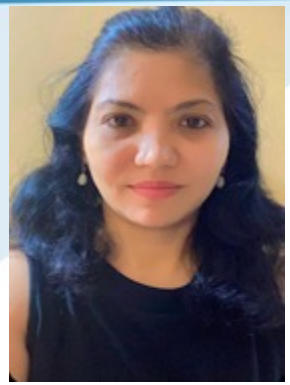
FCIA
FIBRE CHANNEL INDUSTRY ASSOCIATION

Benefits of FC-NVMe for Containerized ML Models

Live Webcast

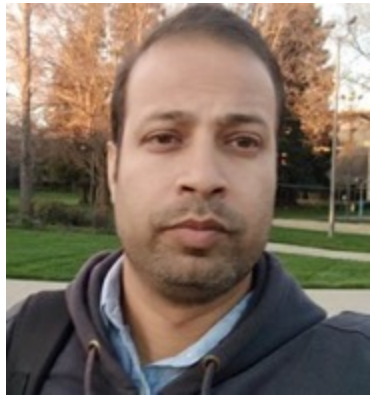
March 9, 2023

Presenters



Ramya Krishnamurthy
Master Technologist [QA Architect HPE]

Nishant Lodha
[Technologist Marvell]



Ajay Kumar
Expert Technologist [QA HPE]

Ashish Neekhra
Senior Specialist [QA HPE]

Namrata M J
[QA Engineer HPE]

Moderator

Rupin Mohan
Strategy, Product Development, CTO HPE

FCIA
FIBRE CHANNEL INDUSTRY ASSOCIATION

About the FCIA

"The Fibre Channel Industry Association (FCIA) is a mutual benefit, non-profit, international organization of manufacturers, system integrators, developers, vendors, industry professionals, and end users."

About the Fibre Channel Industry Association (FCIA)

25+ Years
Promoting Fibre Channel Technology

Industry Leading
Member Companies

150M+ FC Ports
Shipped Since 2001

FCIA
FIBRE CHANNEL INDUSTRY ASSOCIATION

Agenda

Containers

- Overview of Containers using Dockers
- Importance of Dockers in ML

Machine Learning

- ML Fundamentals
- ML Infrastructure Vectors
- ML/DL Workflow

Storage

- ML/DL Storage Access Requirements
- ML Storage architectures

FC-NVMe

- Overview FC-NVMe Architecture
- Advantages of FC-NVMe for ML

Proposed Solution

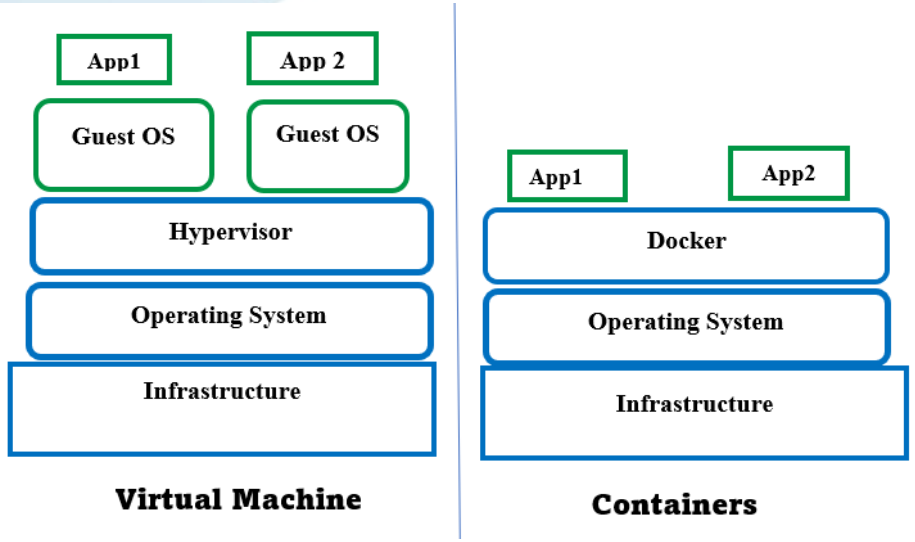
- Containerized ML Model using NVMe over Fibre channel

FCIA
FIBRE CHANNEL INDUSTRY ASSOCIATION

Containers Overview

Ashish Neekhra

Overview of Containers using Dockers



Docker is a container management tool for the creation and execution of containers. It packages the application code, configuration, and all the dependencies into a portable image which can be run on any platform or system

Few advantages of Docker containers are

- lightweight, small, and fast
- smaller in size, often only a few megabytes
- less resource-intensive than virtual machines
- Docker Containers share the same OS kernel, eliminating the need for a separate OS for each application [e.g., databases],
- Docker container packages can run applications in any environment and can be easily ported across different platforms.

FCIA Webcast: Kubernetes and Fibre Channel: A Compelling Case
https://www.youtube.com/watch?v=JBgYIhACQK8

FCIA
FIBRE CHANNEL INDUSTRY ASSOCIATION

Importance of Dockers in ML

Reproducibility in Docker

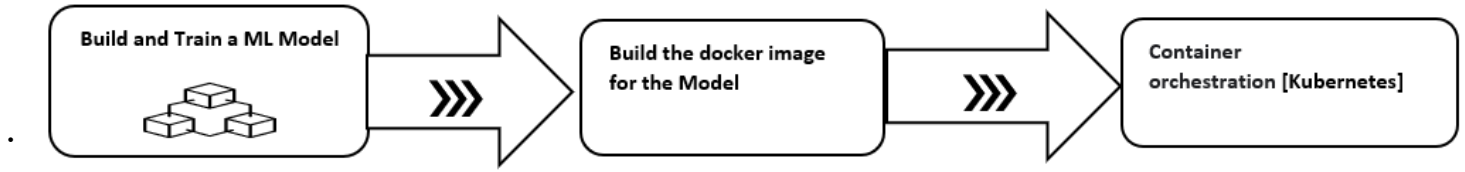
- Docker makes it possible to replicate the same environment quickly and effectively to achieve consistent results over time, which is critical for Machine Learning for data analysis.

Portability in Docker

- Docker helps to bundle your code and dependencies into containers that can later be ported to various platforms. This is useful for porting training models to computers with more resources such as CPU, RAM, and GPUs, regardless of the hardware or operating system.

Deployment in Docker

- Docker simplifies the process of deploying machine learning models



Dockers are useful at various stages of ML such as data gathering, aggregation, pre-processing, data exploration, model training, predict analysis, application deployment, and operational stage where we monitor the ML models for errors, crashes, and latency.

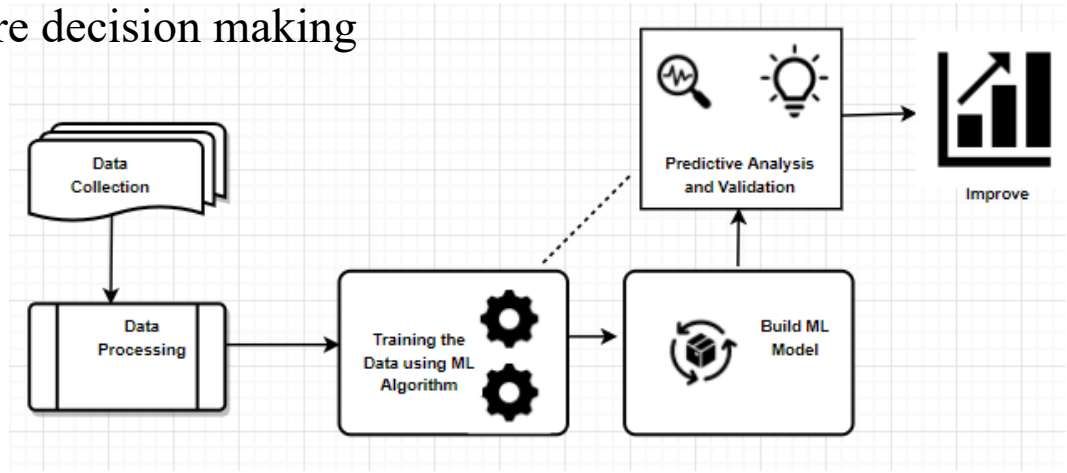
Machine Learning

Ajay Kumar

ML Fundamentals

Machine Learning Algorithm

- Fundamental logic of a Machine Learning model.
- Set of rules and statistical methodologies for extracting useful data and learning patterns from it.
- Uses the training data to build the ML model
- Machine learning models improve as more training data is collected.
- Identifies patterns that are required to predict the outcome.
- Incorporates corrections and improves its future decision making



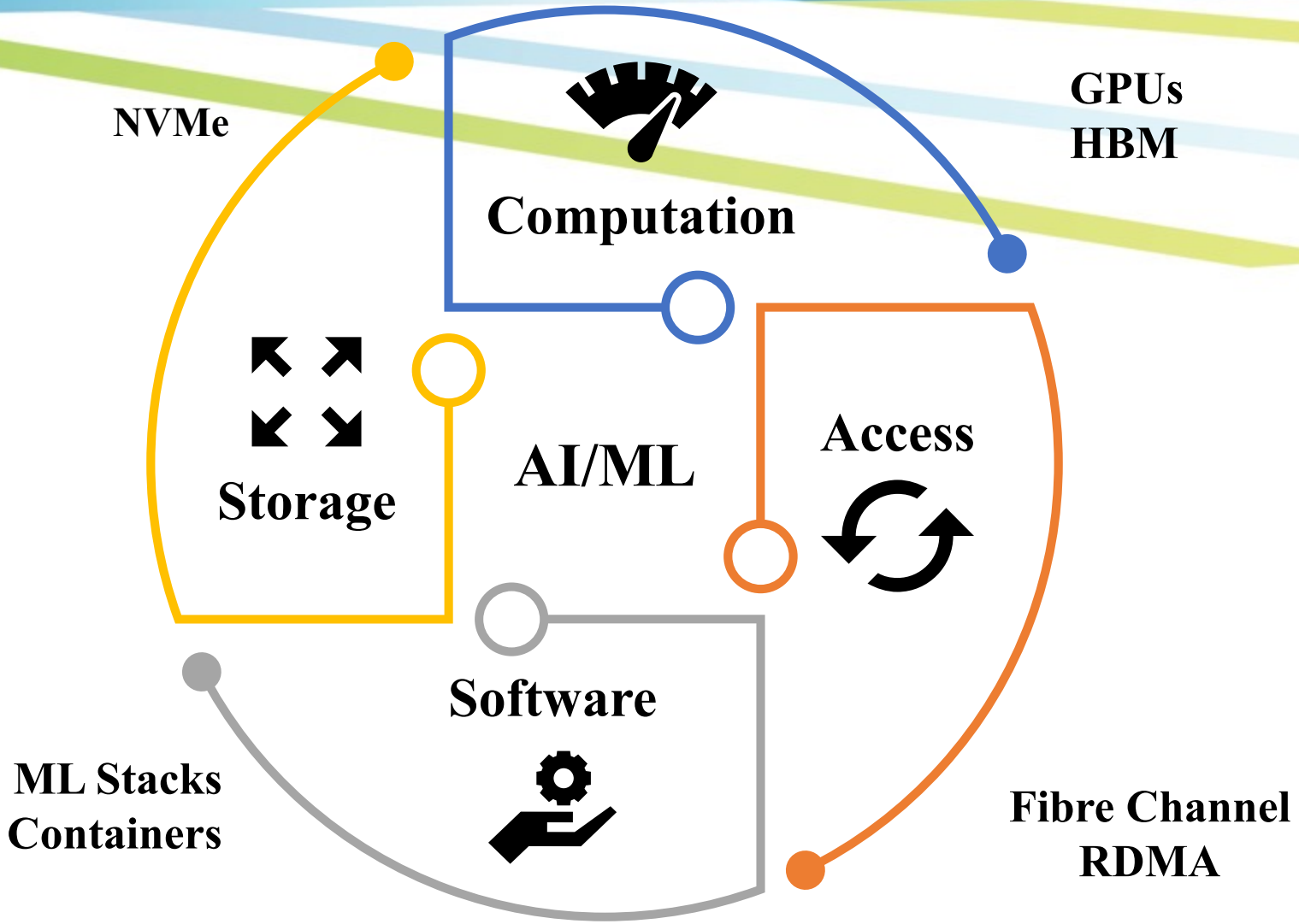
There are different text mining techniques like information extraction, information retrieval. Selecting the appropriate ML model, is a crucial phase in the ML process for data analysis and for the intended outputs

Storage

Nishant Lodha

Machine Learning Infrastructure Vectors

Storage and Storage Access are often after thoughts



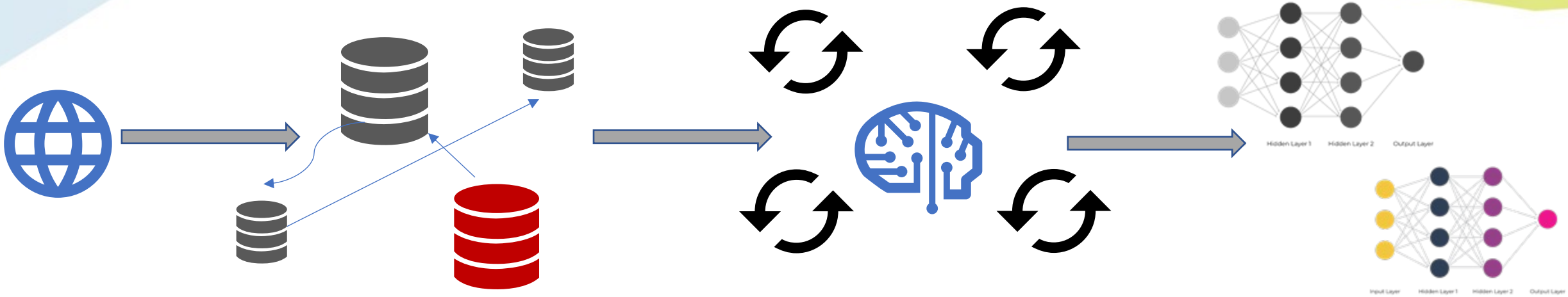
ML/DL Workflow

Instrument and Ingest

Clean and Tag

Train and Validate

Catalogue and Inference



Accumulation

Pre-Processing

Modelling and Inferencing

Access to shared storage is required through the lifecycle

FCIA
FIBRE CHANNEL INDUSTRY ASSOCIATION

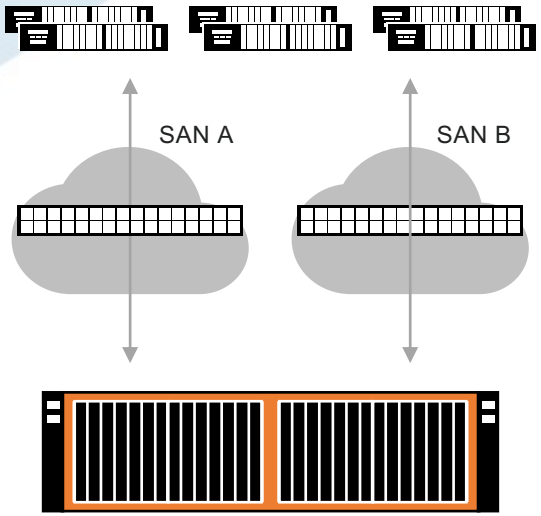
ML/DL Storage Access Requirements

Training speed and accuracy highly dependent on Storage

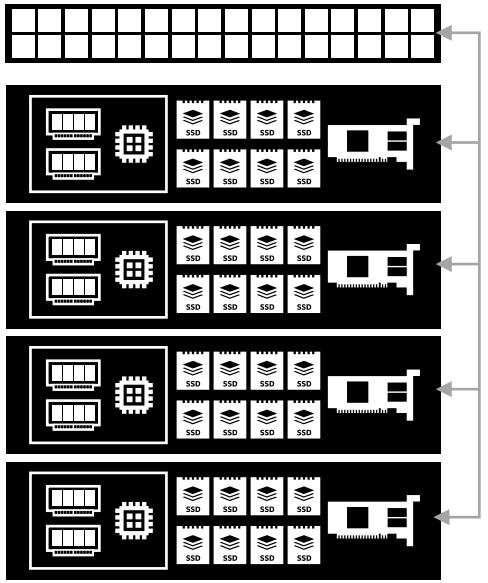
Cost Efficiency	**Storage MLops adds to overall cost (CAPEX) of AI/ML systems. Container native storage with FC delivers the required efficiency**
Scalability	**Capacity and performance. Block Storage and Fibre Channel delivers scale for ML datasets**
Reliability	**Long running workloads like Deep Learning require reliable access to Storage**
Data Locality	**Shared storage architecture delvers the ability to share datasets without unnecessary copies**

ML Storage architectures

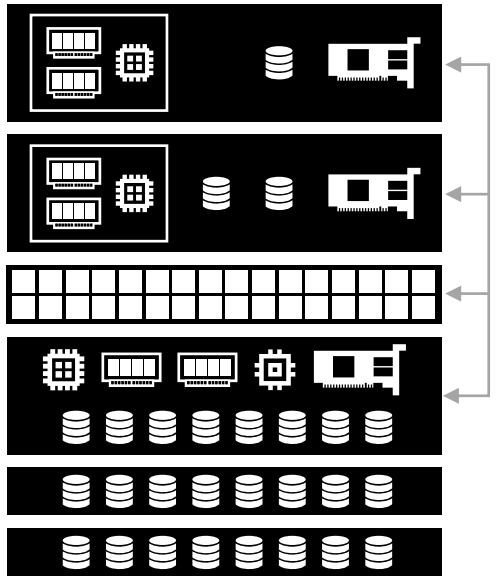
Shared-Storage SAN



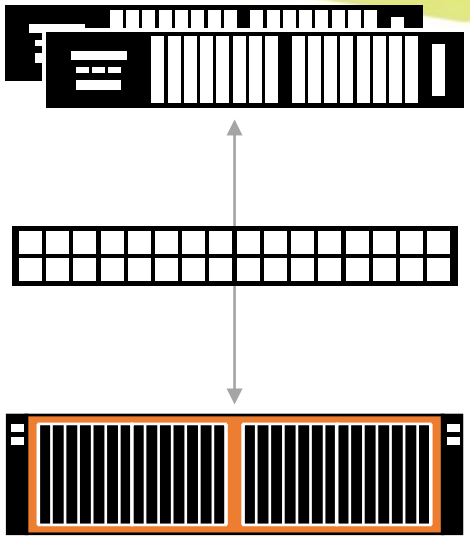
Hyper-converged Infrastructure (HCI)



Software Defined Storage (SDS)



Just a Bunch of Flash (JBOF) and Ethernet BOF



Fibre Channel

Ethernet

FCIA
FIBRE CHANNEL INDUSTRY ASSOCIATION

FC-NVMe

Ramya Krishnamurthy

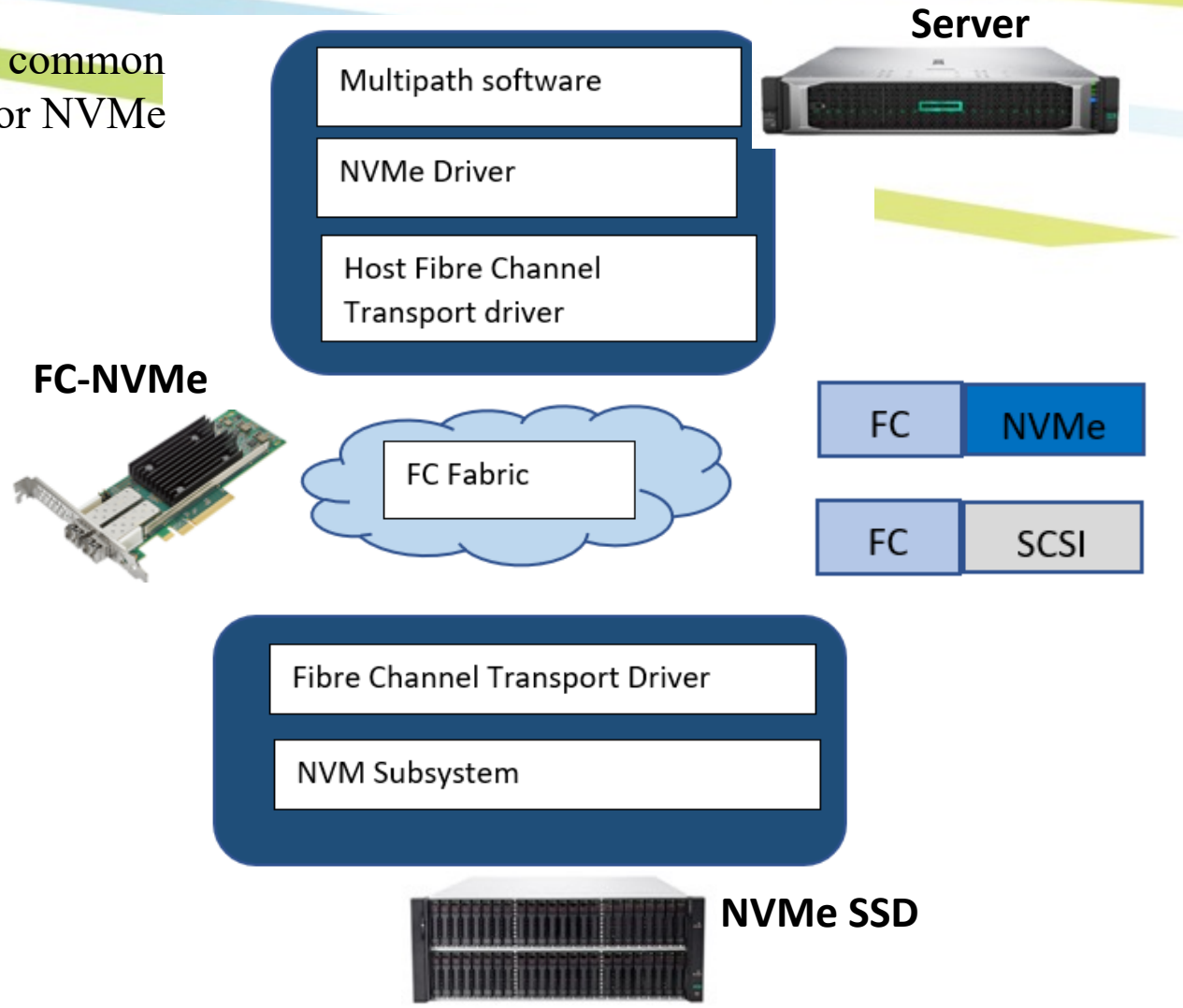
FC-NVMe Architecture

NVMe® over Fabrics using Fibre Channel (FC-NVMe) defines a common architecture that supports a range of storage networking fabrics for NVMe block storage protocol over a storage networking fabric

NVMeoF Objects

Subsystem	• represents the controllers, the namespaces and the actual physical media that backs the namespaces
NQN	• used to uniquely describe a host or NVM subsystem for the purposes of identification and authentication • . An NQN is permanent for the lifetime of the host or NVM subsystem.
Namespace	• is a collection of logical blocks whose logical block addresses range from 0 to the size of the namespace – 1. A namespace ID (NSID) is an identifier used by a controller to provide access to a namespace.
Controllers	• A controller for NVMeoF is a construct that is created when the host sends a NVMeoF connect command to the target NQN for the admin queue

FCIA Webcast: What's New in FC-NVMe
https://www.youtube.com/watch?v=UmzLbpPGIVM



Advantages of FC-NVMe for ML

Modern model training datasets are large, and Machine learning tasks process these massive volumes of data from both organized and unstructured sources.

Why FC-NVMe 2 ?

The next gen FC-NVMe provides better resiliency, improved scalability, and better performance

Artificial intelligence (AI), machine learning, and deep learning involve scanning millions of little files, and they demand sustained, predictable performance and low latency and NVMe over Fibre Channel is the best option for delivering the performance, application response time, and scalability needed for ML workloads

A few of key advantages of NVMe over Fibre channel

- NVMe over Fibre channel enables faster access between hosts and storage systems
- It is lighter than SCSI and has a lean driver stack that allows it to run faster while consuming fewer resources.
- It's a multi-queue model [Supports 64K I/O queues, with up to 64K commands per queue] with built-in multipathing capabilities that help to improve fault tolerance
- The NVMe protocol meets the ever-demanding and dynamic needs of storage environments for high-performance AI/ML workloads which is a key-value proposition when it comes to serving AI, ML, and DL workloads

Proposed Solution

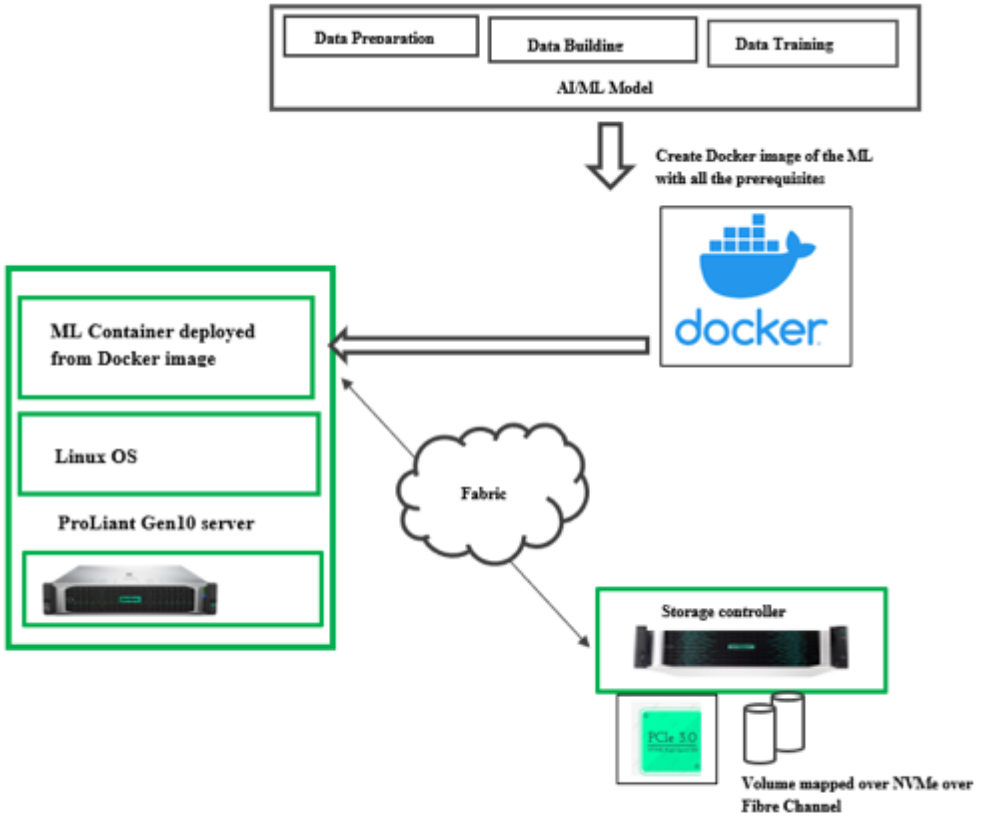
Ashish Neekhra and Ramya Krishnamurthy

Proposed Solution-Containerized ML Model Using NVMe over Fibre Channel

- *Build the model.*
- *Create the requirements file containing all the required libraries.*
- *Create the docker file with necessary environment setup*
- *Build the Docker image of the ML model*

- Linux Operating system (O/S) on the server that provides support for NVMe over Fibre Channel
- The storage controller uses PCIe Gen 3 and supports end-end NVMe [NVMe based flash drives [SSD]and NVMe over Fibre Channel]
- The Storage controller uses Samsung's Multi Level Cell (MLC) Solid State Drive (SAMSUNG MLC 2TB SSD drives)
- Both the server and storage system are configured with Gen 6 FC 32Gb HBAs.

The machine learning model used here is a Deep learning model



Containerized ML model using NVMe End-End [NVMe SSD and NVMe over Fibre Channel]

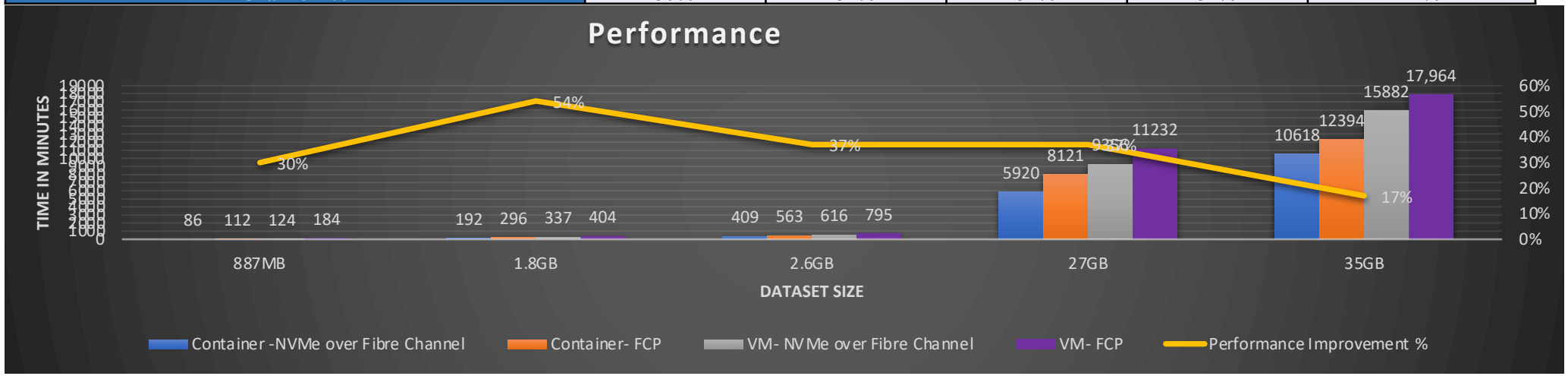
FCIA
FIBRE CHANNEL INDUSTRY ASSOCIATION

Performance Evaluation- NVMe over Fibre Channel

As a baseline, evaluated the performance of our ML model running on the virtual machine and container deployed on server using storage over FCP. Since containers had lower overhead than virtual machines-we see better performance with containers in comparison to the VM

Performance evaluation was then done between VM and container with NVMe over Fibre channel and as we see we get more performance gains with the proposed container solution when using NVMe over Fibre Channel. Performance evaluations were conducted for a variety of training dataset sizes.

Time in minutes					
Dataset size	**887MB**	**1.8GB**	**2.6GB**	**27GB**	**35GB**
Container -NVMe over Fibre Channel	**86**	**192**	**409**	**6920**	**10618**
Container- FCP	**112**	**296**	**563**	**8121**	**12394**
VM- NVMe over Fibre Channel	**124**	**337**	**616**	**9356**	**15882**
VM- FCP	**184**	**404**	**795**	**11232**	**17,964**
Performance Improvement with container on NVMe over Fibre Channel %	**30%**	**54%**	**37%**	**37%**	**17%**



Summary and Q&A

1. Shared storage architectures provide a reliable and efficient mechanism for DL workflows
2. Fibre Channel is the #1 choice for mission critical workloads accessing shared storage
3. NVMe over Fibre Channel reduces the training time for container ML/DL workloads
4. Robust reliability, performance and full offload makes FC an ideal choice for AI/ML

After this Webcast

- Please rate this event – we value your feedback
- We will post a Q&A blog at http://fibrechannel.org/ with answers to today's questions
- Follow us on Twitter @FCIAnews for updates on future FCIA webcasts
- Our library of FCIA on-demand webcasts http://fibrechannel.org/webcasts/ includes:
  - Fibre Channel Fundamentals
  - FC-NVMe
  - Long Distance Fibre Channel
  - Fibre Channel Speedmap
  - FCIP (Extension): Data Protection and Business Continuity
  - Fibre Channel Performance
  - FICON
  - Fibre Channel Cabling
  - 64GFC
  - FC Zoning Basics

Thank You