Storage Systems - Comprehensive Notes

1. Storage Definition

What is Storage?

Storage refers to the retention of retrievable data on computer and electronic systems. It encompasses both the physical media that holds data and the logical systems that manage data access, organization, and retrieval.

Types of Storage Media

- **Primary Storage**: RAM, cache memory (volatile, fast access)
- **Secondary Storage**: Hard drives, SSDs (non-volatile, persistent)
- **Tertiary Storage**: Tape drives, optical media (archival, slow access)

Key Storage Characteristics

- **Capacity**: Amount of data that can be stored
- **Performance**: Speed of read/write operations (IOPS, throughput)
- **Reliability**: Data integrity and system uptime
- **Scalability**: Ability to expand storage as needs grow

2. RAID Technology

RAID Overview

RAID (Redundant Array of Independent Disks) combines multiple physical drives into logical units to improve performance, reliability, or both.

Common RAID Levels

RAID 0 (Striping)

- **Purpose**: Performance improvement
- **Minimum drives**: 2
- **Redundancy**: None
- **Use case**: High-performance applications where data loss is acceptable

RAID 1 (Mirroring)

- **Purpose**: Data redundancy
- **Minimum drives**: 2
- **Capacity**: 50% of total drive space

- **Use case**: Critical data that requires high availability

RAID 5 (Striping with Parity)

- **Purpose**: Balance of performance and redundancy
- **Minimum drives**: 3
- **Fault tolerance**: Can lose 1 drive
- **Capacity**: (n-1) drives worth of space

RAID 6 (Double Parity)

- **Purpose**: Higher fault tolerance
- **Minimum drives**: 4
- **Fault tolerance**: Can lose 2 drives
- **Capacity**: (n-2) drives worth of space

RAID 10 (1+0)

- **Purpose**: High performance with redundancy
- **Minimum drives**: 4
- **Combines**: RAID 1 mirroring with RAID 0 striping
- **Use case**: Database servers, high-transaction systems

RAID Implementation

- **Hardware RAID**: Dedicated RAID controller
- **Software RAID**: OS-level implementation
- **Hybrid RAID**: Combination of hardware and software

3. Direct Attached Storage (DAS)

Definition

DAS refers to storage devices directly connected to a single computer or server without going through a storage network.

Connection Types

- **SATA**: Consumer-grade, up to 6 Gbps
- **SAS**: Enterprise-grade, up to 12 Gbps
- **NVMe**: High-performance, up to 32 Gbps
- **USB/Thunderbolt**: External connections

Advantages

- High performance (no network overhead)
- Simple setup and management
- Lower cost for single-system deployments
- Full control over storage resources

Disadvantages

- Limited scalability
- No sharing between multiple systems
- Single point of failure
- Difficult to manage across multiple servers

Use Cases

- Desktop workstations
- Single-server applications
- Development environments
- Small business solutions

4. Network Attached Storage (NAS)

Definition

NAS is file-level storage connected to a network, providing data access to multiple clients through standard network protocols.

Key Protocols

- **NFS**: Network File System (Unix/Linux)
- **SMB/CIFS**: Server Message Block (Windows)
- **AFP**: Apple Filing Protocol (macOS)
- **FTP/SFTP**: File Transfer Protocol

Architecture Components

- **NAS Head**: Controller with network interface
- **Storage Pool**: Underlying disk arrays
- **File System**: Manages file organization
- **Network Interface**: Ethernet connections

Advantages

- Easy file sharing across networks
- Centralized storage management
- Built-in backup and recovery features
- Cost-effective for file storage
- Simple installation and configuration

Disadvantages

- Performance limited by network bandwidth
- Not suitable for block-level applications
- File-level access only
- Potential network bottlenecks

Use Cases

- File servers
- Home media storage
- Small to medium business data sharing
- Backup and archival storage

5. Storage Area Network (SAN)

Definition

SAN is a dedicated high-speed network that provides block-level access to storage, appearing as locally attached storage to servers.

Key Protocols

- **Fibre Channel**: Traditional SAN protocol (8/16/32 Gbps)
- **iSCSI**: IP-based SAN over Ethernet
- **FCoE**: Fibre Channel over Ethernet
- **NVMe over Fabrics**: High-performance protocol

SAN Components

- **HBA**: Host Bus Adapters (server-side)
- **SAN Switches**: Fabric infrastructure
- **Storage Arrays**: Centralized storage systems
- **Management Software**: SAN administration tools

SAN Topologies

- **Point-to-Point**: Direct connection
- **Arbitrated Loop**: Shared loop topology
- **Switched Fabric**: Full mesh connectivity

Advantages

- High performance and low latency
- Centralized storage management
- Advanced features (snapshots, replication)
- Scalability and flexibility
- Multiple host access to shared storage

Disadvantages

- High cost and complexity
- Requires specialized expertise
- Complex zoning and LUN management
- Vendor lock-in potential

Use Cases

- Enterprise databases
- Virtualization environments
- High-performance computing
- Mission-critical applications

6. Security in SAN

Authentication Methods

- **CHAP**: Challenge Handshake Authentication Protocol
- **Kerberos**: Network authentication protocol
- **Certificate-based**: PKI authentication
- **RADIUS/LDAP**: Centralized authentication

Access Control

- **LUN Masking**: Restricts LUN visibility to specific hosts
- **Zoning**: Controls communication between SAN devices
- **Port Security**: Restricts access based on physical ports
- **RBAC**: Role-based access control

Encryption

- **Data-at-Rest**: Storage array encryption
- **Data-in-Transit**: Network-level encryption
- **Key Management**: Centralized key storage and rotation
- **Self-Encrypting Drives**: Hardware-level encryption

Security Best Practices

- Network segmentation and isolation
- Regular security audits and monitoring
- Firmware and software updates
- Strong password policies
- Multi-factor authentication
- Backup and disaster recovery planning

7. Backups and Restoration

Backup Types

- **Full Backup**: Complete copy of all data
- **Incremental**: Only changed data since last backup
- **Differential**: Changed data since last full backup
- **Synthetic Full**: Combines incremental backups

Backup Strategies

- **3-2-1 Rule**: 3 copies, 2 different media, 1 offsite
- **GFS**: Grandfather-Father-Son rotation
- **Tower of Hanoi**: Mathematical backup rotation
- **Continuous Data Protection**: Real-time backup

Recovery Types

- **File-level Recovery**: Individual files and folders
- **System Recovery**: Complete system restoration
- **Bare Metal Recovery**: OS and application restoration
- **Point-in-Time Recovery**: Restore to specific timestamp
- **Instant Recovery**: VM recovery from backup storage

Recovery Metrics

- **RTO**: Recovery Time Objective (downtime tolerance)
- **RPO**: Recovery Point Objective (data loss tolerance)
- **MTTR**: Mean Time to Recovery
- **MTBF**: Mean Time Between Failures

Backup Technologies

- **Tape Libraries**: Long-term archival storage
- **Disk-to-Disk**: Fast backup and recovery
- **Cloud Backup**: Offsite protection
- **Snapshot Technology**: Point-in-time copies
- **Deduplication**: Reduces storage requirements

8. Data Archiving Solutions

Archive vs. Backup

- **Backup**: Short-term recovery from recent failures
- **Archive**: Long-term retention for compliance/reference
- **Active Archive**: Searchable, accessible archived data
- **Cold Archive**: Infrequently accessed data

Archive Storage Tiers

- **Hot Storage**: Immediate access (SSD, fast HDD)
- **Warm Storage**: Moderate access (standard HDD)
- **Cold Storage**: Infrequent access (tape, cloud glacier)
- **Frozen Storage**: Rare access (offline media)

Archive Technologies

- **Tape Systems**: LTO (Linear Tape-Open) technology
- **Optical Storage**: Blu-ray, M-DISC for longevity
- **Cloud Archive**: Amazon Glacier, Azure Archive
- **Object Storage**: S3-compatible systems
- **Hierarchical Storage Management**: Automated tiering

Compliance Requirements

- **Legal Hold**: Preservation for litigation
- **Regulatory Compliance**: Industry-specific retention

- **Data Governance**: Policies and procedures
- **Chain of Custody**: Evidence preservation
- **Audit Trails**: Access and modification logging

Archive Best Practices

- **Data Classification**: Categorize by importance/sensitivity
- **Retention Policies**: Define lifecycle management
- **Migration Planning**: Technology refresh strategies
- **Integrity Checking**: Regular data verification
- **Metadata Management**: Searchable information catalog
- **Disaster Recovery**: Geographic distribution of archives

Key Takeaways

1. **Choose the right storage type** based on performance, capacity, and budget requirements
2. **Implement appropriate RAID levels** for your reliability and performance needs
3. **Consider network implications** when choosing between DAS, NAS, and SAN
4. **Security should be built-in** from the ground up, not added as an afterthought
5. **Plan for disaster recovery** with comprehensive backup and archive strategies
6. **Regular testing and maintenance** are crucial for all storage systems
7. **Monitor and optimize performance** continuously as workloads evolve