A Comprehensive Study on Cloud Storages Devices and Its Protocol

***Jai Singh Gupta**
****Dr Rajeev Srivastava**

Abstract

Long gone are the days of one file server sitting in a corner humming away while employees save their files to the local hard drives. Today, storage devices are abstracted from the server. IT pros are now required to make important decisions when choosing storage for a particular task. To help with your decision making process, here's an overview of some of the high-level differences between block level storage and file level storage. I also provide information about which type of storage and protocol is used in storage.

Keywords*: NFS, SMB, could storage, block storage.*

Introduction

Cloud storage is an industry term for managed data storage through hosted network (typically Internet-based) service. Several types of cloud storage systems have been developed supporting both personal and business uses. Cloud storage is an term for managed data storage through hosted network (typically Internet-based) service. Several types of cloud storage systems have been developed supporting both personal and business uses. . Cloud storage refers to a virtual storage area that can span across many different physical storage devices. When you use cloud storage, some of your files may be on a physical server in New York while other files are on a physical server in California. Since most users do not know where their physical files are, using cloud storage can be thought of as a vague, untouchable thing – much like a cloud itself!

Advantages of the Cloud

- Syncing: Many cloud storage services offer syncing services. When you set up the software on your local machine, it will create a special folder on your computer and any changes made to the files in that folder will be automatically synced to your cloud storage as long as your computer is connected to the internet. When you access your cloud storage from a different computer, you will see the same files that are in that folder on your personal computer.
- Convenience: One of the biggest benefits of using the cloud is that your data can be accessed from multiple devices. Cloud storage services can usually be accessed from any device with an internet connection. You can check your email, read your documents, view your online photos, or listen to your online music from any computer, tablet, or internet-enabled cell phone.
- Online Backups: One advantage to using a cloud service is that it can serve as an online backup for important files in case your computer stops working. If your computer

crashes or you accidentally delete pictures from your hard drive, you can simply login to the cloud service and download your files again. Some cloud storage services may prompt you to automatically add files to the synced folder as soon as you add them to your computer to serve as a backup.

Disadvantages of the Cloud

- Internet Connection Required: Since data stored on the cloud is on a third party's computer, you will need an internet connection to access it. If you do not have access to the internet, you will not be able access the network of computers that store your data.
- Bandwidth Costs: Bandwidth refers to the amount of data that can be carried over the internet from one point to another. Some internet service providers give users a monthly bandwidth allocation. By transferring data to and from the cloud, you will be using part of the allocation and may have to pay overage charges if you go over that allocation.

There are three main cloud-based storage architecture models: public, private and hybrid.

- Public cloud storage services provide a multi-tenant storage environment that is most suited for unstructured data. Data is stored in global data centres with storage data spread across multiple regions or continents
- Private cloud, or on-premises, storage services provide a dedicated environment protected behind an organization's firewall. Private clouds are appropriate for users who need customization and more control over their data
- Hybrid cloud is a mix of private cloud and third-party public cloud services with orchestration between the platforms for management. The model offers businesses flexibility and more data deployment options

Cloud storage is based on a virtualized infrastructure with accessible interfaces, near-instant elasticity and scalability, multi-tenancy and metered resources. Cloud-based data is stored in logical pools across disparate, commodity servers located on premises or in a data centre managed by a third-party cloud provider. Using the RESTful API, an object storage protocol stores a file and its associated metadata as a single object and assigns it an ID number. When content needs to be retrieved, the user presents the ID to the system and the content is assembled with all its metadata, authentication and security.

Storage Devices: Storage devices can be broadly classified into three categories:

- Block Storage Devices
- File Storage Devices
- Object storage Devices

Block storage: Block-based storage has been around since the invention of the SCSI protocol and today is still one of the most popular ways to connect internal disks and external storage systems. The “unit of consumption” is the LUN or volume, which as its name suggests is logical or

virtual in nature, especially in shared storage systems. Data is written to LUNs in fixed sized blocks and provides another name for this storage format – FBA or Fixed Block Architecture.

File storage: Data is typically stored on a LUN using some kind of structure such as a file system. The file system provides a hierarchical ordering of directories and files that permit data to be accessed in a nice human-readable format. File-based protocols such as NFS and SMB/CIFS extend the idea of the file system across the network, allowing remote hosts to access the data in a hierarchical structured format. They bring additional levels of security (through metadata and other file attributes like ACLs) and data integrity through locking mechanisms that can be distributed across a wide area network.

Object storage: An object store is arguably the purest form of data storage, storing data as binary objects that can range from a few kilobytes to gigabytes in size. There's typically no hierarchical structure in object stores, with each piece of data referenced by an object ID, a system generated string of characters and numbers that requires some external process to keep track of object IDs and what the content is actually for.

Cloud Storage Protocols

All three forms of storage are available in public and private cloud deployments. Due to the increased latency incurred from accessing data across the Internet (or private networks), remotely accessing block storage is not a practical solution; attempting to fix the latency issue using techniques such as caching or parallelizing the I/O results in increased risks to data integrity. As a result, block-based systems are usually accessed by servers that are themselves cloud-based.

Both file and object protocols work well in the cloud (depending of course on the application). File access can usually tolerate greater latency issues through the use of caching and because most updates are made infrequently. This isn't the case of course where file-based systems are used for server or desktop virtualization. Object stores are great for cloud deployments and work well due to their native HTTP support. To date we've seen these platforms used as archives and for backup, primarily because of the performance characteristics of object stores.

Physical storage

At the root of all storage is some set of physical storage protocols, so I'll begin with a quick recap of physical storage? Three major classes of physical storage models are in use today: direct attached storage (DAS), the storage area network (SAN), and network attached storage (NAS).

- **DAS**. Direct attached storage is the simplest storage model. We are all familiar with DAS; this is the model used by most laptops, phones, and desktop computers. The fundamental unit in DAS is the computer itself; the storage for a server is not separable from the server itself. In the case of a phone it is physically impossible to remove the storage from the compute, but even in the case of servers, where it is theoretically possible to pull disk drives, once a drive is separated from the server, it is generally wiped before reuse. SCSI and SATA are examples of DAS protocols.

- **SAN**. Eventually the storage industry recognized the utility of separating storage from the compute. Rather than attaching disks to each individual computer, we placed all the disks on a single cluster of servers and accessed the disk over the network. This simplifies storage management tasks such as backup and failure repair. This division of storage and compute is often called shared storage, since multiple computers will use a single pool of storage. It was most straightforward to communicate between the client and server over the network using the same (or very similar) block protocols that were used to communicate with locally attached disk drives.

  Storage exposed this way is called a storage area network. Fibre Channel and iSCSI are examples of SAN protocols. In a SAN an administrator will group a set of disks (or a portion of a set of disks) into a LUN (logical unit), which then behaves like a single disk drive to outside computers. The LUN is the fundamental unit used to manage SAN storage.

- **NAS**. While SANs allow us to move LUNs between one computer and another, the block protocols they use were not designed to concurrently share data in the same LUN between computers. To allow this kind of sharing we need a new kind of storage built for concurrent access. In this new kind of storage we communicate with the storage using file system protocols, which closely resemble the file systems run on local computers. This kind of storage is known as network attached storage. NFS and SMB are examples of NAS protocols. The file system abstraction allows multiple servers to access the same data at the same time. Multiple servers can read the same file at the same time, and multiple servers can place new files into the file system at the same time. Thus, NAS is a very convenient model for shared user or application data. NAS storage allows administrators to allocate portions of storage into individual file systems. Each file system is a single namespace, and the file system is the primary unit used to manage NAS.

Block technologies:

Fibre Channel: Fibre Channel is a technology for transmitting data between computer devices at data rates of up to 16 Gbps. Fibre Channel is especially suited for connecting computer servers to shared storage devices and for interconnecting storage controllers and drives. Since Fibre Channel is three times as fast, it has begun to replace the Small Computer System Interface (SCSI) as the transmission interface between servers and clustered storage devices. Fibre channel is more flexible; devices can be as far as ten kilometres (about six miles) apart if optical fibre is used as the physical medium. Optical fibre is not required for shorter distances, however, because Fibre Channel also works using coaxial cable and ordinary telephone twisted pair.

Fibre Channel offers point-to-point, switched, and loop interfaces. It is designed to interoperate with SCSI, the Internet Protocol (IP) and other protocols, but has been criticized for its lack of compatibility – primarily because (like in the early days of SCSI technology) manufacturers sometimes interpret specifications differently and vary their implementations.

The world wide name (WWN) in the switch is equivalent to the Ethernet MAC address. As with the MAC address, you must uniquely associate the WWN to a single device. The principal switch selection and the allocation of domain IDs rely on the WWN. The WWN manager, a process-level manager residing on the switch's supervisor module, assigns WWNs to each switch.

FC addresses are FCIDs, which get assigned by a switch, based on its internal representation of its ports. Each node is identified by an 8-bit Port_ID.

ISCSI: iSCSI is a transport layer protocol that describes how Small Computer System Interface (SCSI) packets should be transported over a TCP/IP network. iSCSI, which stands for Internet Small Computer System Interface, works on top of the Transport Control Protocol (TCP) and allows the SCSI command to be sent end-to-end over local-area networks (LANs), wide-area networks (WANs) or the Internet. IBM developed iSCSI as a proof of concept in 1998, and presented the first draft of the iSCSI standard to the Internet Engineering Task Force (IETF) in 2000.

ISCSI works by transporting block-level data between an iSCSI initiator on a server and an iSCSI target on a storage device. The iSCSI protocol encapsulates SCSI commands and assembles the data in packets for the TCP/IP layer. Packets are sent over the network using a point-to-point connection. Upon arrival, the iSCSI protocol disassembles the packets, separating the SCSI commands so the operating system (OS) will see the storage as a local SCSI device that can be formatted as usual.

Today, some of iSCSI's popularity in small to midsize businesses (SMBs) has to do with the way server virtualization makes use of storage pools. In a virtualized environment, the storage pool is accessible to all the hosts within the cluster and the cluster nodes communicate with the storage pool over the network through the use of the iSCSI protocol.

FCoE: FCoE (Fibre Channel over Ethernet) is a storage protocol that enables Fibre Channel communications to run directly over Ethernet. FCoE makes it possible to move Fibre Channel traffic across existing high-speed Ethernet infrastructure and converges storage and IP protocols onto a single cable transport and interface.

The goal of FCoE is to consolidate input/output (I/O) and reduce switch complexity as well as to cut back on cable and interface card counts. Adoption of FCoE been slow, however, due to a scarcity of end-to-end FCoE devices and reluctance on the part of many organizations to change the way they implement and manage their networks

Fibre Channel supports high-speed data connections between computing devices that interconnect servers with shared storage devices and between storage controllers and drives. FCoE shares Fibre Channel and Ethernet traffic on the same physical cable or lets organizations separate Fibre Channel and Ethernet traffic on the same hardware. FCoE uses a lossless Ethernet fabric and its own frame format. It retains Fibre Channel's device communications but substitute's high-speed Ethernet links for Fibre Channel links between devices.

FCIP: Fibre Channel over IP (FCIP or FC/IP, also known as Fibre Channel tunnels or storage tunnelling) is an Internet Protocol (IP)-based storage networking technology developed by the Internet Engineering Task Force (IETF).

FCIP mechanisms enable the transmission of Fibre Channel (FC) information by tunnelling data between storage area network (SAN) facilities over IP networks; this capacity facilitates data sharing over a geographically distributed enterprise. One of two main approaches to storage data transmission over IP networks, FCIP is among the key technologies expected to help bring about rapid development of the storage area network market by increasing the capabilities and performance of storage data transmission.

File Technologies

CIFS: The Common Internet File System (CIFS) is a protocol that gained rapid popularity around the turn of the millennium (the year 2000) as vendors worked to establish an Internet Protocol-based file-sharing protocol. At its peak, CIFS was widely supported by operating systems (OSes) such as Windows, Linux and UNIX.

CIFS uses the client/server programming model. A client program makes a request of a server program (usually in another computer) to access a file or to pass a message to a program that runs in the server computer. The server takes the requested action and returns a response.

CIFS is a public or open variation of the original Server Message Block (SMB) protocol developed and used by Microsoft. Like the SMB protocol, CIFS runs at a higher level and uses the Internet's TCP/IP protocol. CIFS was viewed as a complement to existing Internet application protocols such as the File Transfer Protocol (FTP) and the Hypertext Transfer Protocol (HTTP). Today, CIFS is widely regarded as an obsolete protocol. Although some OSes still support CIFS, newer versions of the SMB protocol — such as SMB 2.0 and SMB 3.0 — have largely taken the place of CIFS.

Some capabilities of the CIFS protocol include:

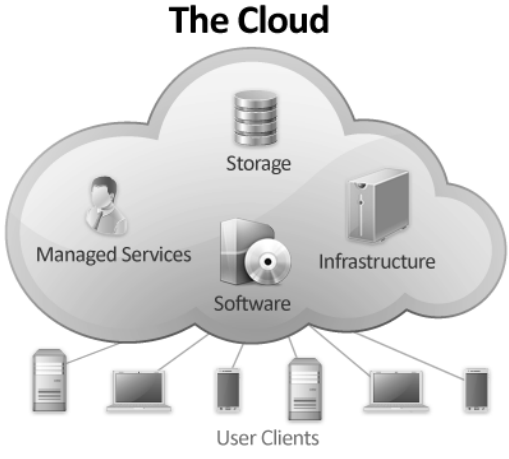
The ability to access files that are local to the server and read and write to them; File sharing with other clients using special locks; Automatic restoration of connections in case of network failure Unicode file names.

NFS: The Network File System (NFS) is a client/server application that lets a computer user view and optionally store and update file on a remote computer as though they were on the user's own computer. The user's system needs to have an NFS client and the other computer needs the NFS server. Both of them require that you also have TCP/IP installed since the NFS server and client use TCP/IP as the program that sends the files and updates back and forth. (However, the User Datagram Protocol, UDP, which comes with TCP/IP, is used instead of TCP with earlier versions of NFS.)

NFS was developed by Sun Microsystems and has been designated a file server standard. Its protocol uses the Remote Procedure Call (RPC) method of communication between computers.

Using NFS, the user or a system administrator can mount all or a portion of a file system (which is a portion of the hierarchical tree in any file directory and subdirectory, including the one you find on your PC or Mac). The portion of your file system that is mounted (designated as accessible) can be accessed with whatever privileges go with your access to each file (read-only or read-write).

SMB: Shares are also referred to as Common Internet File System (CIFS) shares or Windows Shares. With SMB Shares you specify a resource you want to share on a given server, and then the client connects to that resource using something called a Universal Naming Convention (UNC) path. A UNC path may look something like \\MyServer\ShareName or\\MyServer\ Sharename\ my file where My Server is the name of your server, and Share name is the name of the resource that has been shared and my file is the name of the file you wish to access off the network share. SMB can be used for sharing both files and printers and may use the NetBIOS Name, DNS Name, or the IP Address of the server.



**Research Scholar,*
MJRP University Jaipur
& Principal, LBS PG College, Jaipur
***Research Guide, MJRP University Jaipur*

References

1. https://en.wikipedia.org/wiki/Cloud_computing
2. https://teckadmin.wordpress.com/2014/01/19/nfs-cifs-iscsi-file-systems/
3. http://computer.howstuffworks.com/cloudcomputing/cloud-storage.html
4. https://en.wikipedia.org/wiki/Cloud_research
5. https://www.brentozar.com/archive/2012/05/storage-protocol-basics-iscsi-nfs-fibre-channel-fcoe/
6. http://www.doxer.org/difference-between-scsi-iscsi-fcp-fcoe-fcip-nfs-cifs-das-nas-san/