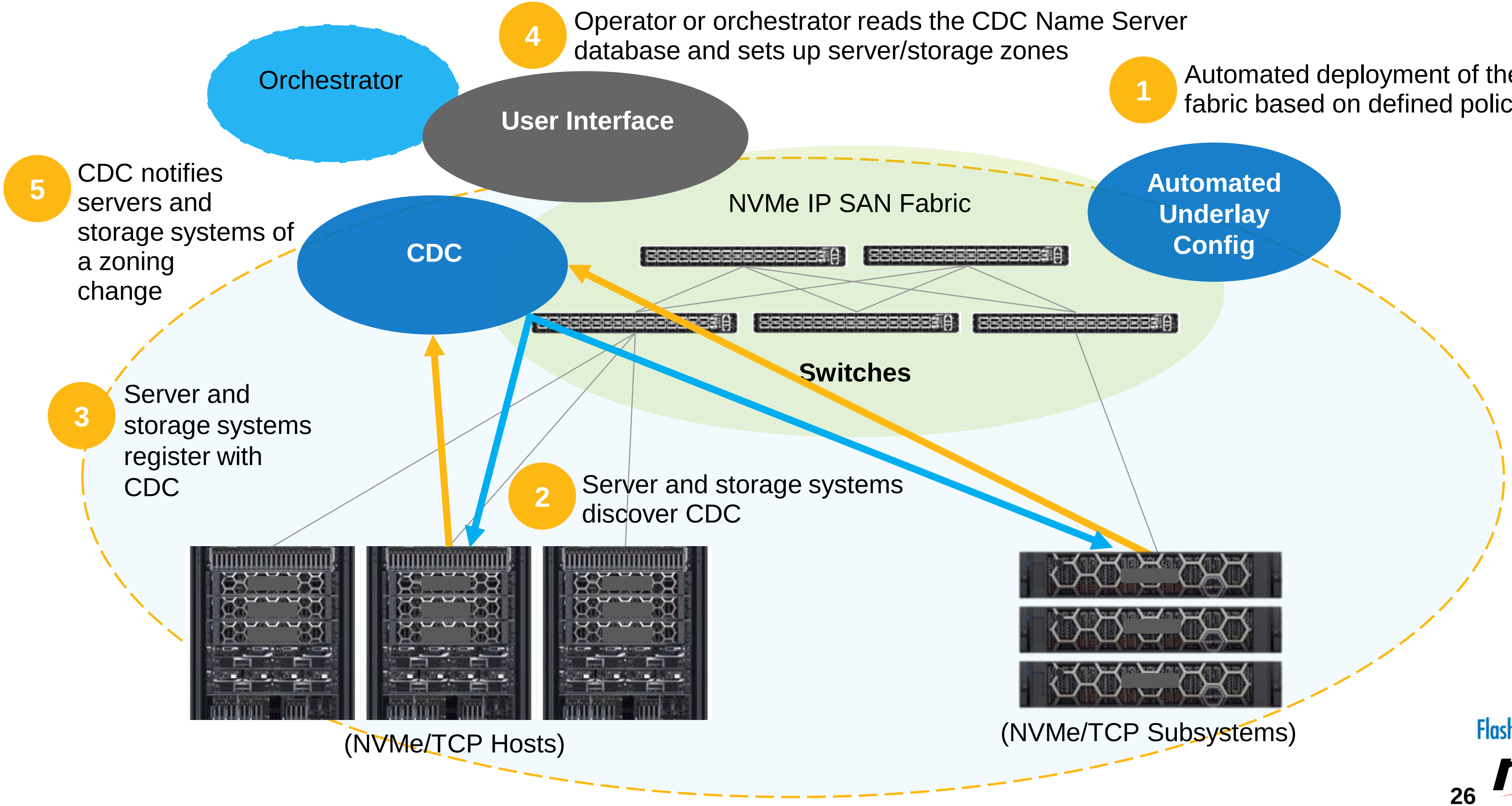
4 Operator or orchestrator reads the CDC Name Server database and sets up server/storage zones

1 Automated deployment of the fabric based on defined policies

3 Server and storage systems register with CDC

2 Server and storage systems discover CDC

NVMe® IP SAN operations



4 Operator or orchestrator reads the CDC Name Server database and sets up server/storage zones

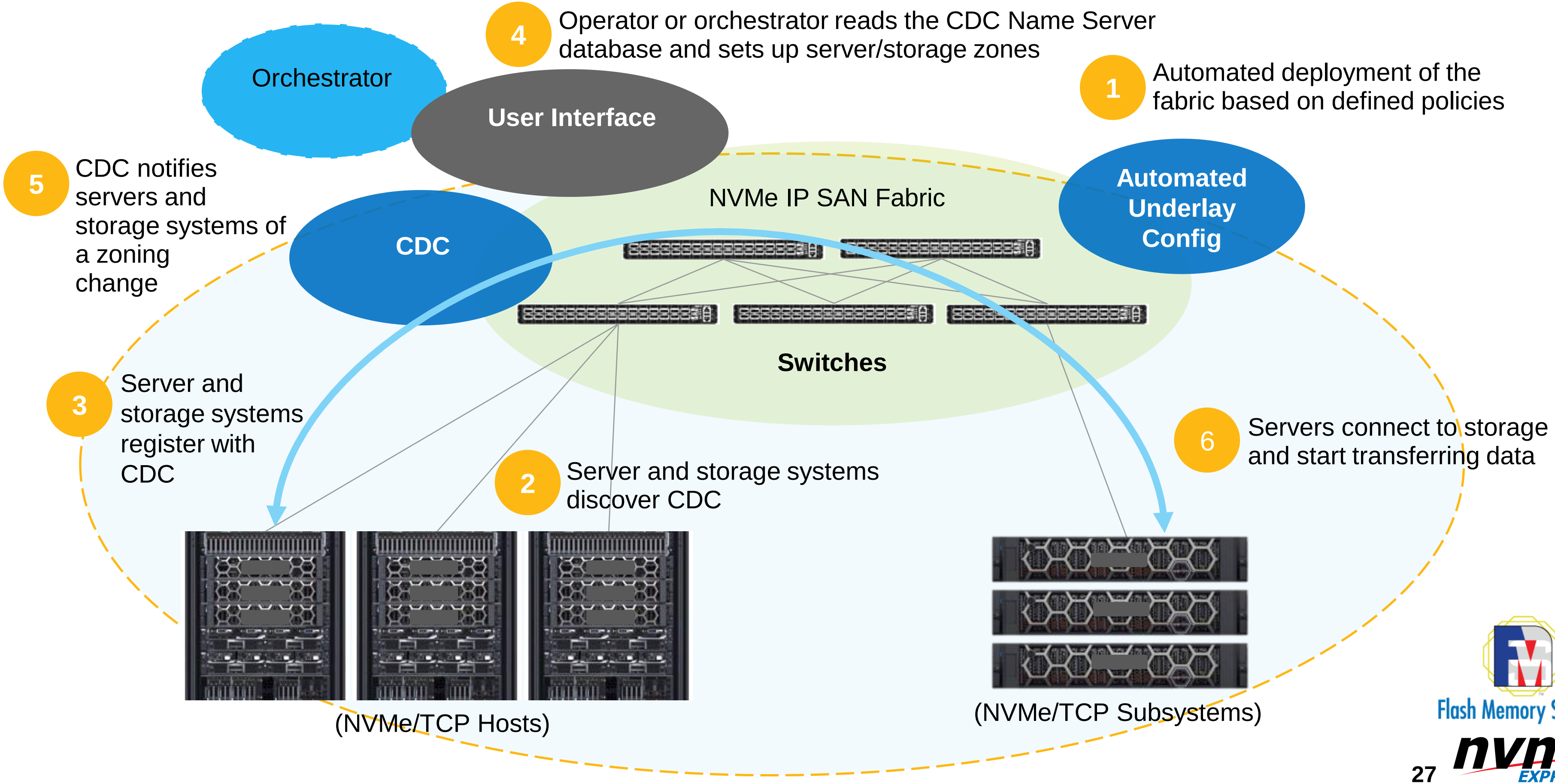
1 Automated deployment of the fabric based on defined policies

5 CDC notifies servers and storage systems of a zoning change

3 Server and storage systems register with CDC

2 Server and storage systems discover CDC

NVMe® IP SAN operations



4 Operator or orchestrator reads the CDC Name Server database and sets up server/storage zones

1 Automated deployment of the fabric based on defined policies

5 CDC notifies servers and storage systems of a zoning change

3 Server and storage systems register with CDC

2 Server and storage systems discover CDC

6 Servers connect to storage and start transferring data

Questions?

nvm EXPRESS®

Architected for Performance