

NVMe 2.x Architecture Training

Let MindShare bring “NVMe 2.x Architecture” to life for you

MindShare's NVMe (Non-Volatile Memory Express) 2.2 course begins with a high-level view of this storage architecture to give the big-picture context for hardware and software interactions. We drill down into details for each aspect of NVMe. Next, we review all the changes and added features to the more recent spec revisions. Since NVMe-oF (NVMe over Fabrics) was combined with the original NVMe definition in the 2.0 spec revision, we include a thorough coverage of that as well. Finally, because many implementations of NVMe will use PCI Express (PCIe), a brief overview of that interface is included.

You Will Learn:

- PCIe architecture brief overview
- Basics of the NVMe Host Controller Interface model
- The steps for device initialization
- How command queues are set up and managed
- The various administrative and I/O commands
- How host software
 - Informs the controller of new commands to execute
 - Learns that commands have been completed
- The list of NVMe commands and how they work
- How error reporting structures are defined
- Power management options
- The new features added with the more recent revisions to the spec (up through 2.2)
- An overview of the storage fabrics in use today, and their strengths and weaknesses
- The changes needed for NVMe to work in a fabric environment
- How an NVMe host finds and accesses drives on a fabric
- How NVMe-oF provides a combination of attractive features for a storage fabric

Who Should Attend?

The course is ideal for RTL-, chip-, system- or system board-level design and validation engineers. Software engineers will benefit because registers programmed to control the hardware are described.

Course Length: 4 Days

Course Outline:

Basic Topics (Covered on Day 1)

- PCIe Architecture Overview
 - Introduction
 - Basic operation
 - Transaction types
 - Layers and responsibilities overview
- Introduction to NVMe
 - NVMe concepts and definitions
 - Comparison to SATA
 - NVMe Host-Controller Interface overview
 - Queue Management
 - Doorbell Head and Tail registers
 - Execution sequence
 - Tracking completion status

- Commands
 - NVMe Admin Commands
 - General Structure of commands
 - Identify command options and results
 - Creation and management of Queues
 - Asynchronous Event Notification
 - NVMe IO Commands
 - Read and Write command structure
 - Optimizing write commands
 - Avoiding write amplification
 - Buffer addressing modes (PRPs vs. SGLs)
 - Priority and Arbitration of commands
 - Completing commands
 - Handling Metadata (DIF/DIX)
 - Get Features / Set Features

Additional Topics (Covered on Day 2)

- Architecture
 - Error reporting, Error Reporting Structures
 - Asynchronous Events
 - Completion status
 - Log pages
 - Other log pages
 - Firmware updates
 - Controller registers
 - Controller initialization
 - Power Management
 - Reservations
- Other NVMe commands
 - Write uncorrectable
 - Write zeroes
 - Compare
 - Dataset management
 - Configuration overview
- Summary of changes for 1.3
 - Boot partitions
 - Thermal management by host
 - Write streams
 - Deallocation
 - Virtualization support
 - Debug support – Telemetry, initiated by host or device
 - Self-test options
 - Sanitize command
 - New commands
- Summary of changes for 1.4
 - NVM Sets
 - I/O Determinism
 - Read recovery levels
 - Persistent Memory Region (PMR)
 - Asymmetric Namespace Access (ANA)
 - Namespace write protect
 - Persistent event log
 - Rebuild assist
 - Endurance groups
 - I/O Performance and Endurance Hints

- Traffic-based Keep Alive
- Vendor-specific UUIDs
- Verify - new command
- Administrative controller
- SQ Associations

Additional Topics (Covered on Day 3)

- Summary of selected changes up to base spec revision 2.2
 - Flexible Data Placement
 - Live Migration
 - Dispersed Namespaces (NS)
 - Device Personalities
 - Rotational (Spinning) media support
 - Telemetry Log Changes
 - Persistent Event Log Enhancements
 - Updated I/O Commands
 - NVM Command Set
 - Computational Command Set Overview
 - Updated Admin Commands
 - Controller Register Enhancements
 - CMB Write Buffer registers
 - Time To Ready
 - Shutdown options
 - Zoned Namespaces Overview
 - Key Value Overview

Additional Topics (Covered on Day 4)

- Introduction to NVMe-oF
 - Motivation for moving NVMe to fabrics
 - Improve storage network performance compared to legacy protocols
 - Avoid limitations of PCIe, extend NVMe protocol to the fabric
 - Optimize access latency across the network
 - Differences from base NVMe
 - Initialization Background
- Fabric Discovery
 - Discovery Service, Discovery Controllers
 - Discovery Mechanism
- NVMe-oF Initialization
 - Fabrics Commands
 - Device Setup
 - Transport mapping/binding RDMA
 - Introduction to RDMA (Remote Direct Memory Access)
 - Protocol Translation
 - RDMA Examples
- RDMA Concepts
 - Motivation: Kernel bypass – reduce CPU involvement
 - Zero-copy operation when possible
 - RNIC (RDMA-aware Network Interface Card)
 - Pinned remote memory and storage
 - Command-based Flow Control
 - Read and Write Command Examples
- NVMe-oF Details
 - Controller Initialization
 - Connections and Associations
 - Static vs. Dynamic Controller model
 - Properties instead of registers

- Authentication
- Encapsulation
- Accessing RDMA Queue pairs
- Data transfer: Command, Response, Data, Flow Control
- Fabrics Commands and Responses
- RNIC Details
 - Software Layers
 - RDMA Verbs
 - Memory Regions and Windows
- Ethernet overview
 - TCP/IP: addressing, layers, packets
 - RDMA models added to it: RoCE and iWARP
 - Overview of iWARP
 - Added to TCP/IP, so packets are routable and don't need DCB
 - Messages
 - Layers
 - Advantages and disadvantages
 - Overview of RoCE
 - Version 2 packets include routing information, version 1 did not
 - Use of Data Center Bridging (DCB) to manage routing
 - Flow control
 - Soft RoCE
 - Advantages and disadvantages

Not Covered

- Topics specific to specs other than the base spec and NVM Command Set.

Course Materials:

1. Downloadable PDF version of the presentation slides
2. NVMe eLearning

Recommended Prerequisites:

A general knowledge of storage architecture would be very helpful, as would previous exposure to PCIe or a similar transport protocol