A Survey on Different Storage Architectures

[1]Rahul Jyoti,[2]Saumitra Kulkarni and [3]Prof. Kirti Wanjale
Department of Computer Engineering
Vishwakarma Institute of Information Technology, Pune, Maharashtra, India

----------***----------

Abstract

This paper surveys different types of data storage architectures used in today's world to store the data of various types and presents an in-depth analysis of the use cases and drawbacks of the various storage architectures in different scenarios. We also examine the limitations of the traditional storage architectures in today's world and also discuss the modern solutions to these problems.
In this survey paper, we have provided a detailed comparison of three different storage architectures namely file storage, block storage, and object storage. This paper provides sufficient information to the reader to be able to choose among these architectures for storing their data as per the use case.

Keywords: *Data storage, Object storage, Scalable storage, Unstructured data storage, Types of storage*

1. Introduction:

In today's world, the rate at which digital data is generated and consumed every day is massive and is expected to grow at an even faster rate in the future. According to the IDC's (International Data Corporation) findings, *"The amount of digital data created over the next five years i.e 2020-2025 will be greater than twice the amount of data created since the advent of digital storage"*. And with the recent advancements in AI, IoT, and cloud technologies data is expected to grow at a huge scale and especially unstructured data. Infact, data creation and consumption have experienced an unusual spike in the year 2020 due to the outbreak of Covid-19 pandemic all over the globe which forced people to stay and work from home and hence leading to a huge amount of data consumption and creation. But the interesting fact is, out of all the new data generated during this pandemic less than 2% of this new data is saved or retained in 2021. This means the vast portion of data that is created is either not stored or overwritten or maybe the storage infrastructure and different storage techniques are not sufficient to store such a huge amount of data.

As per the IDC's reports, *"the data is expected to grow at a CAGR (Compound Annual Growth Rate) of 23% and the storage capacity is expected to grow at a CAGR of 19.2% over the 2020-2025 forecast period"*. This suggests that there is a gap between the amount of data generated and the storage capacity available to store digital data. Not all the data is indeed meant to be stored, different types of data have different use cases and should be stored in a way favorable to the use case. Data storage was quite simple once upon a time when the majority of data generated was structured data. But the exponential growth of unstructured data in recent times has produced another set of challenges to store it efficiently. The traditional architecture used to store the structured data is not efficient enough to store the unstructured data both in terms of cost and performance. There are many use cases in the modern era like big data analytics, image processing, image classification, natural language processing, video analytics, etc. which require unstructured data and that too in huge numbers for better performance of the machine learning models. Hence it is inevitable that we need a better mechanism to store unstructured data efficiently.

There are several data storage architectures present to store data in different ways. In this paper we will be primarily discussing the major three storage architectures namely:

1. File storage

2. Block storage
3. Object storage

All the major cloud service providers in the market provide different kinds of storage services based on these three different architectures. For example, AWS (Amazon Web Services) provides three different services namely EFS (Elastic File System), EBS (Elastic Block Store), and S3 (Simple Storage Service) based on the file storage, block storage, and object storage architectures respectively.

In the following section, we will provide a more comprehensive and detailed analysis of each of these three different storage architectures and will conclude the optimal choice for storage architecture in different use cases and scenarios.

2. Different Types of Storage Systems

There are multiple storage systems available like files, blocks, and object storage which store data in distinct ways - with their potential and constraints. File Storage uses a hierarchical structure in files and folders to store data; Block storage divides data into chunks and stores them in an organized manner in equal-sized blocks; Object storage governs data and associated metadata.

2.1. What Is File Storage?

Also known as file-based storage. Data is stored in one single unit of the information under the hierarchy of folders, just like the papers in the binder. It needs a precise path to access that piece of data in the hierarchy. This storage system is more organized. The data retrieval involves the use of a limited amount of metadata that directs to the exact path.

It is the most traditional and widely used storage system for direct and NAS (Network-attached storage) systems. It has broad capabilities and can store a massive variety of data. It has been a popular technique for decades and also convenient for transactional data or more structured volumes.

File storage works seamlessly with an organized load of structured data. But as the data grows, the process of retrieval can become cumbersome and time-demanding. Scaling needs more hardware devices or replacing drives with higher storage capacity both of which can be expensive. File storage can function well with millions of files but it can't handle billions of files. It is best suited for a limited number of files locally.

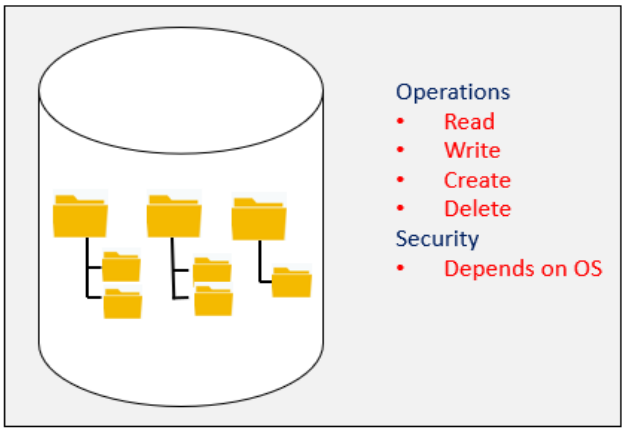

Figure 1: High-level overview of File Storage

2.1.1. Benefits of File Storage

1. Simple storage system

This is the simplest and familiar and elementary approach for managing files and folders on drives or NAS (Network Attached Storage). Files have names and limited metadata which are stored under a hierarchical structure. It is a highly abstracted storage system from an end-user perspective.

2. Sharing files

It is the most ideal way of sharing and centralizing files on LAN (Local Area Network). These files stored on NAS (Network Attached Storage) can be accessed across the network with the right set of access permissions.

3. Security

LAN (Local Area Network) based storage provides a level of security. Cloud-connected storage provides additional data security and disaster recovery by data replication across the globe in different regions.

4. Uses familiar protocols.

Standard Operating Systems like Linux and Windows uses SMB (Server Message Block) and NFS (Network FileSystem) and permits to read and write files to servers over the network.

2.1.2. Drawbacks of File Storage

1. Data Redundancy

Since data management is a file storage system that is manual and tedious it might result in data redundancy and can occupy unwanted space. This can increase the need for scalability and also increases cost.

2. File searching

This is one of the major drawbacks, when large amounts of data are stored, it impacts file search performance, and also due to data redundancy problems it has high latency for searching files.

2.2.What Is Block Storage?

Block storage is a data storage approach in which each storage block acts as an independent hard drive that is configured by the administrator. Data is stored to storage media in equal-sized blocks in this block storage model. Each and every block is associated with a distinct address and the address is only metadata associated with that block.

Block storage can decouple data from user environments, permitting that data to broadcast across multiple environments. This generates multiple paths to the same data and permits the user to retrieve it faster. When data is queried, block storage associates all the blocks and presents the whole data to the user.

To govern block storage, a piece of software independent of media controls the placement and organization of blocks. The same software handles the retrieval of data using metadata to find the desired blocks and manage the data in them.

Block storage plays an important role in many enterprise applications.

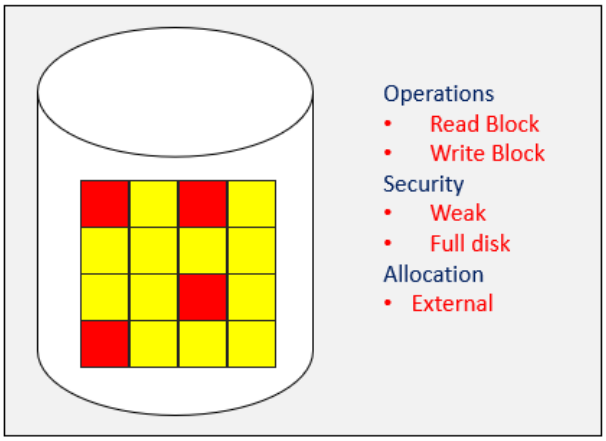

Figure 2: High-Level Overview of Block Storage

2.2.1. Benefits of Block Storage

1. High Performance

When there are a large number of IO's happening block storage most ideal storage which has very low latency and fast retrieval. Block storage is mostly used in database servers where there are a large number of queries simultaneously.

2. Flexibility and Scalability

Scaling block storage is simply adding new block storage volumes without impacting performance. Block storage is very mobile, they are easily movable between servers by swapping destination routes.

3. Easy file modification

From a user perspective-changing, a specific part of a file is very easy. In block storage, a specific block can be accessed and change can be added to that specific block. Unlike Object storage here just a chunk of files can be accessed.

4. Bootability in OS

Operating systems can be easily booted directly from blocks using SAN (Storage Area Network).

2.2.2. Drawbacks of Block Storage

1. Metadata limitation

Block Storage uses a constrained amount of metadata, much lesser than other types of storage. This can affect application storage retrieval and searching operations because applications cannot recognize the location and may have to perform a scan on a large number of blocks.

2. Server Binding

Block storage is tightly integrated into the server and cannot be accessed by other nodes at the same time. Some software can handle this scenario but it adds overhead to the file system.

3. Cost

Block storage is the most expensive of all storage since it is used for high IO operations. SAN (Storage Area Network) requires hefty investment and demands highly trained individuals for maintenance.

2.3. What Is Object Storage?

Object Storage stores each piece of data as an object. Data is divided into chunks and stored on shared storage with its accompanying metadata and distinctive identifier. At the time of retrieval, these chunks and metadata are identified with a custom identifier. Metadata is essential to success for Object Storage.

Unlike other storage systems, Object storage provides a high level of abstraction to the end-user. The end-user just interacts with the simple interface to store, retrieve and delete objects. It reduces the overhead of managing the files. Instead of highlighting the logical block it just provides abstraction in the form of a bunch of objects. In the domain of object store, space allocation is done by object store itself instead of any high-level file system software. All the operations are carried through requests and every time an end-user queries object-store all the requests are authenticated (authentication of end-user) and authorized (check privileges and access control to end-user). This per-request authentication and authorization provide security at the granular level.

Object storage is compliant with a massive variety of unstructured data like videos, songs, pictures, etc. This makes it more flexible and useful from the perspective of the end-user.

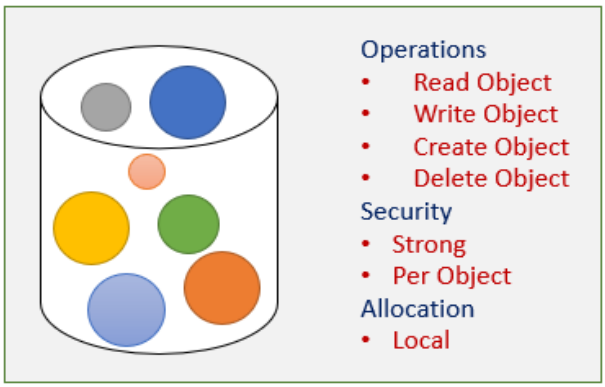

Figure 3: High-Level Overview of Object Storage

2.3.1. Benefits of Object Storage

1. Infinite Scalability

Object Storage is developed to meet the need for escalating scalability and extensibility. It evades the constraints of hierarchical storage architecture by binding the data and associated metadata in a single container with a global and distinct identifier. Multiple storage servers can be added to existing clusters to scale object storage.

2. Security

Object storage is deployed over private networks and each request from the end-user comes through the authentication and authorization filter, which adds security at the granular level. Space allocation and access control are accomplished by the IAM (Identity and Access Management) roles by the end-user.

3. Cost Reduction

Organizations can accomplish major savings by mobilizing the stagnant data to object storage with zero impact on user data access. Due to its highly scaling factor, it makes object storage more cost-effective than other storage systems. This reduces backup costs and time demand.

4. Data recalls at faster rates

Object Storage uses metadata and custom identifiers to retrieve data and the absence of a hierarchical structure of files makes it faster. They handle lower amounts of traffic simultaneously since it

is used as a data archival facility in most of the use cases.

2.3.2 Drawbacks of Object Storage

1. Less compatibility with traditional database

Since object storage is focused on handling unstructured data hence, rolling back to traditional databases which hold data in a more structured way in tables can cause compatibility issues.

2. Can't alter the chunk of the data blob

It forbids to alter the specific chunk of an object, due to this the entire object must be read at once. Thus, this adds an extra overhead of handling complete objects and adds extra latency. Hence, object storage is mostly used for archival processes.

3. Can't be mounted on Operating Systems

Object storage can't be mounted on Operating Systems like a normal disk, it requires some clients and plugins. Thus, unlike other storage systems, searching becomes a tedious task in object storage.

3. Comparison Table

Parameter	**File Storage**	**Block Storage**	**Object Storage**
Scalability	Limited scalability	High but adds cost	Very High
Performance	Low performance	High performance	High performance only for retrieval
Cost	Low	High	Low
Security	Depends on OS	Weak	High
File Modification	Easy when the path is known	Easy due to block access	Hard - Need to read the entire object

4. Conclusion

We analyzed all three storage architectures in detail. To conclude, typically there is no best storage architecture that works for all the use cases. Each architecture has its own pros and cons and is uniquely designed to cater to a specific set of problems. Before choosing the suitable storage one needs to assess whether this storage option fulfills your requirement. If you are looking for some organized way of storing data with NAS devices then file storage is probably the best option. If your requirements demand frequent access to your data then block storage is a good choice. If you require huge volumes of storage and your data is expected to scale in the future then you may opt for block storage.

Following are the some of most suitable use cases for each of this storage architecture:

1. File storage:
   - Web content serving
   - Local file sharing
   - Centralized file sharing
   - Financial Modelling
   - Data Analytics
2. Block storage:
   - Storage Disk for VM's
   - Database systems storage
   - Cache memory storage
   - Application scratch disk
3. Object storage:
   - Streaming videos
   - Image and video storage
   - Backup and archival systems
   - Genomics
   - Disaster recovery storage systems

5. References

[1] S. Samundiswary, Nilma M Dongre, Object storage architecture in cloud for unstructured data, In ICISC, 2017

[2] M. Factor, K. Meth, D. Naor, O. Rodeh and J. Satran, Object storage: the future building block for storage systems, In IEEE International Symposium on Mass Storage Systems and Technology, 2005.

[3] Yuanyuan Guo, Jianjun Hao
, Yijun Guo, Tao Luo, Research on data block storage strategy in cloud storage system, In ICCC, 2017

[4] IDC Annual Datasphere Report https://www.idc.com/getdoc.jsp?containerId=prUS47560321

[5] Vlad Bucur, CatalinDehelean, LiviuMiclea, Object storage in the cloud and multi-cloud: State of the art and the research challenges, In AQTR, 2018

[6]Mark Laurence, Blockfileobjectstorage LinkedIn Article, 2020

[7]Mike Mesnier, Carnegie Mellon and Intel Gregory R. Ganger, Carnegie Mellon Erik Riedel, Object-Based Storage, 2003