SNIA | NETWORKING
NSF | STORAGE

Everything You Wanted to Know About Throughput, IOPs, and Latency But Were Too Proud to Ask

Live Webinar

February 7, 2024

10:00 am PT / 1:00 pm ET

Today's Presenters

Erik Smith
Distinguished Engineer
Dell Technologies

Bill Martin
Co-Chair SNIA Technical Council
Principal Engineer, SSD IO Standards
Samsung

Krishnakumar Gowravaram
Senior Principal Engineer,
Connectivity and Cloud
Solutions, Celestica

SNIA NSF | NETWORKING STORAGE

The SNIA Community

200 Corporations, universities, startups, and individuals	**2,500** Active contributing members	**50,000** Worldwide IT end users and professionals

Ethernet, Fibre Channel, InfiniBand®

iSCSI, NVMe-oF™, NFS, SMB

Virtualized, HCI, Software-defined Storage

Storage Protocols (block, file, object)

Securing Data

Technologies We Cover

SNIA. NSF | NETWORKING STORAGE

SNIA. NSF | NETWORKING STORAGE

SNIA Legal Notice

- The material contained in this presentation is copyrighted by SNIA unless otherwise noted.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced in their entirety without modification
  - SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of SNIA.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

  NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

Today's Agenda

- Definition of Terms
- Storage system bottlenecks
- Real-World Examples
  - OLTP Application
  - AI/ML Application
- Q&A

Definition of Terms

- Bandwidth (data transfer capacity)
  - The maximum rate at which data can be transmitted
  - Often incorrectly used when referring to throughput
- Throughput (data transfer rate)
  - The amount of data per unit time actually moved across an interconnect
- Latency
  - A time period between two events
  - Examples:
    - for I/O request execution time:
      - the time between the making of an I/O request and completion of the request's execution
    - for rotational latency:
      - the time between the completion of a seek and the instant of arrival of the first block of data to be transferred at the disk's read/write head
- IOPs
  - I/O Operations per second
  - Measured at the source of the I/O Operations

What Impacts Throughput and Latency?

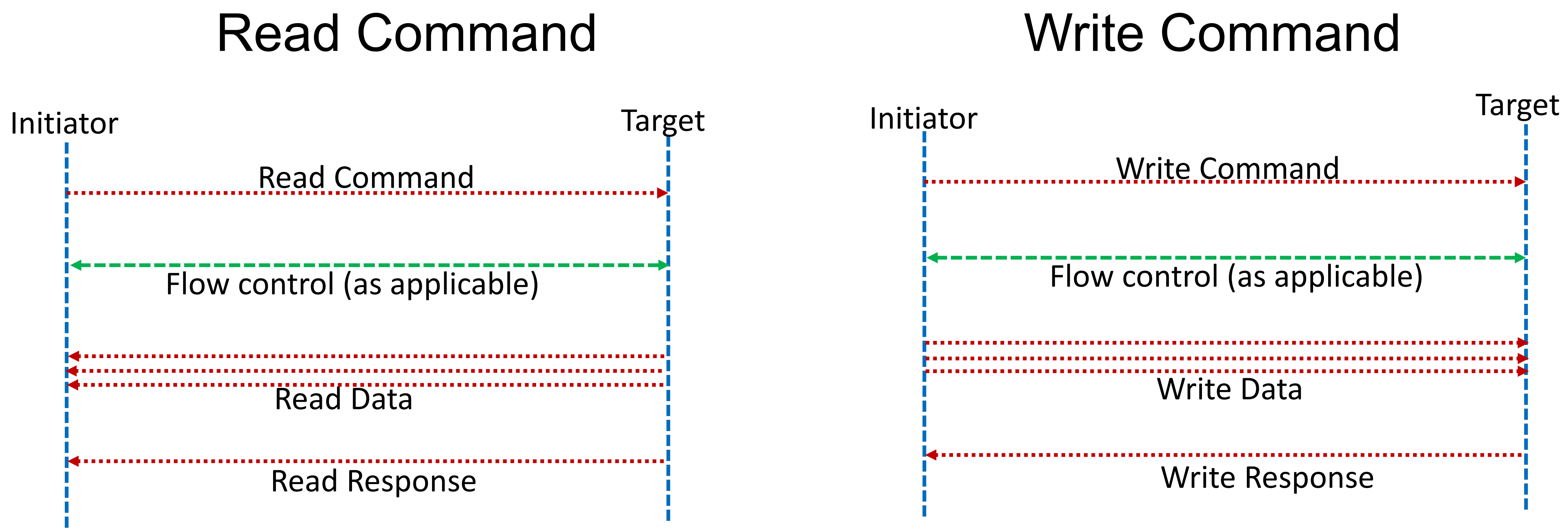
- Throughput (data transfer rate)
  - The maximum throughput is equal to the bandwidth of the interconnect
  - Reduction may occur due to congestion
    - Multiple hosts and devices on the interconnect
- Latency
  - Latency may be increased due to contention for resources in a device
  - Examples:
    - Multiple requests for reading or writing data to the same physical media
    - Reads and writes that utilize the same controller or bus to physical media
- IOPs
  - IOPs may be reduced by both throughput reduction and increased latency

Basics of Storage IO

Storage I/O

- Moves data from/to Storage client to/from Storage device
  - Storage client is called a Initiator & Storage Device is called a Target
- Consists of a Command phase, one or more Data phases & a Response phase
- Every I/O has a unique ID called IOTag.
  - Unique IO-tags are allocated by Storage client and Storage device
- I/O is considered complete when either a Response is received by Initiator or a Timeout occurs
- Many protocols for Storage I/O
  - ATA, SCSI, NVMe
- Different transport protocols move storage data
  - SAS, FC, Ethernet, PCIe

Storage I/O Sequence



Steps in Storage I/O

- Initiator
  - Allocate IoTag,
  - Create command buffer
  - Allocate data buffers
  - Send command to target
  - Starts I/O timer
  - Process I/O Response Or
  - Handle I/O timeout
- Target
  - Receive & parse IO Command
  - Allocates IoTag
  - Allocate Data buffers
  - Initiate one or more Data phases
  - Create Response buffer and transmit Response to the Initiator

Storage I/O Parameters

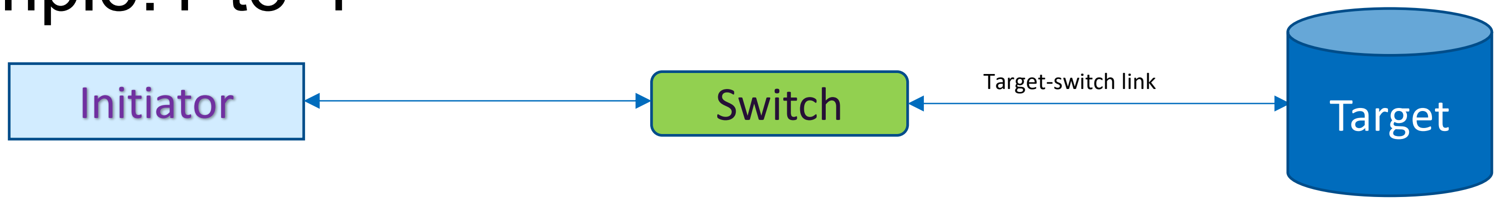
- Data Direction: Read or Write
  - Read – Initiator receives data
  - Write – Initiator sends data
- Block Size:
  - Data payload transferred as a single unit
  - Small blocks: 4KB, 8KB, 16KB etc.
  - Large blocks: 256KB, 512KB, 1MB etc.
- Access pattern:
  - Sequential Access:
    - Reading or Writing data is performed from/to contiguous blocks.
    - Has less overhead when seeking data
  - Random Access:
    - Reading and Writing data is performed from/to non-contiguous blocks.
    - Adds overhead when seeking data
    - Less of a issue with SSD media

Storage I/O Parameters

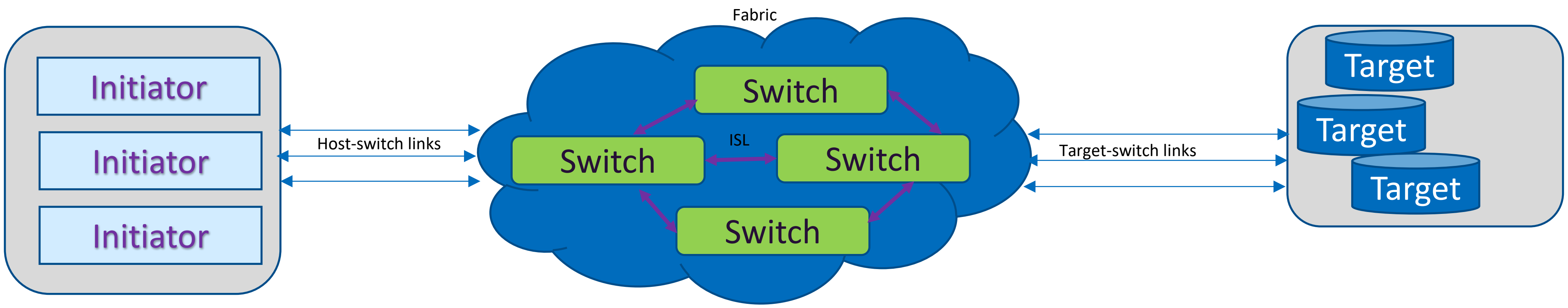
- Distance (Hop count)
  - Initiator and target can be connected to each other
    - Ex. Local disk drives in a server
  - Initiator and target are connected over a network
    - Ex. SAN (Storage Area Network) or NAS (Network Attached Storage)
  - Complexity grows with multiple components and over distance
    - multiple initiators, sharing multiple targets over a multi-switch fabric

Storage Network: Initiator --Target

- Simple:1-to-1



- Complex: Many-to-Many



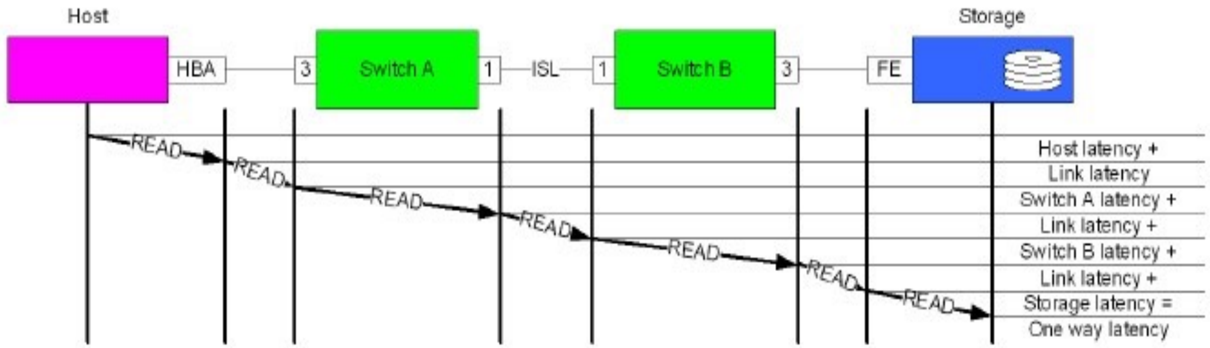
SNIA NSF | NETWORKING STORAGE

Storage Bottlenecks

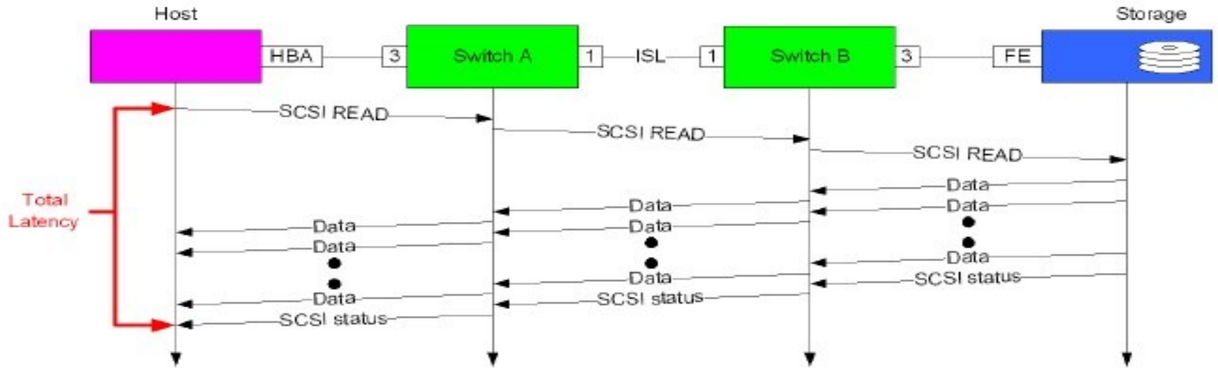
- Bottlenecks generally move from 1 point to another, much like traffic.
- Bottlenecks can at the component level:
  - Drives, Controllers (NIC, HBA), Interconnects (Switches, Expanders), SW stacks
  - HDD → SSD
  - Faster Interconnects → FC: 2,4,8..128Gbps & Ethernet: 1,10,100,800Gbps
  - SCSI →NVMe
- Key is to match component capabilities, much like synchronized/timed traffic signals
- Queue Depth
  - Host queue depth: Number of I/O requests Initiator stack can issue.
  - Target queue depth: Number of I/O requests that Target can service at a time.
  - Host queue depth > Target queue depth – Queue full status
    - I/O request needs to be retried – adds to complete round trip

Storage Bottlenecks

- Distance – Hop count



- Latency adds up with every hop & every phase



SNIA NSF | NETWORKING STORAGE

Storage Bottlenecks

- Storage architecture
  - RAID configuration, Data distribution, Data replication
  - Garbage collection, Write amplification

- Block size and access pattern
  - Large block IO is better suited for Sequential access
  - Small block IO is better suited for Random access

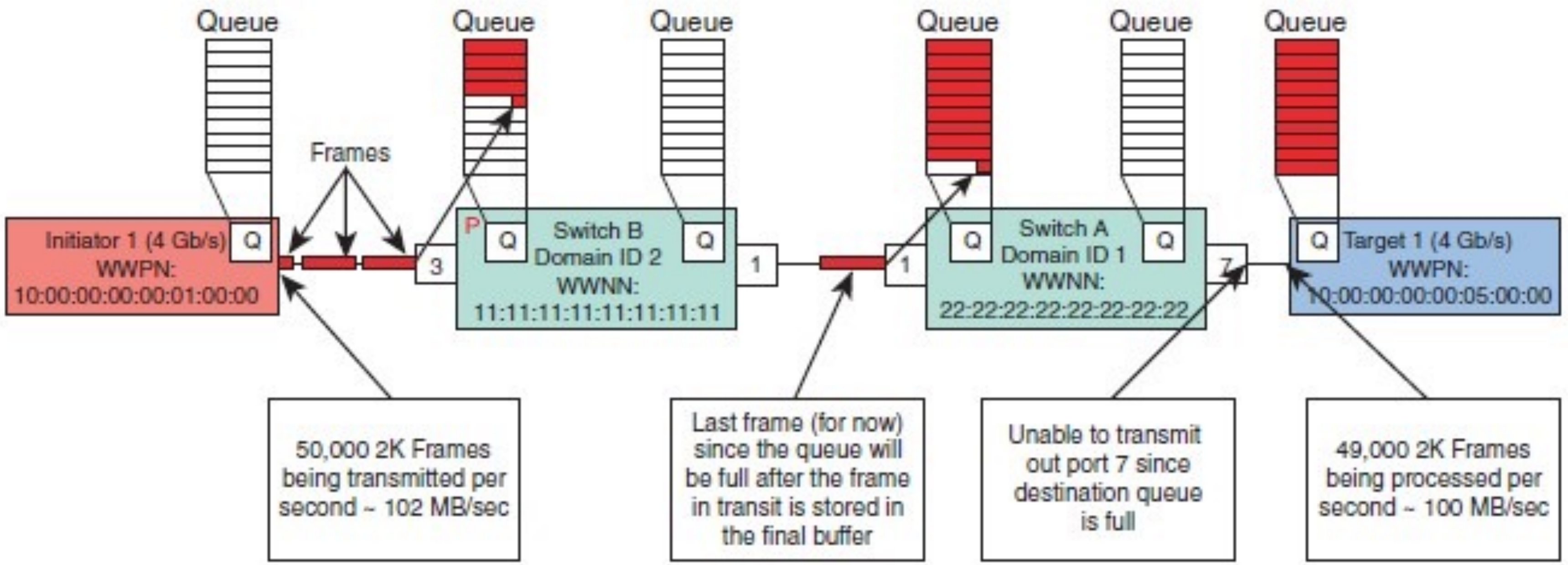
- I/O Blender effect
  - I/O is no longer predictable
  - Virtualization & Multi-Tenancy
  - Virtual machines & storage virtualization
  - Common storage hardware shared across multiple applications

Storage Bottlenecks

- Different applications, Different I/O needs
  - Inefficient caches
  - Read/Write scheduling
  - Large/small block sizes
  - Sequential & random IO
- Network Configuration
  - Networks play a big role in performance
    - Network bandwidth, congestion, latency, packet loss
    - Slow drain, oversubscription

Storage Bottlenecks

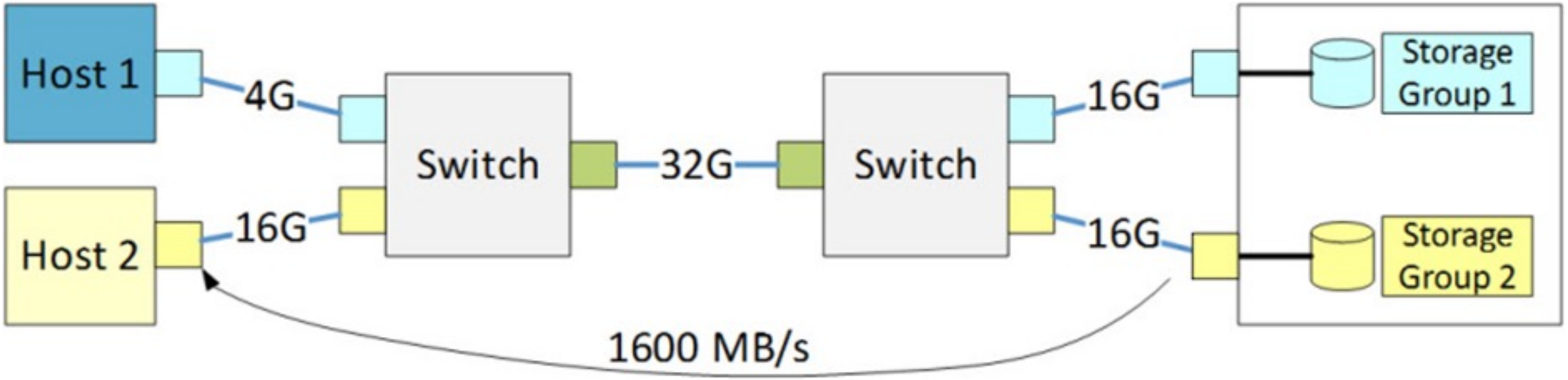
- Slow Drain
  - One slow drain device can cause congestion on a shared link



SNIA NSF | NETWORKING STORAGE

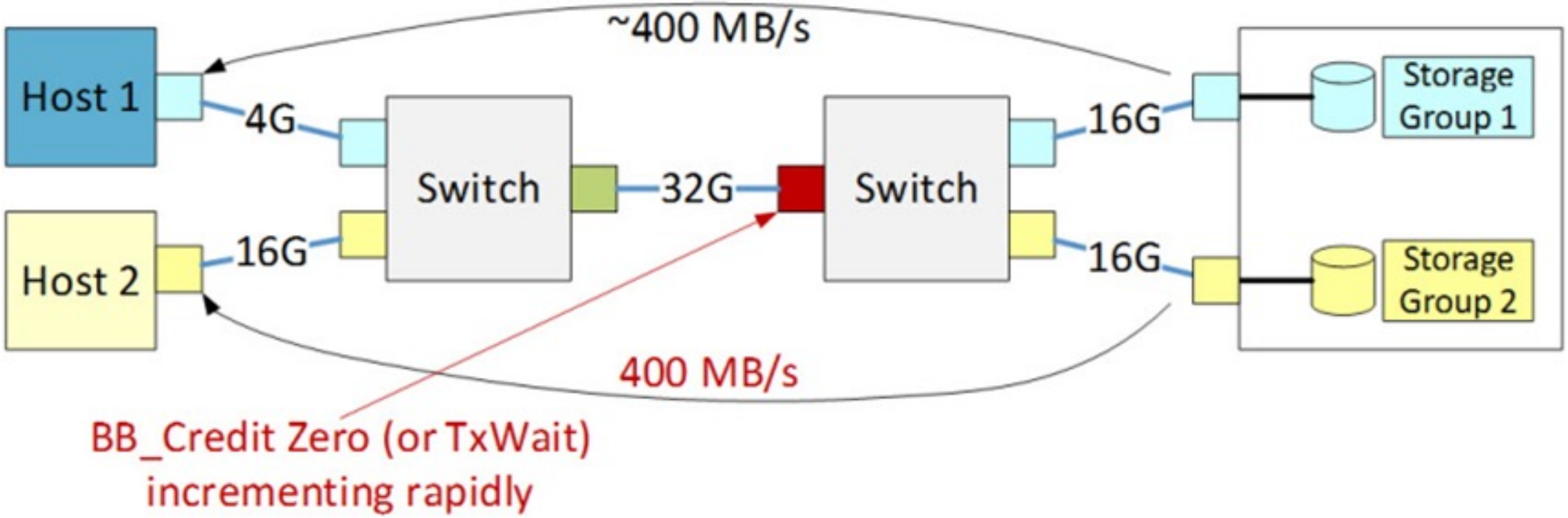
Storage Bottlenecks

- Oversubscription
  - Host 2 & Storage Group 2 are bandwidth matched – providing a 32Gbps end-to-end link.



Storage Bottlenecks

- Host 1 performs a large block Read 512K
- Host -2 Read throughput drops to 400 MB/s



SNIA NSF | NETWORKING STORAGE

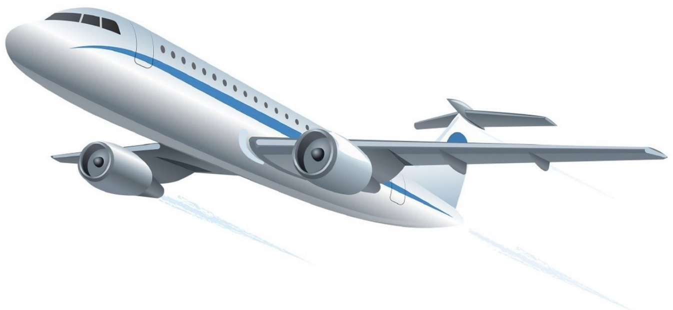
Application – Storage

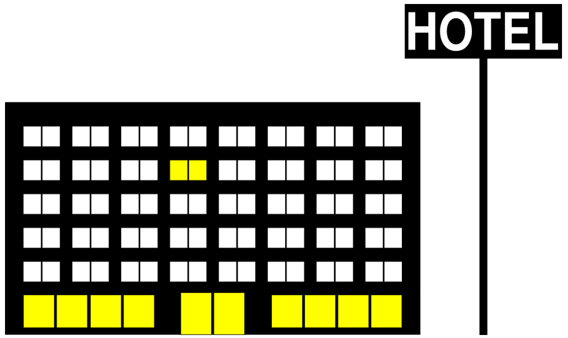
- Data: Application - Storage relationship
  - Not all data is same
    - Data Tiers
    - Hot, Warm, Cold
  - Data Type
    - FS metadata, KV store, audio/video, pics,
  - Combination of I/O parameters: Data direction, Block size, Access Patterns
    - Large blocks & Sequential access offer better throughput
    - Small blocks & Random access offer better IOPs.

OLTP Application

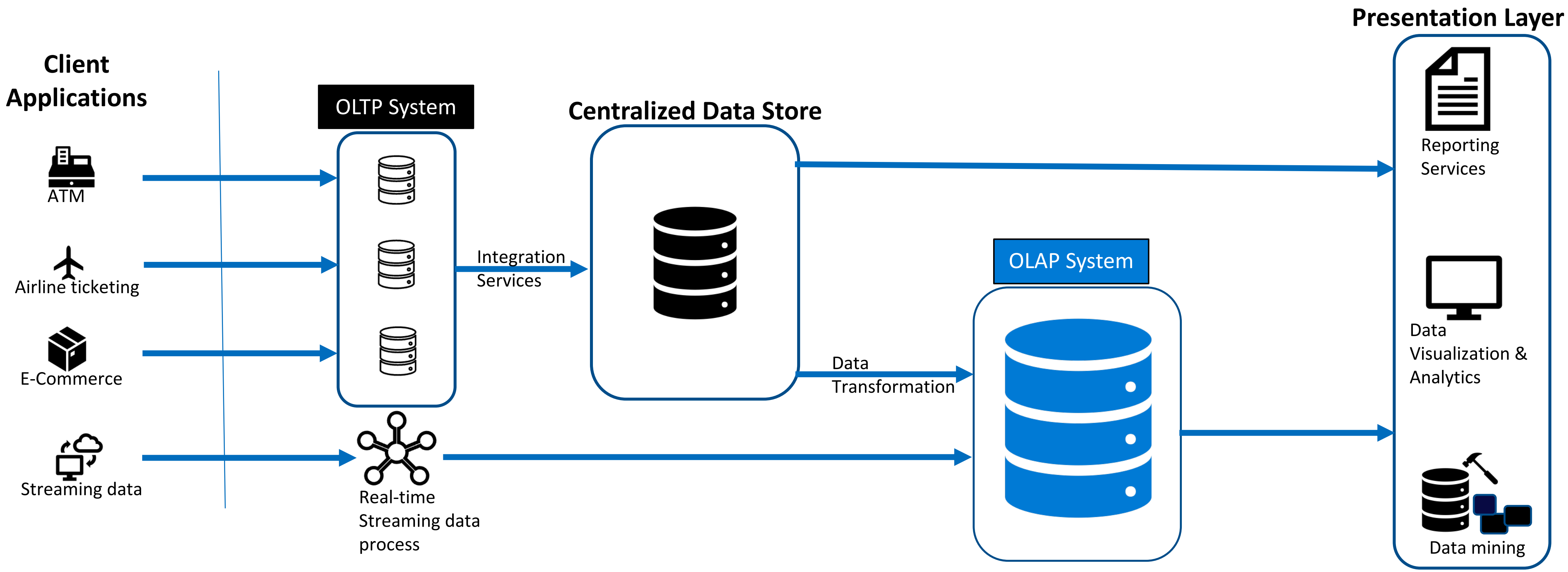
OLTP

- Use cases: Banking, airline/hotel ticketing, e-commerce, ATMs
- OLTP is generally deployed with OLAP
- Both applications work with the same data-set
  - But I/O needs are different
  - Look into OLTP I/O requirements





OLTP Deployment

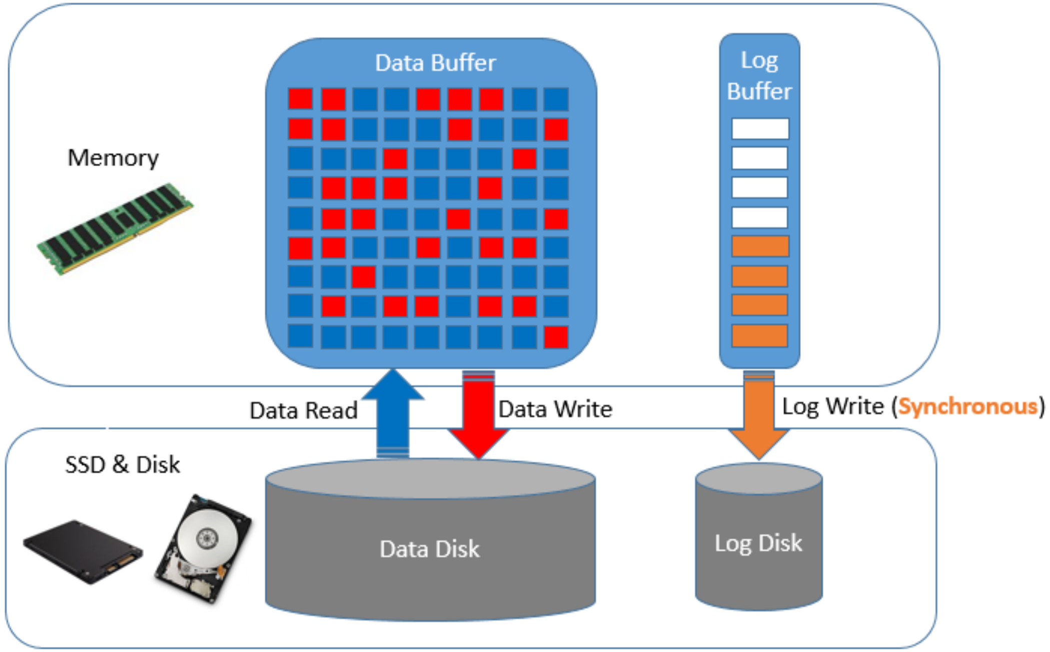


OLTP I/O Patterns

- Large number of IO
  - Support large IOPs
  - Support large queues
- Small IO
  - Handle Random IO
  - Handle Read/Write mix
  - Handle head-of-line blocking – Write holding Reads
- Real-time access
  - Response time in milliseconds
- Requests from multiple sites
  - Data integrity – multiple concurrent access
  - Data replication – may require synchronous replication
- Atomic & stateful
  - Changes are persisted, inserts, deletes, modify
  - Manage Cache effectively

OLTP I/O Patterns

- Data files
  - Accessed Random reads/writes
  - Generally 4K – 64K IOs
- Archive logs
  - Sequential writes
  - Large I/O size - 128K – 1M writes
- Redo logs
  - Sequential writes
  - Small I/O size – align I/O size to physical disk block size (512 bytes, 4K)
- Separate Data volume & Log volume

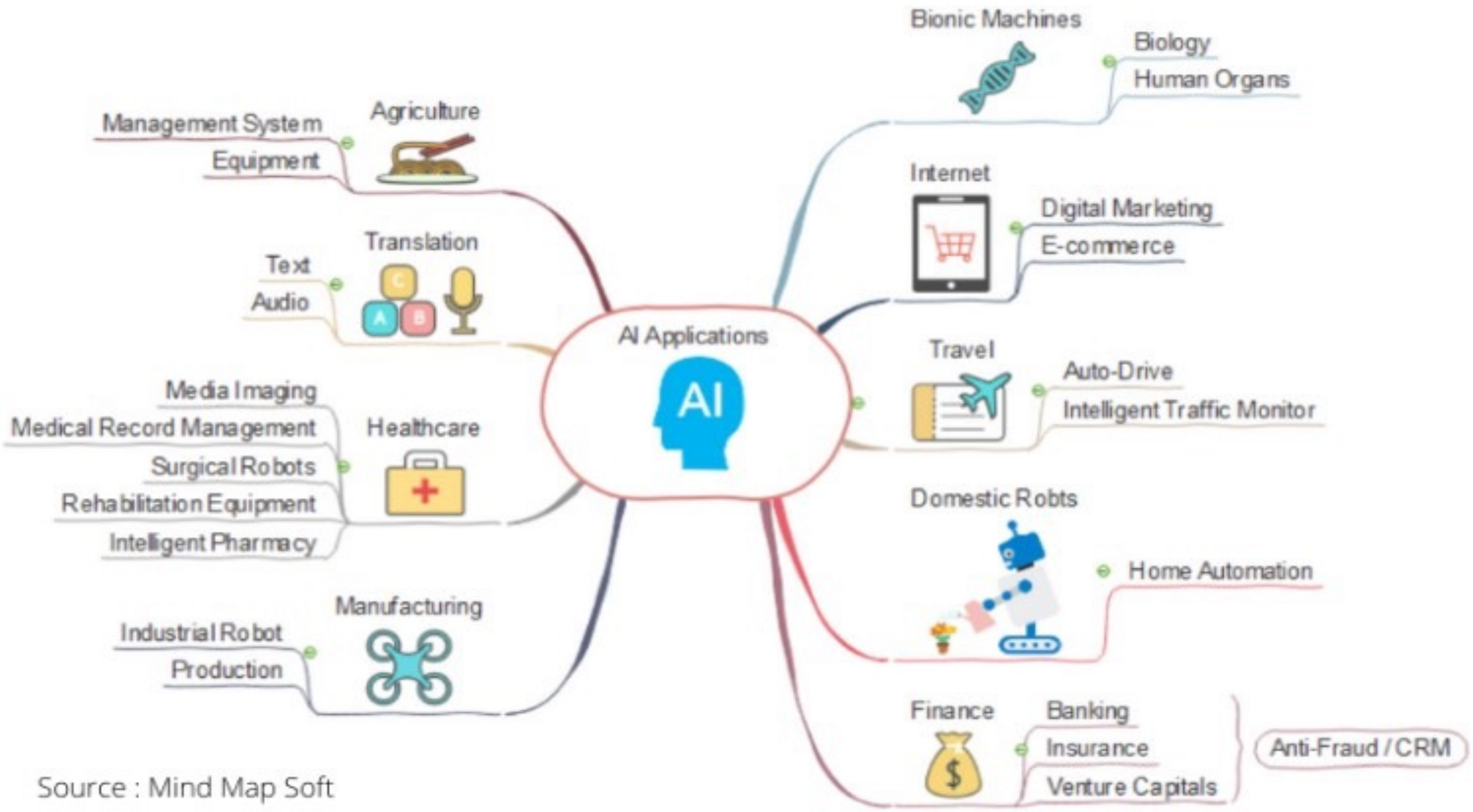


SNIA NSF | NETWORKING STORAGE

AI & ML Application

AI / ML Use Cases

- AI/ML is revolutionizing the way we interact with technology

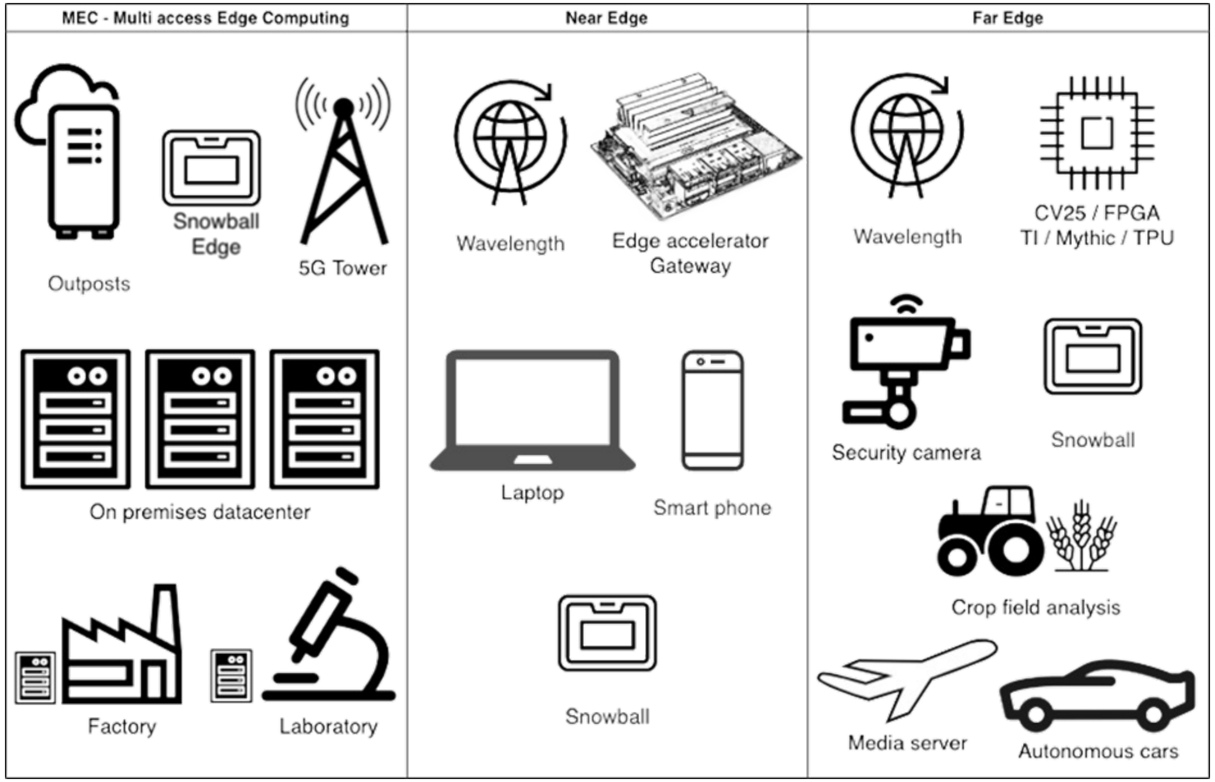


Source : Mind Map Soft

SNIA NSF | NETWORKING STORAGE

AI / ML Use Cases

- Data centers, Mobile phones, Appliances, Robots, Sensors

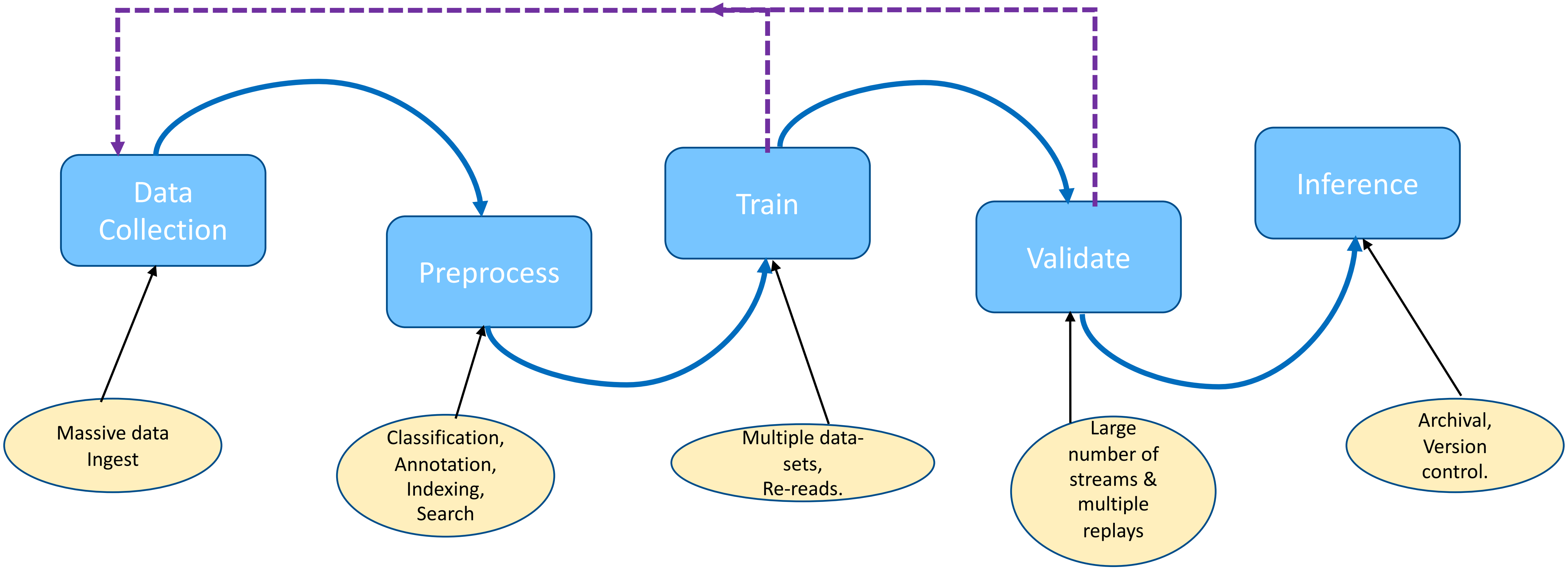


AI / ML Use Cases

- Is not limited to sandbox for data scientists.
  - Every company/organization need to be AI-ready
- Small clusters for basic model training provide 80% accurate predictions.
- 80% → 99.x% accuracy adds significant load on infrastructure.
- Large datasets are required for more accurate predictions
  - Datasets of several petabytes
  - Storage often becomes the bottleneck

AI / ML Workflow

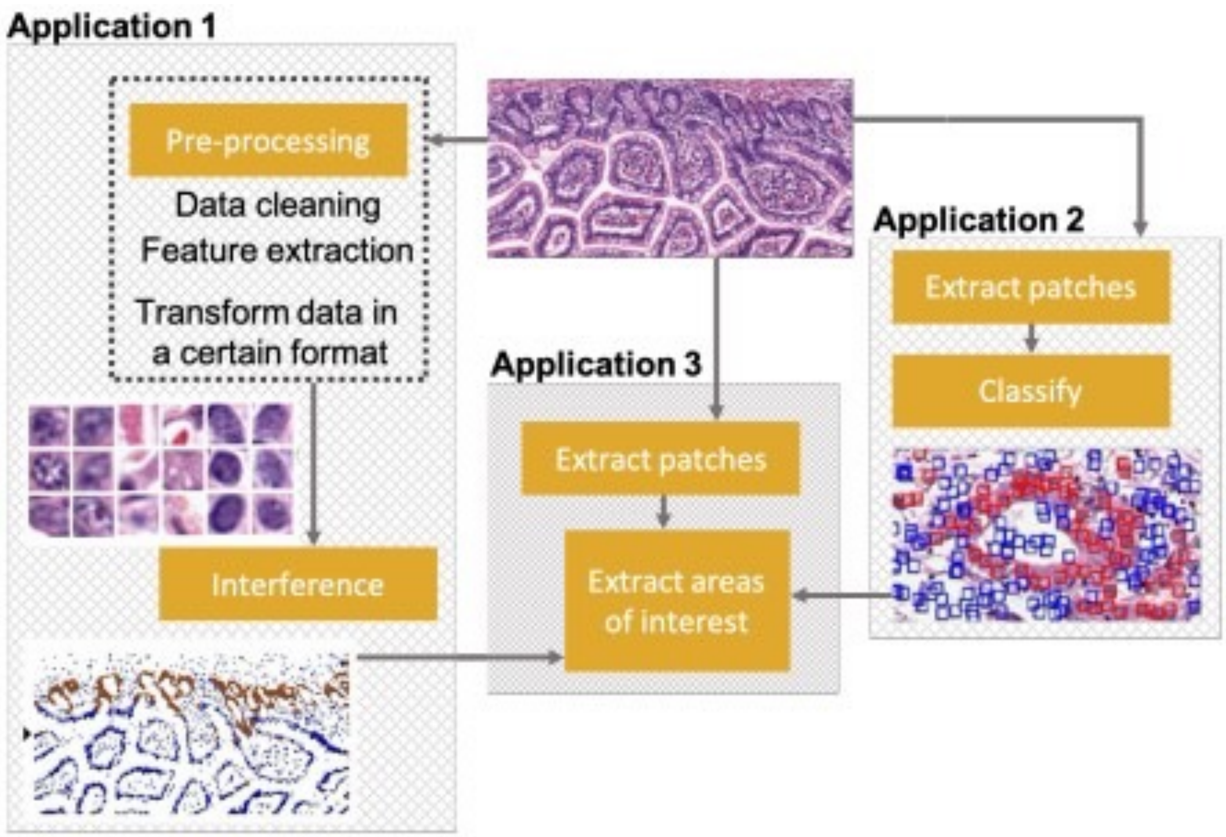
- AI/ ML application different phases



SNIA NSF | NETWORKING STORAGE

AI / ML Workflow

- Multiple phases – Multiple applications



AI / ML Data Characteristics

- Scale with performance
  - Handle large data-sets (Petabyte scale)
  - Throughput & Low latency
  - Random Reads & Writes
  - Large IO & Concurrent IO
- Unstructured
  - Datasets can be images, audio, video, tabular with millions of rows
- Immutable
  - Training needs repeatable experiments
- Large number of files
  - Data-sets from different sources over a period of time
- Portable
  - Need to moved across different environments
- Distributed and resilient
  - Collaboration, availability
- RESTful APIs
  - Communication between different services
- Secure
  - Lock data, version control, retention policy

AI / ML Storage Needs

- Data collection:
  - Large capacity – data lakes
  - Need high write throughput
  - High queue depth
- Preprocess
  - High Read/Write throughput
  - Sequential writes
  - High queue depth
  - Handle multiple streams
- Model Training
  - Needs low latency – high latency will slow training
  - Requires throughput
  - Support different IO sizes
    - Separate data stores
  - Large file count

AI / ML Storage Needs

- Validation
  - IOPS
  - Large queue depths
  - Good Cache management
- Inference
  - Low latency
  - Smaller IO
  - Multiple data sets
  - Mixed workloads

Summary

- Storage performance metrics
  - What they mean and why they are important
- Storage bottlenecks are hard to fix
  - Bottlenecks move from one point to another
  - Need a holistic view of storage ecosystem
- Different applications have different storage needs
  - Application may have different I/O characteristics at different points
  - Understanding I/O perf metrics helps in designing and choosing correct Storage

Q&A

After this Webinar

- Please rate this webinar and provide us with your feedback
- This webinar and a copy of the slides are available at the SNIA Educational Library https://www.snia.org/educational-library
- A Q&A from this webinar, including answers to questions we couldn’t get to today, will be posted on our blog at https://sniansfblog.org/
- Follow us on Twitter @SNIANSF

SNIA NSF | NETWORKING STORAGE

Thank You

SNIA NSF | NETWORKING STORAGE