Software Defined Systems

Week 4 Software Defined Storage (SDS)

Lecturer: Biniam Behailu
Addis Ababa Science and Technology
Addis Ababa, Ethiopia

Contents

01 Understanding Data Explosion

02 Challenges of Traditional Storage Systems

03 Introduction to SDS

04 SDS Architecture

05 Use Cases for SDS

06 Popular SDS Tools

07 Future Trends in SDS

Software Defined Storage (SDS)

Learning objectives

- Define Software Defined Storage (SDS) and its significance.
- Identify sources and characteristics of data explosion.
- Analyze limitations of traditional storage systems.
- Recognize benefits and use cases of SDS.
- Identify challenges in implementing SDS and strategies to overcome them.

Data Explosion

- Data explosion refers to the exponential increase in the volume of data generated, collected, and stored in various forms across multiple platforms.
- Data can be generated from various sources, including social media, IoT devices, and enterprise applications.
- The total volume of data created worldwide is expected to reach a staggering 175 zettabytes by the year 2025. (source: IDC) [1]

Sources of Data Explosion

- **Social Media** - Massive user-generated content.
- **IoT Devices** - Billions of devices transmitting data continuously.
- **Enterprise Applications** - CRM and ERP systems producing extensive data.

Characteristics of Data Explosion

- **Volume** - Unprecedented amounts of data generated.
- **Velocity** - Real-time generation and processing.
- **Variety** - Structured, semi-structured, and unstructured data.

Traditional Storage Systems

- Traditional storage systems refer to legacy data storage technologies commonly used in enterprises before the rise of cloud and modern storage solutions.
- These systems are typically characterized by their physical infrastructure and specific technologies.

Traditional Storage Systems

- **Direct Attached Storage (DAS) - Storage directly connected to a server (e.g., internal hard drives). It offers high speed but lacks network access.**
  - **Hard Disk Drives** (HDDs) - Magnetic storage devices that use spinning disks to read/write data.
  - **Solid State Drives** (SSDs) - Faster than HDDs, SSDs use flash memory to store data.
  - **Tape Drives** - Used primarily for archival storage, tape drives are economical for long-term data retention but have slower access times compared to disk-based systems.

Traditional Storage Systems

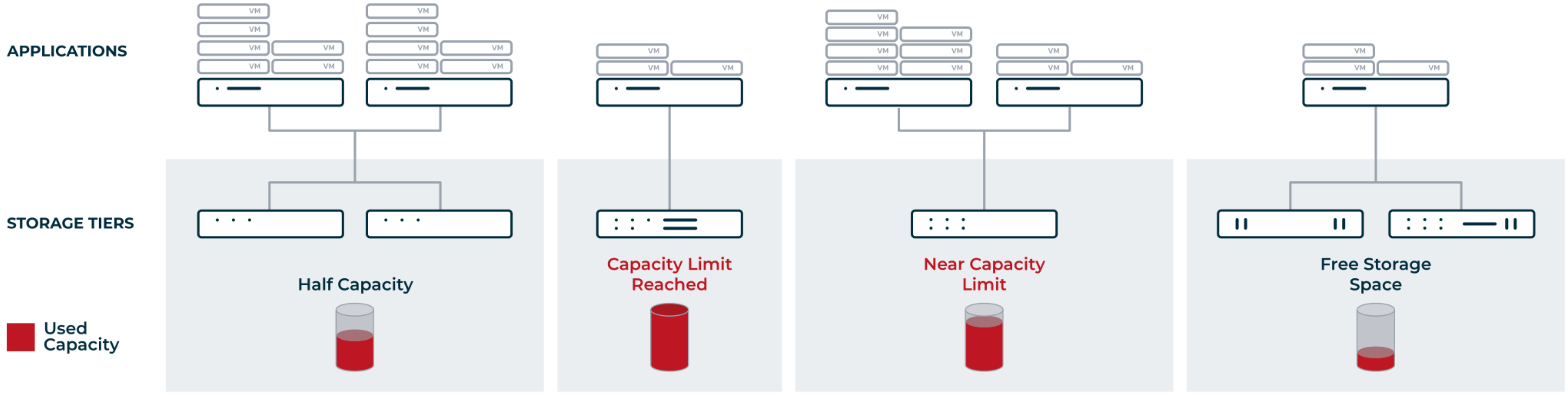
- **Networked Storage -** Includes NAS and SAN, providing shared access and scalability.
  - **Storage Area Networks (SAN) -** A dedicated network that provides access to consolidated block-level storage, often used for high-performance applications.
  - **Network Attached Storage (NAS) -** A file-level storage device connected to a network, allowing multiple users to access data concurrently.

Limitations of Traditional Storage Systems

- **Scalability Issues** - Difficult to scale horizontally.
- **Cost Inefficiency** - High capital and operational costs.
- **Performance Bottlenecks** - Struggles with high IOPS and latency.

Limitations of Traditional Storage Systems

- Isolated Storage Capacity Across Unlike Storage Systems



Source: https://www.datacore.com/software-defined-storage/

Software Defined Storage(SDS)

- Software Defined Storage (SDS) is an architecture that abstracts and virtualizes storage resources, allowing for more flexible and efficient management of data storage.

- This approach decouples the storage hardware from the software that manages it, enabling organizations to optimize their storage environments without being tied to specific hardware vendors.

SDS Key Components

- **Software Layer** - Manages resources and provides APIs.
- **Hardware Layer** - Utilizes various storage hardware.
- **Management Layer** - Tools for monitoring and orchestration.

SDS Architecture

Management Plane

Control Plane

Data Plane

Source: https://blogs.vmware.com/virtualblocks/2015/04/14/sds-the-missing-link-storage-automation-for-application-service-catalogs/

Benefits of SDS

- **Scalability** - Seamless addition of storage nodes.
- **Flexibility** - Supports multiple storage types and protocols.
- **Cost Efficiency** - Reduces capital and operational costs.
- **Improved Performance** - Optimizes resource utilization.

Understanding Data Storage

Understanding Data Storage

- Data storage is the method by which digital information is preserved for later use.
- Different storage architectures cater to various needs in terms of performance, scalability, and cost.
- Understanding these differences is key to choosing the right storage solution for a given task.
- There are different data storage architectures, such as object storage, block storage, and file storage.
- Each of these approaches has its own unique way of handling data, making them suitable for different purposes.

Object Storage "Everything in a Bucket" Approach

- Treats data as objects stored in a flat structure called a bucket.
- Each object contains the data, metadata, and a unique identifier.
- Accessed via HTTP/HTTPS using RESTful APIs.
- Focuses on scalability, durability, and metadata richness

Object Storage "Everything in a Bucket" Approach

- Treats data as objects stored in a flat structure called a bucket.
- Each object contains the data, metadata, and a unique identifier.
- Accessed via HTTP/HTTPS using RESTful APIs.
- Focuses on scalability, durability, and metadata richness

Block Storage "Building Blocks" Approach

- Divides data into fixed-size blocks, each with a unique identifier.
- Blocks are stored independently and can be located anywhere on the storage medium.
- The operating system manages the blocks and organizes them with a file system.
- Focuses on high performance, low latency, and random access

File Storage "Familiar Filing Cabinet" Approach

- Organizes data in a hierarchical structure of files and folders (directories).
- Data is accessed through file paths (e.g., /home/user/documents/report.docx).
- Uses file-level protocols like NFS and SMB/CIFS for sharing.
- Focuses on simplicity, organization, and file sharing.)

SDS Use Cases

SDS Use Cases

- **Cloud Storage** - Efficient management of cloud resources.
- **Big Data Analytics** - Supports large-scale data processing.
- **Virtual Desktop Infrastructure (VDI)** - Simplifies management for virtual desktops.
- **Edge Computing** – Reduces the distance data must travel, enhancing performance.

Cloud Storage

- Cloud storage is a service that allows individuals and organizations to store data remotely on servers managed by third-party providers.
- This technology offers several advantages, including scalability, accessibility, and cost-effectiveness.
- Users can access their data from any device with an internet connection, facilitating collaboration and remote work.
- The types of cloud solutions include public, private, and hybrid clouds.

Big Data and SDS

- Big Data refers to the vast volumes of structured and unstructured data that organizations generate, collect, and analyze.
- Managing this data effectively is crucial for deriving insights and making informed decisions.
- Software Defined Storage (SDS) plays a vital role in this context by providing a flexible and scalable storage solution tailored to the needs of big data environments.

VDI and SDS

- Virtual Desktop Infrastructure (VDI) is a technology that allows users to access desktop environments hosted on centralized servers rather than on local machines.
- Software Defined Storage (SDS) complements VDI by providing a flexible and scalable storage solution that can adapt to the varying demands of virtual desktop environments.
- With SDS, the storage resources are abstracted from the hardware, allowing for dynamic allocation and management based on usage patterns.

SDS and Edge Computing

- Edge computing refers to the practice of processing data closer to its source, rather than relying solely on centralized data centers.
- This approach minimizes latency, reduces bandwidth usage, and improves response times, making it ideal for applications that require real-time data processing and analysis.

Popular SDS Tools

Ceph

- Ceph is an open-source, distributed storage system designed to provide highly scalable object, block, and file storage in a unified platform.
- It is widely used in cloud environments and big data applications.

VMware vSAN

- VMware vSAN (Virtual SAN) is a software-defined storage solution integrated with VMware vSphere.
- It enables organizations to create a highly scalable and resilient storage infrastructure that leverages existing server hardware.

Red Hat Gluster Storage

- Red Hat Gluster Storage is a scalable, software-defined storage solution designed for handling large amounts of unstructured data.
- It provides a flexible and robust platform suitable for various workloads, including cloud storage, big data, and containerized environments.

OpenStack Swift

- OpenStack Swift is an object storage system that is part of the OpenStack cloud computing platform.
- It is designed to store and manage large amounts of unstructured data, providing a highly scalable, durable, and secure storage solution.

IBM Spectrum Scale

- IBM Spectrum Scale (formerly known as GPFS - General Parallel File System) is a high-performance, scalable file storage solution designed for managing large volumes of data in a distributed environment.
- It is suitable for various workloads, including big data analytics, cloud storage, and high-performance computing (HPC).

Popular SDS Tools Comparison

Feature/Tool	Ceph	VMware vSAN	Red Hat Gluster Storage	OpenStack Swift	IBM Spectrum Scale
Architecture	Distributed object and block storage	Hyper-converged	Distributed file system	Object storage	Parallel file system
Scalability	Elastic scalability	Horizontal scaling	Horizontal scaling	Petabyte-scale	Petabyte-scale
Performance	High performance for varied workloads	High performance for VMs	Good for large files	Optimized for unstructured data	High throughput for HPC
Management	Ceph Dashboard, CLI	Integrated with vSphere	Centralized management	Open API for integration	Comprehensive tools
Cost	Open-source	Requires VMware ecosystem	Commodity hardware	Open-source	Enterprise license
Best Use Cases	Cloud storage, virtualization, and big data	Virtual desktops, VDI	Cloud storage, big data	Cloud-native apps	Big data, HPC

Integration with Existing Infrastructure

- Integrating Software Defined Storage (SDS) with existing infrastructure involves several key steps to ensure a seamless transition and optimal performance.

- In order to effectively implement SDS, organization should focus on:

Assessment of Current Infrastructure

Choosing the Right SDS Solution

Phased Implementation

Data Migration Strategy

Training and Support

Monitoring and Optimization

Disaster Recovery with SDS

- Disaster recovery (DR) is a critical aspect of IT strategy, ensuring data availability and business continuity in the event of unexpected disruptions.
- Software Defined Storage (SDS) plays a pivotal role in enhancing disaster recovery capabilities.

Security in SDS

- Security is a critical aspect of Software Defined Storage (SDS), ensuring that data is protected against unauthorized access, breaches, and loss.
- Key features include data encryption both at rest and in transit, ensuring that information remains secure during storage and transmission.
- Access control mechanisms, such as role-based access control (RBAC) and multi-factor authentication (MFA), help restrict data access to authorized users only.

Cost Analysis of SDS

- One of the primary advantages of SDS is its ability to leverage commodity hardware, which significantly reduces capital expenditures compared to traditional storage solutions that often require proprietary equipment.
- This reduction in hardware costs is complemented by the scalability of SDS, allowing organizations to expand storage capacity as needed without major upfront investments.

Vendor Landscape

- The Software Defined Storage (SDS) market features a diverse range of vendors offering various solutions tailored to different organizational needs.
- Some key players includes Vmware, RedHat, OpenStack, IBM, Dell, IBM

SDS in Multi-Cloud Environments

- SDS enables organizations to manage storage resources across multiple cloud providers from a single interface.
- This unified management simplifies operations, reduces complexity, and allows for consistent policies and practices across different environments.
- With SDS, data can be easily moved between on-premises and various cloud environments.

Future of SDS

- SDS solutions are expected to incorporate artificial intelligence (AI) and machine learning (ML) for enhanced data management.
- These technologies can optimize storage allocation, predict performance issues, and automate routine tasks, leading to greater efficiency and reduced operational overhead.

SDS and Data Privacy

- Organizations face significant hurdles in understanding and complying with various data privacy regulations.
- To mitigate these challenges, organizations should adopt best practices such as implementing robust encryption methods and strict access controls.
- Additionally, access controls limit data availability to authorized personnel only, reducing the risk of data breaches and ensuring compliance with privacy regulations.

Software Defined Storage Advantages

- With SDS, businesses can meet the ever-increasing demands of modern data-intensive applications.
    - Improved data security
    - More efficient data management
    - High availability of stored information
    - Enhanced IT disaster recovery capabilities

Test Your Knowledge

1. Which of the following is a benefit of using SDS?

    A) Increased hardware dependence

    B) Simplified data management

    C) Manual resource allocation

    D) Rigid architecture

Test Your Knowledge

1. Which of the following is a benefit of using SDS?

   A) Increased hardware dependence

   B) Simplified data management

   C) Manual resource allocation

   D) Rigid architecture

Reason: SDS simplifies data management by abstracting storage resources from hardware, allowing for flexible and automated handling. This reduces manual configurations, enabling quicker responses to changing needs.

Test Your Knowledge

2. VMware vSAN is best known for its integration with which type of infrastructure?

   A) Virtualization platforms

   B) Physical servers

   C) Legacy storage systems

   D) Network Attached Storage (NAS)

Test Your Knowledge

2. VMware vSAN is best known for its integration with which type of infrastructure?

A) Virtualization platforms

B) Physical servers

C) Legacy storage systems

D) Network Attached Storage (NAS)

Reason: VMware vSAN is designed for seamless integration with VMware's virtualization solutions, allowing organizations to manage compute and storage as a unified system.

References

[1] A. Patrizio, "IDC: Expect 175 zettabytes of data worldwide by 2025," Network World, Dec. 3, 2018. [Online]. Available: https://www.networkworld.com/article/966746/idc-expect-175-zettabytes-of-data-worldwide-by-2025.html. [Accessed: Apr. 10, 2025].

Thank you!

Lecturer: Biniam Behailu
Addis Ababa Science and Technology
Addis Ababa, Ethiopia