STORAGE TECHONOLOGIES FUNDAMENTALS

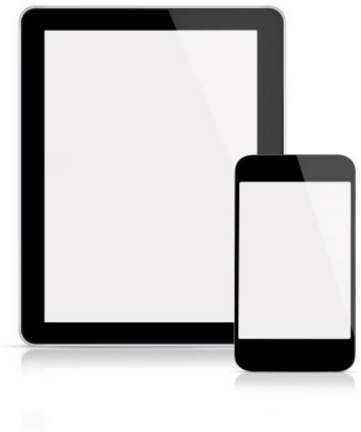
LOGISTICS

Class Hours:

- Class starts at X:00 AM/PM Time and Finish at XX:00 AM/PM
- There will be regular breaks in class.

Telecommunication:

- Turn off or set electronic devices to silent (not vibrate)
- Reading or attending to devices can be distracting to other students
- Try to delay until breaks or after class



COURSE STRUCTURE

Day 1: Storage Fundamentals & Architecture

Day 2: Flash Storage & Pure Storage

Day 3: Data Protection & Management

Day 4: Kubernetes Integration & Modern Applications

PREREQUISITES & EXPECTATIONS

- Basic understanding of computer systems, networking, and programming fundamentals
- One or more years of technical experience

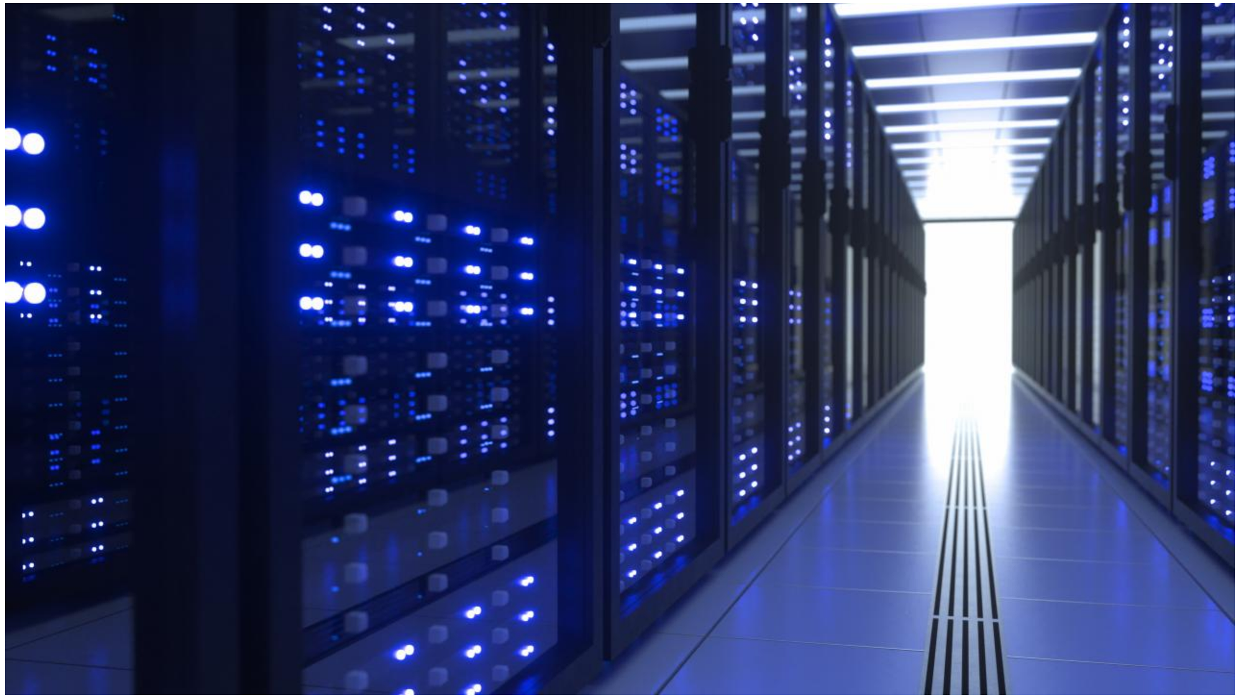
AGENDA FOR DAY 1

1. Storage Fundamentals & Architecture
2. Storage Evolution & Business Context
3. Modern Storage Architectures
4. Storage Architectures in Practice
5. Storage Types & Selection

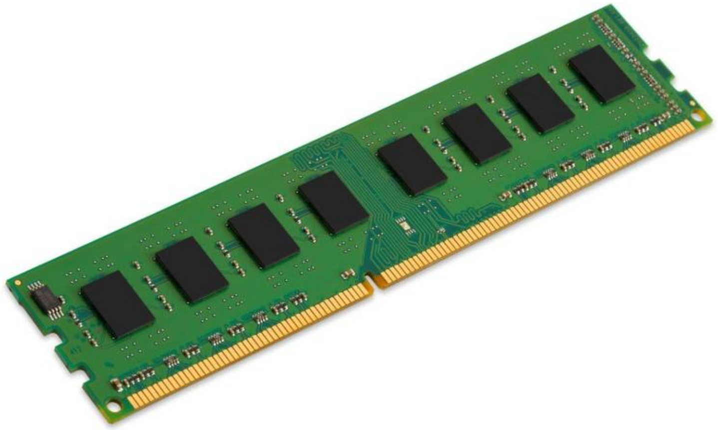
STORAGE FUNDAMENTALS & ARCHITECTURE

WHAT IS DATA STORAGE?

- Storage refers to the process of retaining digital data so it can be processed, retrieved, or backed up when needed. It is a foundational component of any computing system, determining how information is accessed, protected, and preserved over time



TYPES OF STORAGE



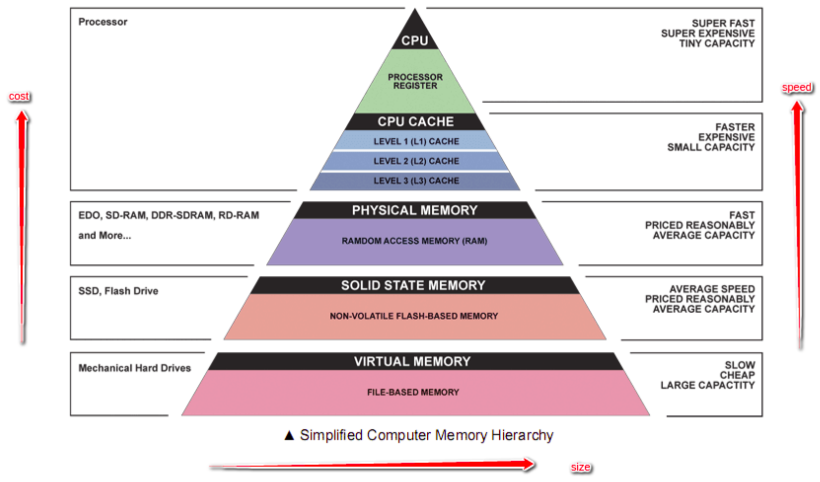
- Volatile Storage (Temporary):

Example: Random Access Memory (RAM)
  - Data is lost when power is turned off
  - Very fast read/write speed
  - Used for active processes and temporary data storage
  - Purpose: Speed up processing and improve system performance

- Non-Volatile Storage (Persistent):

Examples: Solid State Drives (SSD), Hard Disk Drives (HDD)
  - Data retained even when the system is powered off
  - Slower than RAM but offers permanent data storage
  - Used for long-term data retention

STORAGE HIERARCHY



- Primary Storage:
  - Located inside the CPU or directly accessible by it
  - Examples: Registers, Cache, RAM
  - Purpose: Immediate access for processing
- Secondary Storage:
  - External to CPU, used for long-term data retention
  - Examples: SSDs, HDDs
  - Purpose: Store operating systems, applications, and user data
- Tertiary Storage:
  - Used for archiving and backups
  - Examples: Magnetic tapes, Optical discs, Cloud archival storage
  - Purpose: Disaster recovery and long-term data preservation

STORAGE TIERS PURPOSE

Performance: Ensures fast access to frequently used data
Persistence: Retains information beyond power cycles
Protection: Provides data redundancy, backup, and recovery options

Tier	Example	Use Case	Cost	Speed
Tier 0	NVMe	Critical databases	$$$	Fastest
Tier 1	SSD	Production workloads	$$	High
Tier 2	SAS HDD	File servers	$	Medium
Tier 3	Tape / Cloud Archive	Backups	$-	Slow

STORAGE METRICS

- IOPS: Input/Output Operations per Second
- Throughput: MB/s or GB/s of data transfer
- Latency: Delay in I/O completion
- Capacity: Total usable space
- Utilization: % of used resources

IOPS (INPUT/OUTPUT OPERATIONS PER SECOND)

- Meaning: Measures how many read/write operations a storage device can perform in one second.

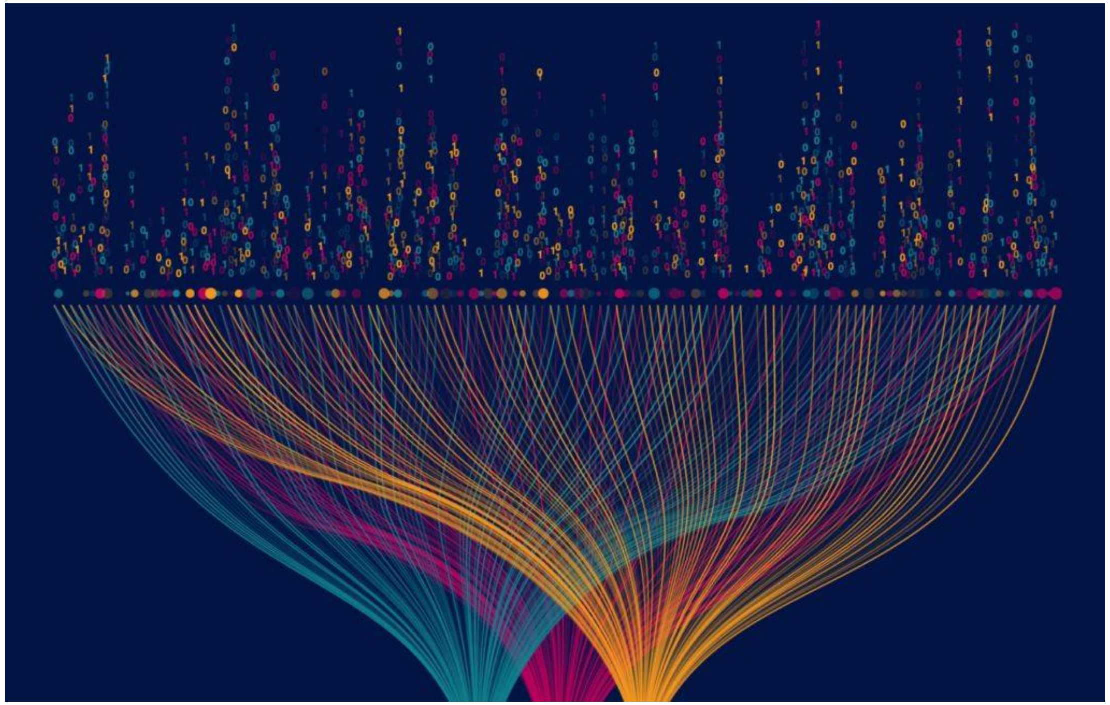
- Usage: Critical for workloads with many small, random I/O requests, such as databases and virtualization.

- Comparison:
  - HDD: ~100 IOPS
  - SSD: ~100,000 IOPS
  - NVMe SSD: ~1,000,000 IOPS
- Goal: Higher IOPS = Faster access to data and better performance for transaction-heavy systems.

THROUGHPUT (DATA TRANSFER RATE)

- Meaning: Measures how much data can be transferred per second.
- Units: MB/s or GB/s
- Best For: Sequential workloads like video editing or large file transfers.

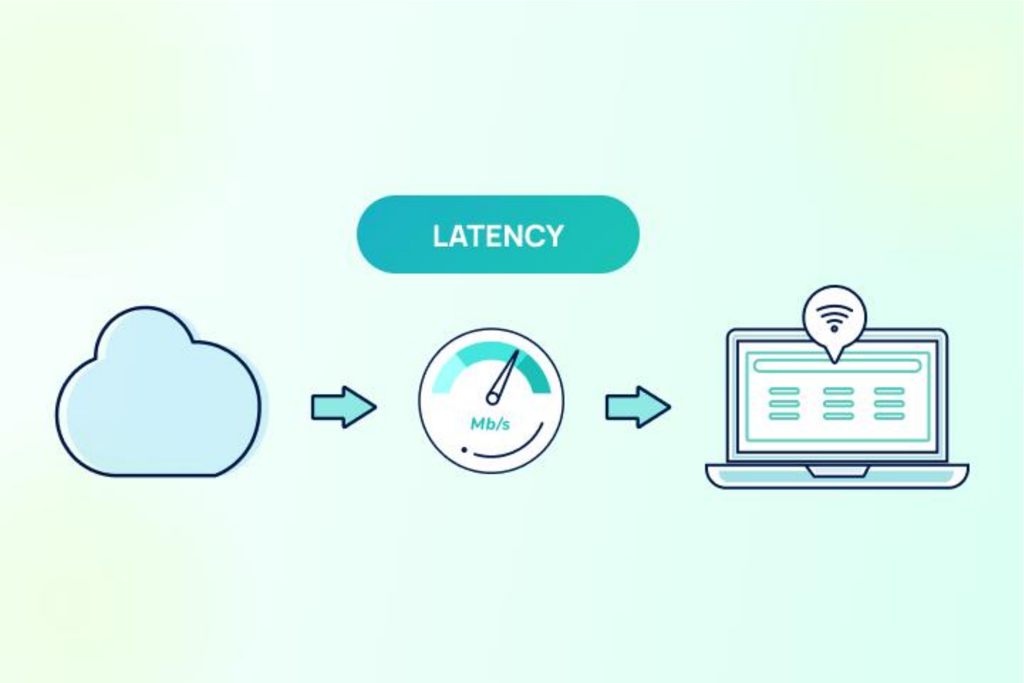
- Example:
  - HDD: ~150 MB/s
  - SSD: ~500 MB/s
  - NVMe: >3 GB/s



LATENCY (RESPONSE TIME)

- Meaning: Time delay between a request and completion of an I/O operation.
- Units: Measured in milliseconds (ms) or microseconds (µs).
- Lower latency means faster system response.

- Example:
  - HDD: ~5–10 ms
  - SSD: ~0.1 ms
  - NVMe: <0.02 ms



CAPACITY & UTILIZATION

- Capacity: Total amount of usable storage space available.
- Utilization: Percentage of storage or I/O resources currently being used
- Goal: Keep utilization below thresholds to avoid performance degradation.

POP QUIZ:

What is the main purpose of digital storage in computing systems?

A. To perform computations
B. To retain data for processing, retrieval, and backup
C. To transmit data across networks
D. To compress files

POP QUIZ:

What is the main purpose of digital storage in computing systems?

A. To perform computations
B. To retain data for processing, retrieval, and backup
C. To transmit data across networks
D. To compress files

POP QUIZ:

Which of the following is an example of volatile storage?

A. Hard Disk Drive (HDD)
B. Solid-State Drive (SSD)
C. Random Access Memory (RAM)
D. Optical Disc

POP QUIZ:

Which of the following is an example of volatile storage?

A. Hard Disk Drive (HDD)
B. Solid-State Drive (SSD)
C. Random Access Memory (RAM)
D. Optical Disc

POP QUIZ:

What is the primary function of tertiary storage?

A. Fast data access
B. Archiving and backups
C. Operating system boot-up
D. Temporary caching

POP QUIZ:

What is the primary function of tertiary storage?

A. Fast data access
B. Archiving and backups
C. Operating system boot-up
D. Temporary caching

POP QUIZ:

What is the main goal of monitoring utilization in storage systems?

A. To track user activity
B. To prevent overuse that degrades performance
C. To increase redundancy
D. To reduce hardware cost

POP QUIZ:

What is the main goal of monitoring utilization in storage systems?

A. To track user activity

B. To prevent overuse that degrades performance

C. To increase redundancy

D. To reduce hardware cost

STORAGE EVOLUTION & BUSINESS CONTEXT

STORAGE EVOLUTION TIMELINE

1950s-1980s:
Tape and
Drum Storage

- 1980s-1990s:
Hard Disk
Drives (HDD)

- 1990s-2000s:
RAID and
Storage
Networks

- 2000s-2010s:
Flash Storage
Emergence

- 2010s-Present:
All-Flash Arrays, NVMe,
Cloud Storage

- Future:
Storage Class
Memory, DNA
Storage

TRADITIONAL STORAGE CHALLENGES

- Performance Bottlenecks:
  - Spinning disk latency (5-15ms)
  - IOPS limitations
  - Unpredictable performance
- Management Complexity:
  - Multiple vendors and protocols
  - Manual provisioning
  - Capacity planning difficulties
- Reliability Issues:
  - Single points of failure
  - Complex disaster recovery
  - Long backup windows

BUSINESS VALUE OF STORAGE MODERNIZATION

- Agility: Faster provisioning and scaling for digital initiatives
- Resilience: Built-in redundancy, rapid recovery
- Cost Efficiency: Optimized use of tiers and cloud economics
- Compliance: Secure, auditable, and governed data lifecycle management

POP QUIZ:

What is a major challenge of traditional storage systems?

A. Too much automation
B. Limited redundancy and performance bottlenecks
C. Lack of physical space
D. Overuse of flash storage

POP QUIZ:

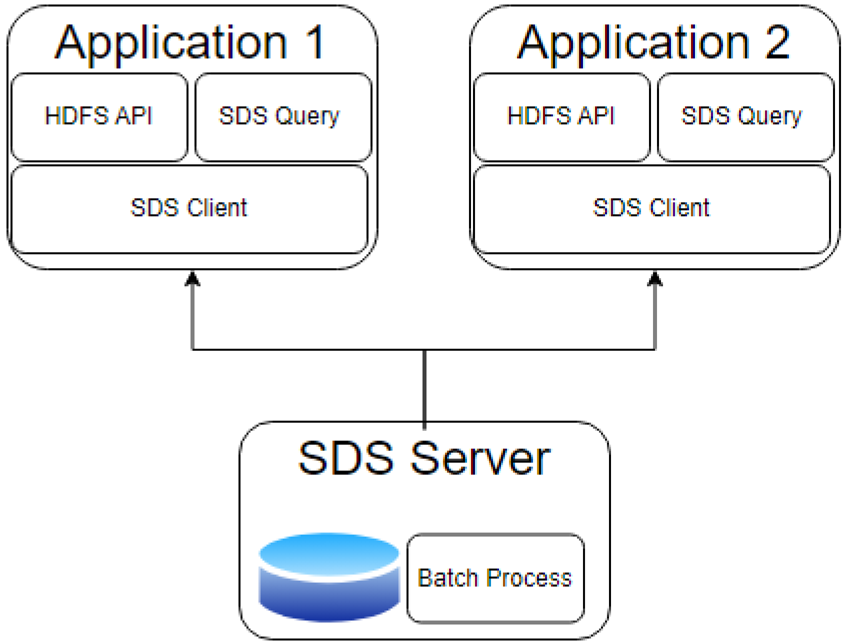
What is a major challenge of traditional storage systems?

A. Too much automation
B. Limited redundancy and performance bottlenecks
C. Lack of physical space
D. Overuse of flash storage

MODERN STORAGE ARCHITECTURES

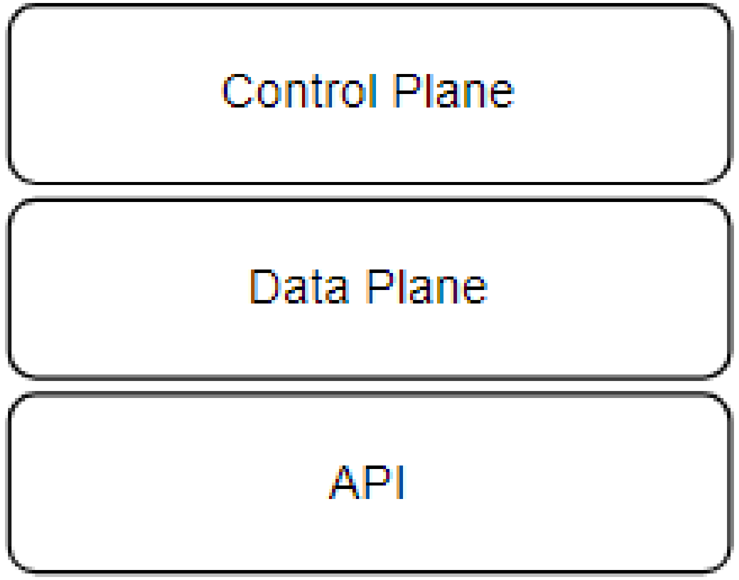
SOFTWARE-DEFINED STORAGE (SDS)

- Abstracts storage from physical hardware
- Managed via software intelligence
- Scales horizontally across commodity servers
- Examples: VMware vSAN, Ceph, Dell PowerFlex



SDS ARCHITECTURE OVERVIEW

- Control Plane: Manages policies, automation, and orchestration
- Data Plane: Handles actual I/O operations and data services
- Separation of Software & Hardware: Allows flexibility across vendor platforms
- APIs & Integration: Connects with cloud, hypervisors, and container platforms



SDS BENEFITS

- Hardware Independence: Works across commodity x86 servers
- Scalability: Add nodes seamlessly to increase capacity and performance
- Automation: Simplifies provisioning and lifecycle management
- Cost Efficiency: Reduces capital expenditure through commodity components
- Flexibility: Unified management across hybrid and multi-cloud environments

SDS CHALLENGES

- Complex Implementation: Requires careful integration and network design
- Performance Overheads: Virtualized layers may introduce latency
- Data Migration: Transitioning from legacy storage can be time-consuming
- Skills Gap: Need for expertise in automation, virtualization, and distributed systems
- Vendor Lock-in (Software Side): Proprietary SDS platforms can still limit portability

SDS USE CASES

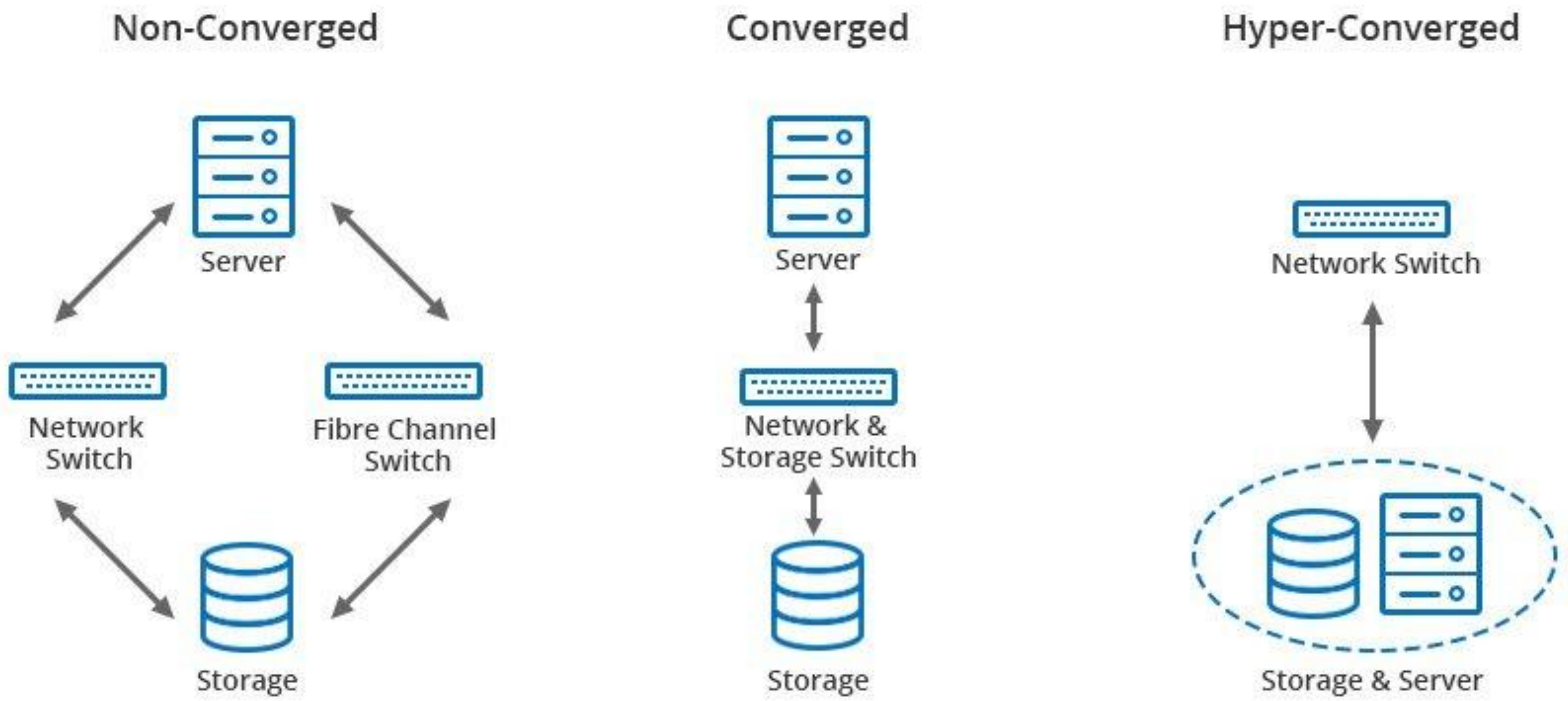
- Private Cloud Infrastructure: Dynamic storage for VMs and containers
- Backup & Disaster Recovery: Flexible, policy-based replication
- Big Data & Analytics: Scalable, high-throughput architecture
- Edge Deployments: Lightweight SDS nodes at distributed sites
- Test/Dev Environments: Rapid provisioning and teardown

HYPERCONVERGED INFRASTRUCTURE (HCI)

- Definition : Hyperconverged infrastructure is an approach to data storage that aims to minimize compatibility issues between storage systems, servers and network devices.

- HCI is a software defined IT infrastructure that includes:
  - Virtualized Computing
  - Software-defined Storage
  - Software-defined networking

NC VS CI VS HCI ARCHITECTURE



HCI BENEFITS

- Simplified Operations: Unified management reduces administrative overhead
- Scalability: Expand capacity by adding nodes
- Resilience: Built-in redundancy and fault tolerance
- Cost Efficiency: Reduces infrastructure complexity and vendor sprawl
- Optimized for Virtualization: Seamless with hypervisors and VM platforms

HCI CHALLENGES

- Initial Cost: Higher upfront for integrated nodes
- Vendor Lock-in: Proprietary ecosystems can limit flexibility
- Performance Overheads: Resource sharing may affect demanding workloads
- Network Dependency: Cluster stability relies heavily on inter-node communication
- Limited Customization: Tight hardware/software integration limits tuning

HCI USE CASES

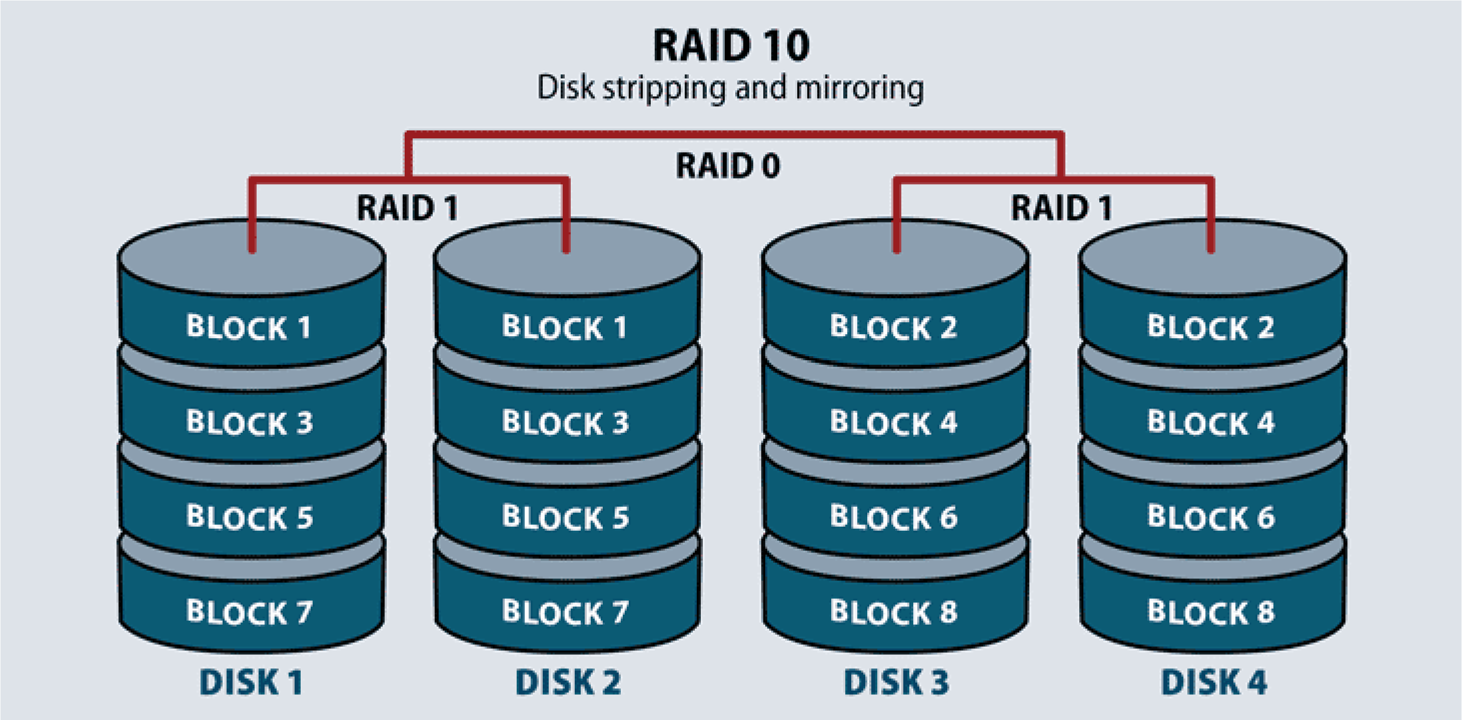
- Virtual Desktop Infrastructure (VDI): Simplifies provisioning and scaling desktops
- Private & Hybrid Cloud: Foundation for cloud-native and virtualized environments
- Edge Computing: Compact, self-contained infrastructure for remote locations
- Backup & DR: Integrated replication and failover capabilities
- Dev/Test Environments: Rapid deployment and simplified rollback

RAID STORAGE

- RAID = Redundant Array of Independent Disks is a data storage virtualization technology that combines multiple physical data storage components into one or more logical units for the purposes of data redundancy, performance improvement, or both

RAID ARCHITECTURE OVERVIEW

- Data Striping : Splits data across multiple disks to improve performance
- Mirroring : Duplicates data on separate disks for redundancy
- Parity: Stores error-checking information to rebuild data after failure

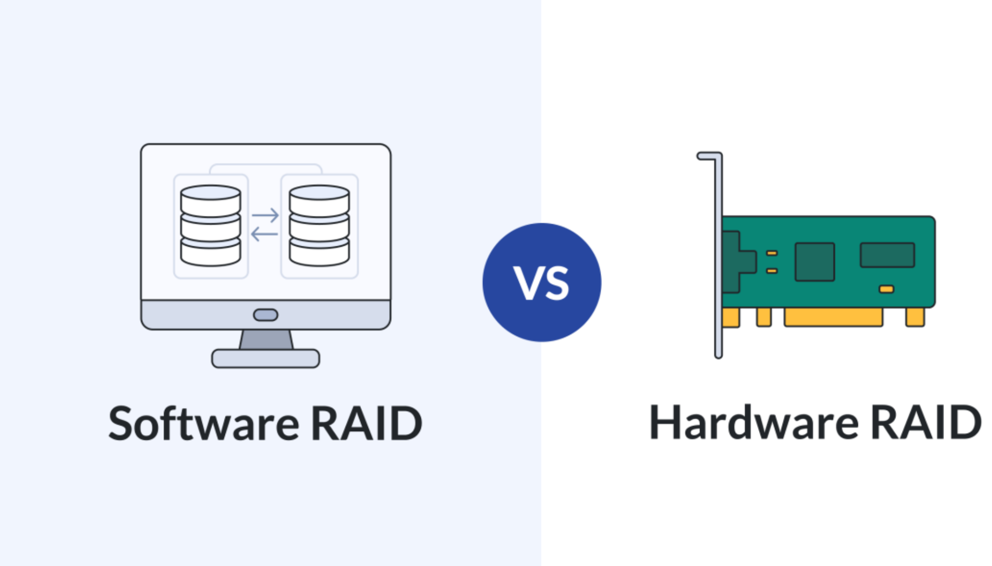


RAID LEVEL COMPARISON

RAID Level	Redundancy	Performance	Capacity Efficiency	Use Case
RAID 0	❌ None	✅ High	100%	Temporary data, cache
RAID 1	✅ Full	⚠ Moderate	50%	OS disks, critical systems
RAID 5	✅ Parity	✅ Balanced	~67%	File servers, general workloads
RAID 6	✅ Dual Parity	⚠ Slightly lower	~50%	Large capacity arrays
RAID 10	✅ Mirroring + Striping	✅ High	50%	Databases, high IOPS workloads

SOFTWARE VS HARDWARE RAID

- Software RAID:
  - Managed by the OS (e.g., Windows Storage Spaces, mdadm in Linux)
  - Cost-effective, flexible, but consumes CPU resources
- Hardware RAID:
  - Uses dedicated RAID controller cards
  - Higher performance, battery-backed cache, independent of OS
  - Hybrid Approaches: Smart NICs or NVMe RAID in enterprise systems



RAID LIMITATIONS

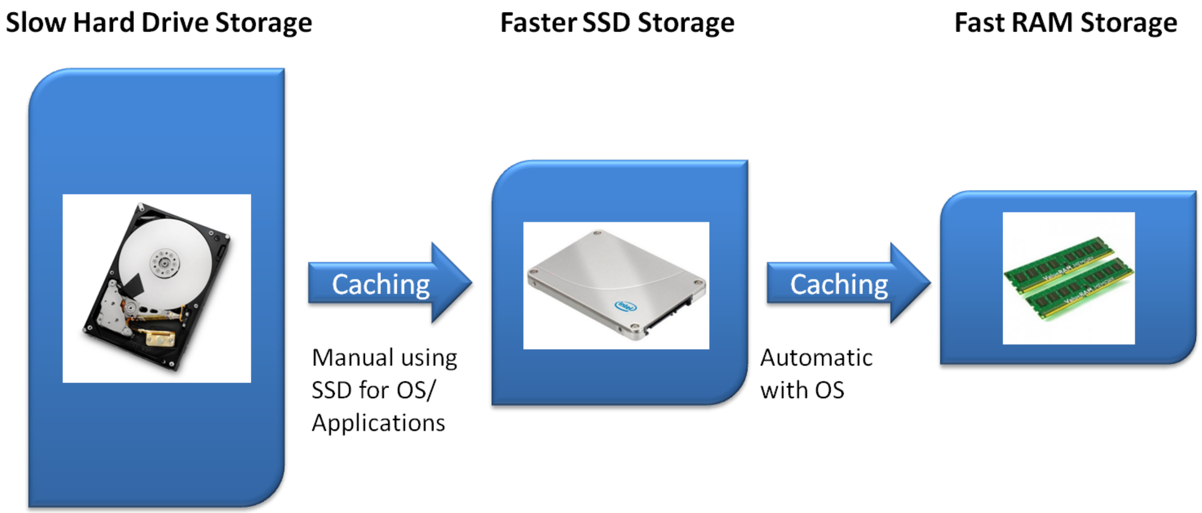
- No Protection Against Data Corruption or Ransomware
- Limited Scalability beyond fixed drive counts
- Rebuild Times Increase with large-capacity drives
- Controller Failures can still cause data loss
- Modern Alternatives: Distributed or object-based storage (e.g., Ceph, GlusterFS)

RAID IN MODERN STORAGE CONTEXT

- Still foundational in enterprise arrays (SAN/NAS)
- Integrated into virtualization and SDS platforms (vSAN, PowerFlex)
- Enhanced by erasure coding and data striping across nodes
- Transitioning toward cloud-native redundancy models

CACHING AND TIERING

- Caching: Stores hot data in faster media (SSD, DRAM)
- Tiering: Moves data dynamically between storage tiers
- Benefits:
  - Improves I/O performance
  - Reduces latency
  - Optimizes cost per GB



CACHING ARCHITECTURE

- Cache Layers:
  - L1: DRAM (ultra-fast, volatile)
  - L2: SSD (persistent, slower but larger)

- Cache Policies:
  - Write-through (safe, slower writes)
  - Write-back (fast, risk of data loss on failure)

- Cache Placement:
  - Host-level (server cache)
  - Array-level (storage controller cache)

TIERED STORAGE ARCHITECTURE

- Tier Types:
  - Hot Tier: SSD/NVMe for frequently accessed data
  - Warm Tier: High-performance HDDs for regular data
  - Cold Tier: Low-cost storage for infrequently accessed data
- Automated Tiering:
  - Monitors I/O patterns
  - Moves data between tiers dynamically

CACHING VS. TIERING

Aspect	Caching	Tiering
Purpose	Speed up frequent I/O	Optimize cost/performance balance
Data Lifespan	Short-term (temporary)	Long-term (persistent)
Media Type	DRAM, SSD	SSD, HDD, Cloud
Movement Trigger	Access frequency	Policy-based or automated
Examples	Read/write cache in SSDs	Auto-tiering in Dell EMC, NetApp

POP QUIZ:

What is the key concept behind SDS?

A. Hardware-managed storage
B. Software abstraction of storage from hardware
C. Centralized tape management
D. Manual provisioning

POP QUIZ:

What is the key concept behind SDS?

A. Hardware-managed storage
B. Software abstraction of storage from hardware
C. Centralized tape management
D. Manual provisioning

POP QUIZ:

What does HCI integrate into a single platform?

A. Only storage
B. Compute, storage, and networking
C. Networking and monitoring
D. Applications only

POP QUIZ:

What does HCI integrate into a single platform?

A. Only storage

B. Compute, storage, and networking

C. Networking and monitoring

D. Applications only

POP QUIZ:

What does RAID stand for?

A. Random Array of Indexed Drives
B. Redundant Array of Independent Disks
C. Repetitive Allocation of Indexed Data
D. Reliable Access for Integrated Databases

POP QUIZ:

What does RAID stand for?

A. Random Array of Indexed Drives
B. Redundant Array of Independent Disks
C. Repetitive Allocation of Indexed Data
D. Reliable Access for Integrated Databases

POP QUIZ:

What distinguishes tiering from caching?

A. Caching is long-term, tiering is temporary
B. Tiering moves data automatically between storage tiers
C. Tiering is faster than caching
D. Caching stores cold data

POP QUIZ:

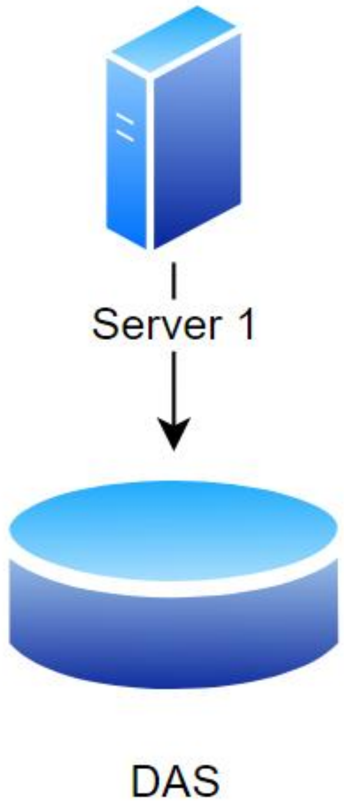
What distinguishes tiering from caching?

A. Caching is long-term, tiering is temporary
B. Tiering moves data automatically between storage tiers
C. Tiering is faster than caching
D. Caching stores cold data

STORAGE ARCHITECTURES IN PRACTICE

DAS - DIRECT ATTACHED STORAGE

- Definition: Storage directly connected to a single server
- Types:
  - Internal drives (SATA, SAS, NVMe)
  - External enclosures (USB, eSATA, SAS)
- Advantages:
  - High performance (no network overhead)
  - Simple configuration
  - Low cost for single server
- Disadvantages:
  - No sharing between servers
  - Limited scalability
  - Single point of failure

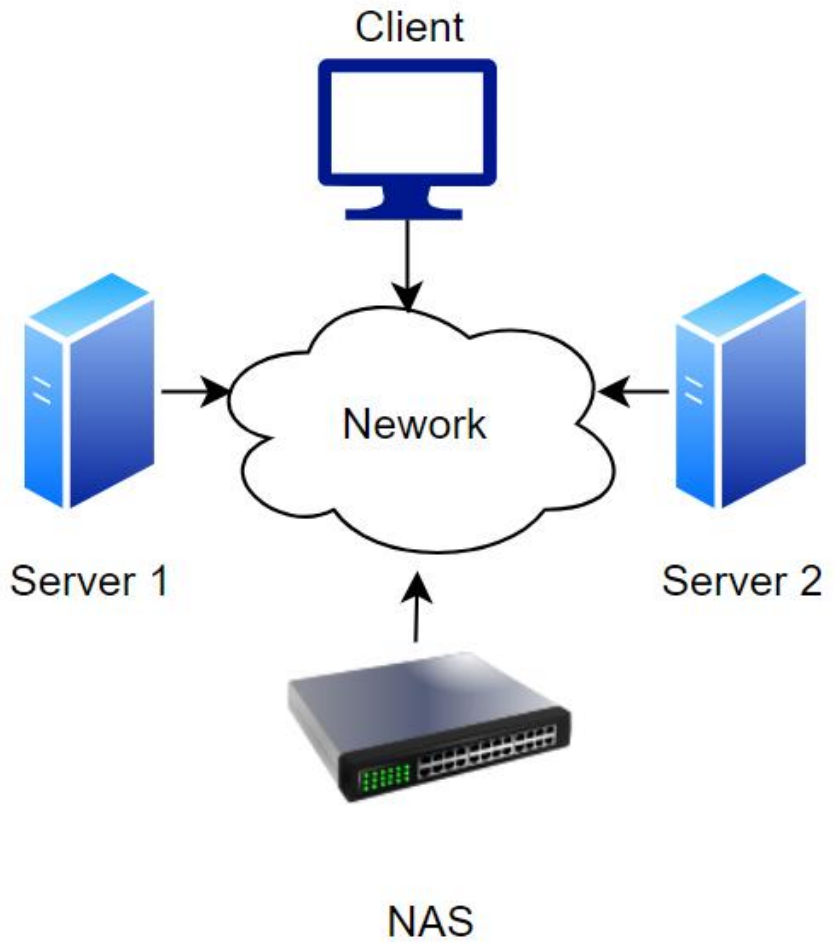


DAS IMPLEMENTATION EXAMPLE

- Use Case: Database Server with High IOPS Requirements
- Configuration:
  - Server: Dell PowerEdge R750
  - Storage: 8x NVMe SSDs in RAID 10
  - Capacity: 16TB usable
  - Performance: 800,000 IOPS
- Benefits: Predictable low latency, dedicated resources
- Limitations: Cannot share with other applications

NAS - NETWORK ATTACHED STORAGE

- Definition: File-level storage accessible over network
- Key Characteristics:
  - File sharing protocols (NFS, SMB/CIFS)
  - Network connectivity (Ethernet)
  - Multiple simultaneous clients
- Components:
  - NAS head (controller)
  - Storage pools
  - Network interfaces
  - File system



NAS PROTOCOLS DEEP DIVE

- NFS (Network File System):
  - UNIX/Linux native
  - Versions: NFSv3, NFSv4, NFSv4.1
  - Stateless protocol
- SMB/CIFS (Server Message Block):
  - Windows native
  - Versions: SMB1, SMB2, SMB3
  - Stateful protocol
- AFP (Apple Filing Protocol):
  - macOS native (deprecated)

NAS ARCHITECTURE COMPONENTS

- NAS Head:
  - CPU and memory for file operations
  - Multiple network ports (1GbE, 10GbE, 25GbE)
  - Storage controllers
- Storage Backend:
  - Disk shelves or flash arrays
  - RAID configurations
  - Hot spare drives
- File System:
  - Manages files and directories
  - Handles permissions and quotas
  - Provides data services (snapshots, replication)

NAS USE CASES & BENEFITS

Primary Use Cases:

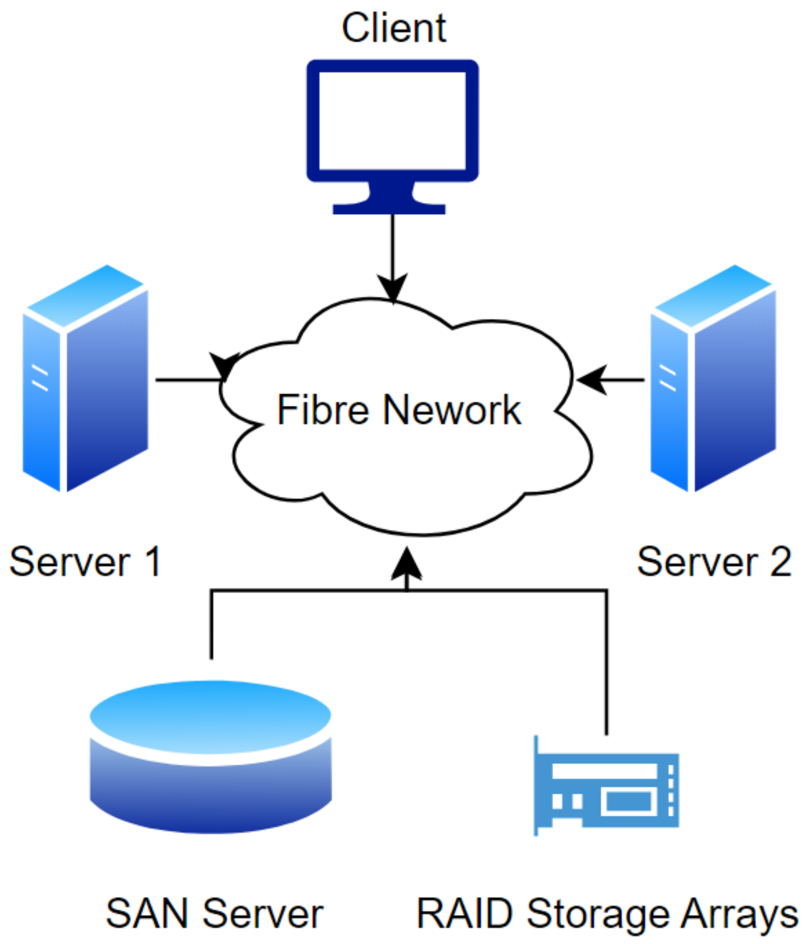
- File shares and home directories
- Content repositories
- Backup targets
- Development environments

Benefits:

- Easy to manage and provision
- Cross-platform file sharing
- Built-in data services
- Good price/performance for file workloads

STORAGE AREA NETWORK

- Definition: Block-level storage network connecting servers to storage
- Key Characteristics:
  - Dedicated storage network
  - Block-level access
  - High performance and availability
  - Centralized management
- Components:
  - Storage arrays
  - SAN switches
  - Host Bus Adapters (HBAs)
  - Storage fabric

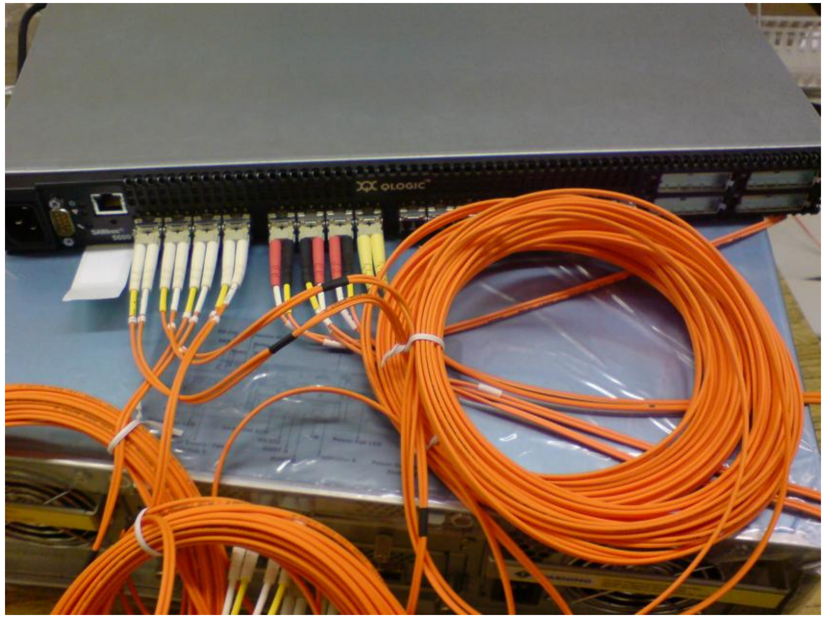


SAN TECHNOLOGIES

- Fibre Channel (FC):
  - Speeds: 8, 16, 32, 64, 128 Gbps
  - Low latency, high reliability
  - Dedicated network infrastructure
- iSCSI (Internet SCSI):
  - SCSI over Ethernet
  - Lower cost than FC
  - Uses existing network infrastructure
- FCoE (Fibre Channel over Ethernet):
  - FC frames over Ethernet
  - Converged network infrastructure

FIBRE CHANNEL DEEP DIVE

- Architecture:
  - Point-to-point, switched fabric, or arbitrated loop
  - WWN (World Wide Name) addressing
  - Zoning for security and performance
- Components:
  - FC switches (directors, edge switches)
  - HBAs in servers
  - Storage controllers
- Advantages:
  - Deterministic performance
  - Low CPU overhead
  - High availability features



ISCSI IMPLEMENTATION

- Components:
  - iSCSI initiator (client)
  - iSCSI target (storage)
  - Ethernet network
- Addressing:
  - IQN (iSCSI Qualified Name)
  - IP-based addressing
- Performance Considerations:
  - Network bandwidth and latency
  - TCP/IP overhead
  - Jumbo frames

POP QUIZ:

What best defines Direct Attached Storage (DAS

A. Storage connected over a network
B. Storage directly attached to a single server
C. Cloud-based shared storage
D. Virtualized storage over IP

POP QUIZ:

What best defines Direct Attached Storage (DAS

A. Storage connected over a network

B. Storage directly attached to a single server

C. Cloud-based shared storage

D. Virtualized storage over IP

POP QUIZ:

NAS provides storage access at which level?

A. Block level
B. File level
C. Object level
D. Memory level

POP QUIZ:

NAS provides storage access at which level?

A. Block level
B. File level
C. Object level
D. Memory level

POP QUIZ:

What is a key advantage of SAN over NAS?

A. Higher latency
B. Dedicated high-performance storage fabric
C. File-based access
D. Easier to configure

POP QUIZ:

What is a key advantage of SAN over NAS?

A. Higher latency

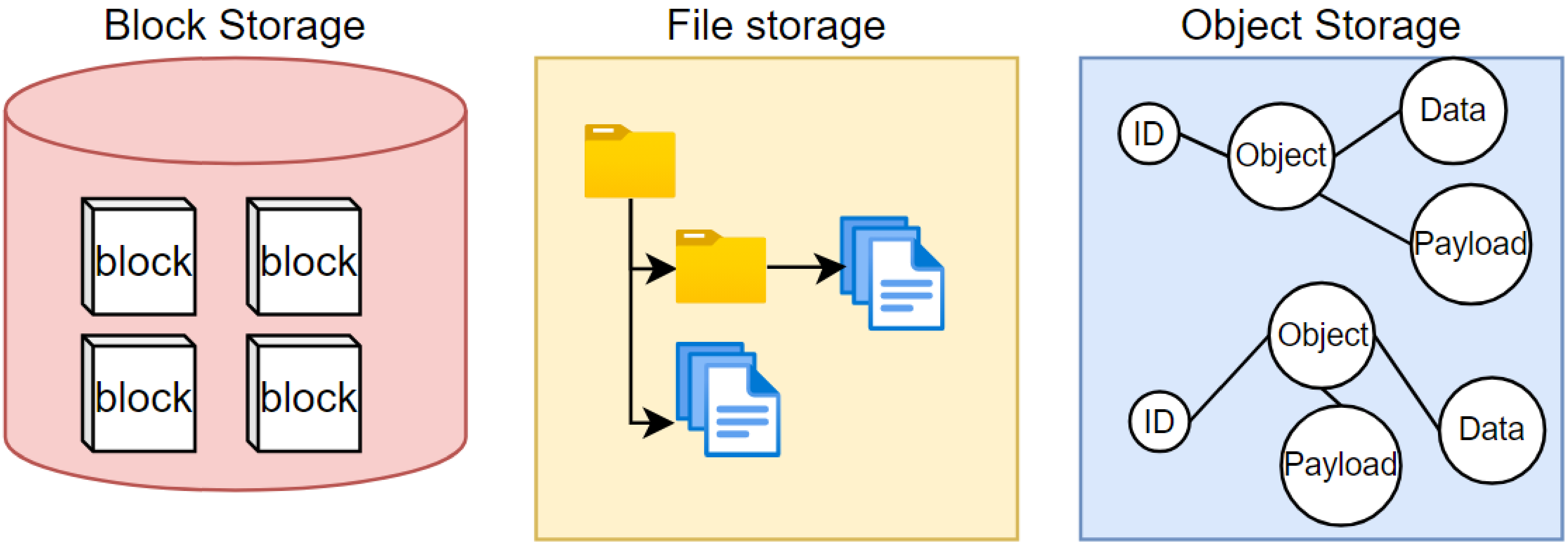
B. Dedicated high-performance storage fabric

C. File-based access

D. Easier to configure

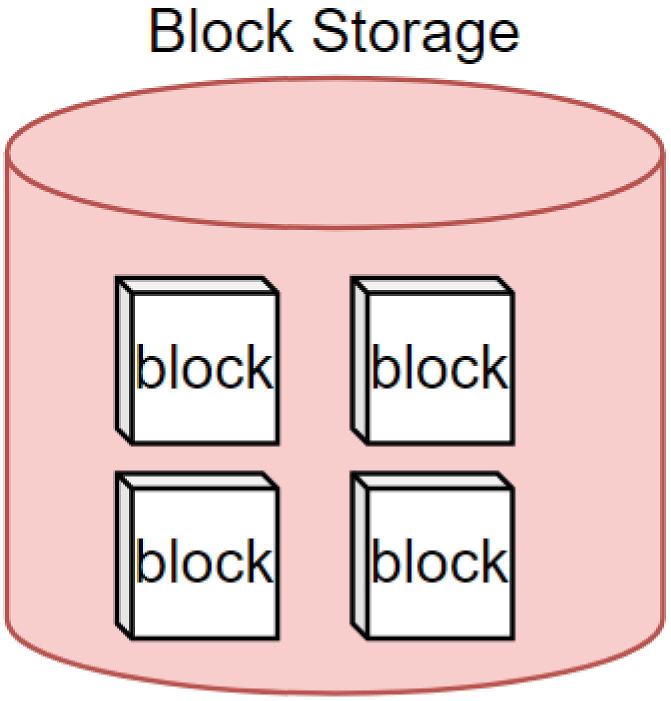
STORAGE TYPES & SELECTION

BLOCK VS FILE VS OBJECT STORAGE



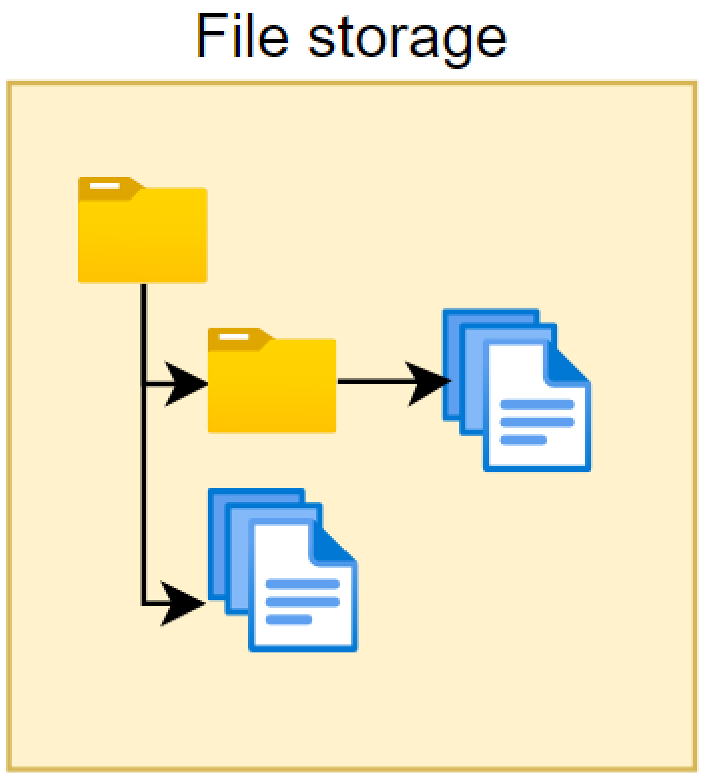
BLOCK STORAGE CHARACTERISTICS

- Features:
  - Fixed-size blocks (512B, 4KB, etc.)
  - Direct attachment to operating system
  - High performance for databases
  - Supports any file system
- Use Cases:
  - Database storage
  - Virtual machine storage
  - High-performance applications
- Limitations:
  - Requires file system management
  - Limited sharing capabilities



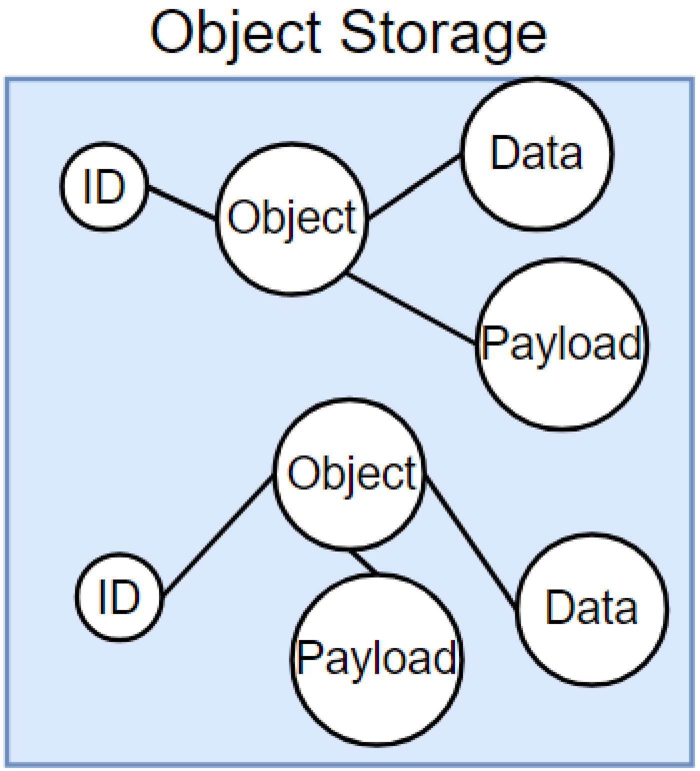
FILE STORAGE CHARACTERISTICS

- Features:
  - Hierarchical directory structure
  - File-level permissions and metadata
  - Network accessibility
  - Built-in sharing protocols
- Use Cases:
  - User home directories
  - Shared application data
  - Content management
  - Backup and archival
- Advantages:
  - Easy to manage and share
  - Cross-platform compatibility
  - Built-in data services



OBJECT STORAGE CHARACTERISTICS

- Features:
  - Flat namespace (no directories)
  - Rich metadata
  - REST API access
  - Virtually unlimited scalability
- Use Cases:
  - Cloud applications
  - Content distribution
  - Data archival and backup
  - Big data analytics
- Advantages:
  - Massive scalability
  - Built-in redundancy
  - Application integration



STORAGE SELECTION CRITERIA

- Performance Requirements:
  - IOPS and throughput needs
  - Latency sensitivity
  - Concurrent user count
- Capacity and Growth:
  - Current and future capacity
  - Growth rate and patterns
- Availability and Reliability:
  - Uptime requirements
  - Disaster recovery needs
  - Data protection requirements

Amazon S3

Azure Storage

POP QUIZ:

Which type of storage is best for databases requiring high IOPS?

A. File storage
B. Object storage
C. Block storage
D. Tertiary storage

POP QUIZ:

Which type of storage is best for databases requiring high IOPS?

A. File storage
B. Object storage
C. **Block storage**
D. Tertiary storage

POP QUIZ:

Which of the following is true about object storage?

A. Uses directories and paths
B. Provides virtually unlimited scalability
C. Offers lowest latency for small files
D. Is limited to local access

POP QUIZ:

Which of the following is true about object storage?

A. Uses directories and paths

B. Provides virtually unlimited scalability

C. Offers lowest latency for small files

D. Is limited to local access

LAB 1 : CONFIGURE ISCSI STORAGE

- Set up Linux iSCSI target
- Configure Windows iSCSI initiator
- Create and mount iSCSI LUN
- Test block-level access

LAB 2 : SET UP NFS FILE SHARING

- Configure NFS server on Linux
- Export file systems with different permissions
- Mount NFS shares on multiple clients
- Test concurrent file access