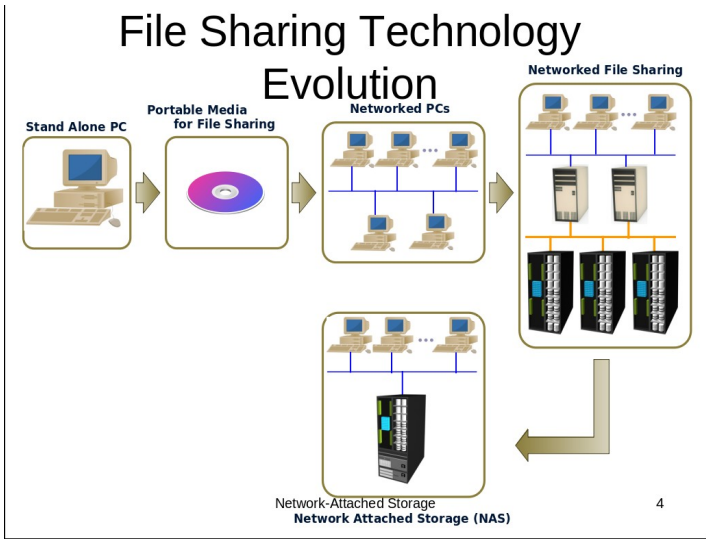
Information Storage Management
Module 4: Assignment

1 : Briefly describe about File Sharing Technology Evolution

Ans:

File sharing is the practice of distributing or providing access to digital media, such as computer programs, multimedia (audio, images and video), documents or electronic books. File sharing may be achieved in a number of ways. Common methods of storage, transmission and dispersion include manual sharing utilizing removable media, centralized servers on computer networks, World Wide Web-based hyperlinked documents, and the use of distributed peer-to-peer networking.



Initialy all the files are stored on our PC itslef or copying files to some portable medias (like floppy,CD,DVD,USB) , But this method is not suited for large organisation spread across different regions.
To rectify that problem Networked sharing came to play.Due to tremendous growth of data, organisers set up many file servers.They connected to DAS or SAN.

But such environment have poor scalability,high cost and complexity.Network-attached storage (NAS) as emerged as solution..Which can be considered as a server dedicated for file sharing alone.

2 : What is NAS? Write its advantages?

Ans:

Network-attached storage (NAS) is a file-level (as opposed to block-level) computer data storage server connected to a computer network providing data access to a heterogeneous group of clients. NAS is specialized for serving files either by its hardware, software, or configuration. It is often manufactured as a computer appliance – a purpose-built specialized computer.

Advantages :

- Improved Flexibility: compatabile with both UNIX and Windows clients .It can server request from diff. types of clients from same source.
- NAS are mainly speed and convenience. Instead of a hard drive connecting to your computer, NAS connects to your wireless router – enabling multiple users from multiple devices to access the files on the network
- Comprehensive access to information: Enables efficient file sharing and supports many-to-one and one-to-many configurations.
- Improved efficiency: NAS uses an operating system specialized for file serving.
- Simplified management
- Private Cloud Storage
- Remote Access

3 : Differentiate General Purpose Servers and NAS Devices with figure?

Ans:

A NAS device is optimized for file-serving functions such as storing, retrieving, and accessing fi les for applications and clients.A general-purpose server can be used to host any application because it runs a general-purpose operating system. Unlike a general-purpose server, a NAS device is dedicated to file-serving. It has specialized operating system dedicated to file serving by using industry-standard protocols.

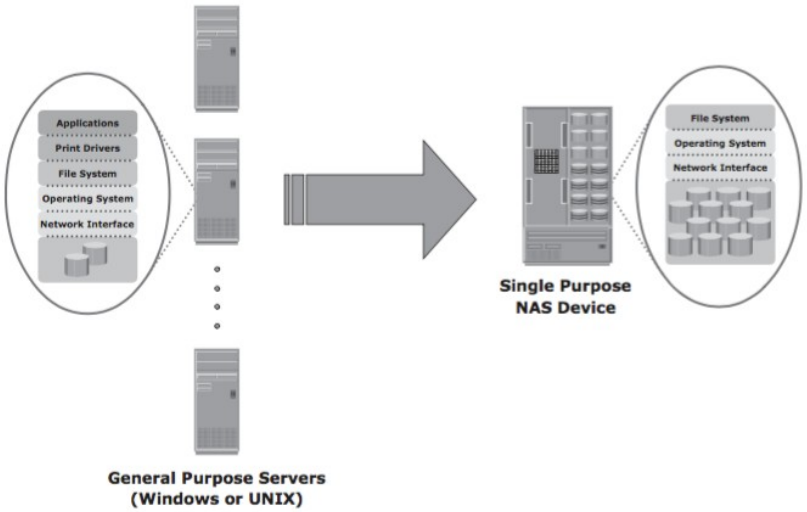


Figure 7-1: General purpose server versus NAS device

4 : Write about Benefits of NAS ?

Ans:

- Comprehensive access to information: Enables effi cient fi le sharing and supports many-to-one and one-to-many confi gurations. The many-to-one configuration enables a NAS device to serve many clients simultaneously.The one-to-many configuration enables one client to connect with manyNAS devices simultaneously.
- Improved efficiency: NAS delivers better performance compared to a general-purpose fi le server because NAS uses an operating system spe-cialized for file serving.
- Improved flexibility: Compatible with clients on both UNIX and Windows platforms using industry-standard protocols. NAS is fl exible and can serve requests from different types of clients from the same source.
- Centralized storage: Centralizes data storage to minimize data duplication on client workstations, and ensure greater data protection
- Simplified management: Provides a centralized console that makes it possible to manage file systems efficiently
- Low cost: NAS uses commonly available and inexpensive Ethernet components.
- Ease of deployment: Configuration at the client is minimal, because the clients have required NAS connection software built in.

5. Write notes on File Sharing Environment?

Ans:

A file sharing environmnet consist of multiple servers or NAS devices. Some times it needed to move files from one device to another ,File level virtualization is implemented in file sharing environment with high mobility.Even if the files are accessed, it enable movement of files accross NAS device.

6. Write about Components of NAS?

Ans:

Mainly Two Components: **NAS head and storage**

NAS head includes:

- CPU and memory
- One or more network interface cards (NICs), which provide connectivity to the client network. Examples of network protocols supported by NIC include Gigabit Ethernet, Fast Ethernet, ATM, and Fiber Distributed Data Interface (FDDI).
- An optimized operating system for managing the NAS functionality. It translates fi le-level requests into block-storage requests and further converts the data supplied at the block level to file data..
- NFS, CIFS, and other protocols for file sharing
- Industry-standard storage protocols and ports to connect and manage physical disk resources.

The storage could be external to the NAS device and shared with other hosts.

7. How NAS Implementation is done?

Ans:

Three common NAS implementations are unified, gateway, and scale-out.

-

Unified NAS -

- Unified NAS performs fi le serving and storing of fi le data, along with providing access to block-level data. It supports both CIFS and NFS protocols for file access and iSCSI and FC protocols for block level access. Due to consolidation of NAS-based and SAN-based access on a single storage platform, unified NAS reduces an organization’s infrastructure and management costs.

Gateway NAS -

- A gateway NAS device consists of one or more NAS heads and uses external and independently managed storage. Similar to unifi ed NAS, the storage is shared with other applications that use block-level I/O. Management functions in this type of solution are more complex than those in a unified NAS environment because there are separate administrative tasks for the NAS head and the storage. A gateway solution can use the FC infrastructure, such as switches and directors for accessing SAN-attached storage arrays or direct- attached storage arrays.

Scale-out NAS -

- Both unified and gateway NAS implementations provide the capability to scale-up their resources based on data growth and rise in performance requirements.Scaling up these NAS devices involves adding CPUs, memory, and storage to the NAS device. Scalability is limited by the capacity of the NAS device to house and use additional NAS heads and storage. Scale-out NAS enables grouping multiple nodes together to construct a clustered NAS system. A scale-out NAS provides the capability to scale its resources by simply adding nodes to a clustered NAS architecture. The cluster works as a single NAS device and is managed centrally.

8. NAS File Sharing Protocols?

a) Common Internet File System (CIFS)

b) Network File System

Ans:

a) **CIFS** is abbreviation for "**Common Internet File System**" used by Windows operating systems for file sharing. CIFS also uses the client-server methodology where A client makes a request of a server program for accessing a file .The server takes the requested action and returns a response. CIFS is a open standard version of the Server Message Block Protocol (SMB) developed and used by Microsoft and it uses the **TCP/IP protocol.**

CIFS provides the following features to ensure data integrity:

- It uses file and record locking to prevent users from overwriting the work of another user on a file or a record.
- It supports fault tolerance and can automatically restore connections and reopen files that were open prior to an interruption.

b) NFS is the "Network File System" specifically used for Unix and Linux operating systems. It allows files communication transparently between servers and end users machines like desktops & laptops. NFS uses client- server methodology to allow user to view read and write files on a computer system. A user can mount all or a portion of a file system via NFS.

The NFS protocol provides a set of RPCs to access a remote file system for the following operations:

- Searching files and directories
- Opening, reading, writing to, and closing a file
- Changing file attributes
- Modifying file links and directories

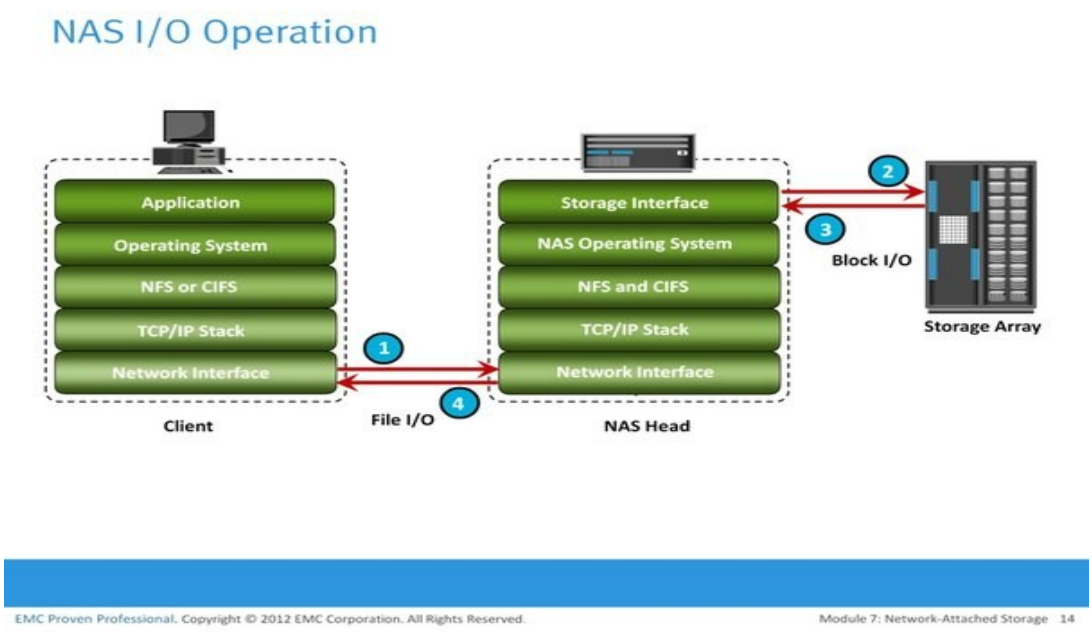
9. Explain the NAS I/O Operation?

Ans:

NAS provides fi le-level data access to its clients. File I/O is a high-level request that specifies the file to be accessed.The NAS operating system keeps track of the location of files on the disk volume and converts client file I/O into block-level I/O to retrieve data. The process of handling I/Os in a NAS environment is as follows:

- The requestor (client) packages an I/O request into TCP/IP and forwards it through the network stack. The NAS device receives this request from the network

- The NAS device converts the I/O request into an appropriate physical storage request, which is a block-level I/O, and then performs the operation on the physical storage.
- When the NAS device receives data from the storage, it processes and repackages the data into an appropriate fi le protocol response.
- The NAS device packages this response into TCP/IP again and forwards it to the client through the network.



10 . Factors Affecting NAS Performance in detail?

Ans:

NAS uses IP network; therefore, bandwidth and latency issues associated with IP affect NAS performance. Network congestion is one of the most signifi can't sources of latency in a NAS environment. Other Factors:

- **Number of hops:** A large number of hops can increase latency because IP processing is required at each hop, adding to the delay caused at the router.
- **Authentication with a directory service such as Active Directory or NIS:** The authentication service must be available on the network with enough resources to accommodate the authentication load. Otherwise, a large number of authentication requests can increase latency.

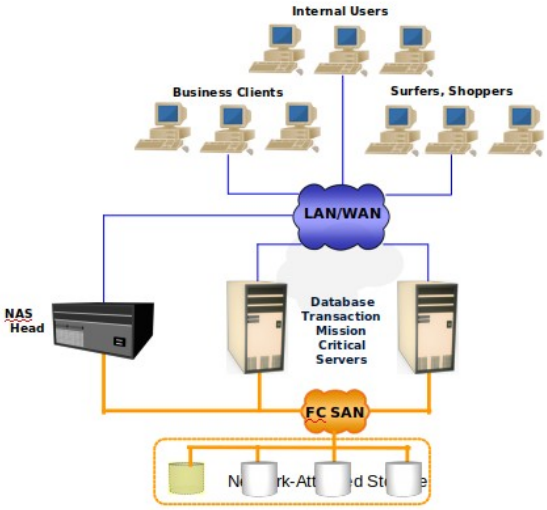
- **Retransmission:** Link errors and buffer overfl ows can result in retransmission. This causes packets that have not reached the specifi ed destination to be resent.
- **Overutilized routers and switches:** The amount of time that an over- utilized device in a network takes to respond is always more than the response time of an optimally utilized or underutilized device. Network administrators can view utilization statistics to determine the optimum utilization of switches and routers in a network.
- **File system lookup and metadata requests:** NAS clients access fi les on NAS devices. The processing required to reach the appropriate file or directory can cause delays. Sometimes a delay is caused by deep directory structures and can be resolved by fl attening the directory structure.
- **Over utilized NAS devices:** Clients accessing multiple fi les can cause high utilization levels on a NAS device, which can be determined by viewing utilization statistics. High memory, CPU, or disk subsystem utilization levels can be caused by a poor fi le system structure or insuffi cient resources in a storage subsystem.
- **Over utilized clients:** The client accessing CIFS or NFS data might also be over utilized. An overutilized client requires a longer time to process the requests and responses. Specifi c performance-monitoring tools are available for various operating systems to help determine the utilization of client resources.

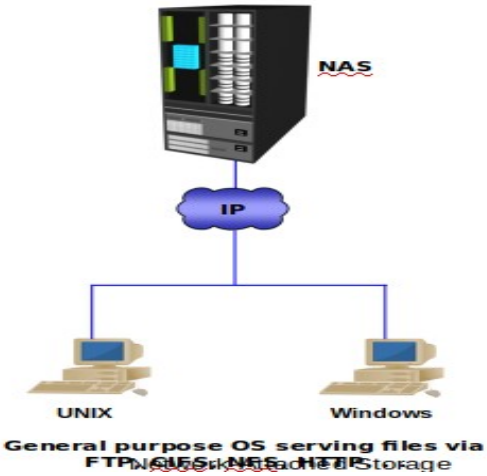
11 . Illustrate Server and storage Consolidation with NAS with figure?

Ans:

- Server consolidation is the process of migrating network services and application from multiple computers to a singular computer
- This consolidation can include multiple physical computers to multiple virtual computers on one host computer. You can consolidate computer for several reason ,such as minimizing power consumption ,simplifying administration duties ,or reducing overall cost
- Storage Consolidation also called storage convergence is a method of centralizing data among muliple servers.The objective is facilitate data backup and archiving for

all subscribers in an enterprises while minimizing the time required to access and store data





12. Briefly describe about File-level Virtualization. Also draw a comparison figure with explanation to show the environment Before and After File-level Virtualization

Ans:

File-level virtualization eliminates the dependencies between the data accessed at the file level and the location where the files are physically stored. Implementation of file-level virtualization is common in NAS or file-server environments. It provides non-disruptive file mobility to optimize storage utilization.

File-level virtualization simplifies file mobility. It provides user or application independence from the location where the fi les are stored. File-level virtualization creates a logical pool of storage, enabling users to use a logical path, rather than a physical path, to access files. File-level virtualization facilitates the movement of files across the online file servers or NAS devices. This means that while the files are being moved, clients can access their files nondisruptively.Clients can also read their files from the old location and write them back to the new location without realizing that the physical location has changed. A global namespace is used to map the logical path of a file to the physical path names.

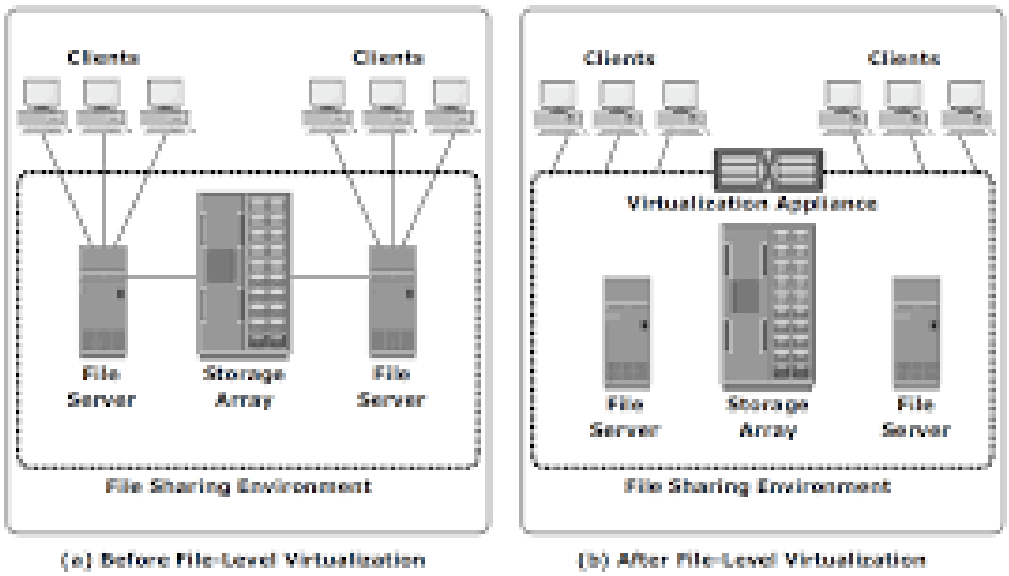
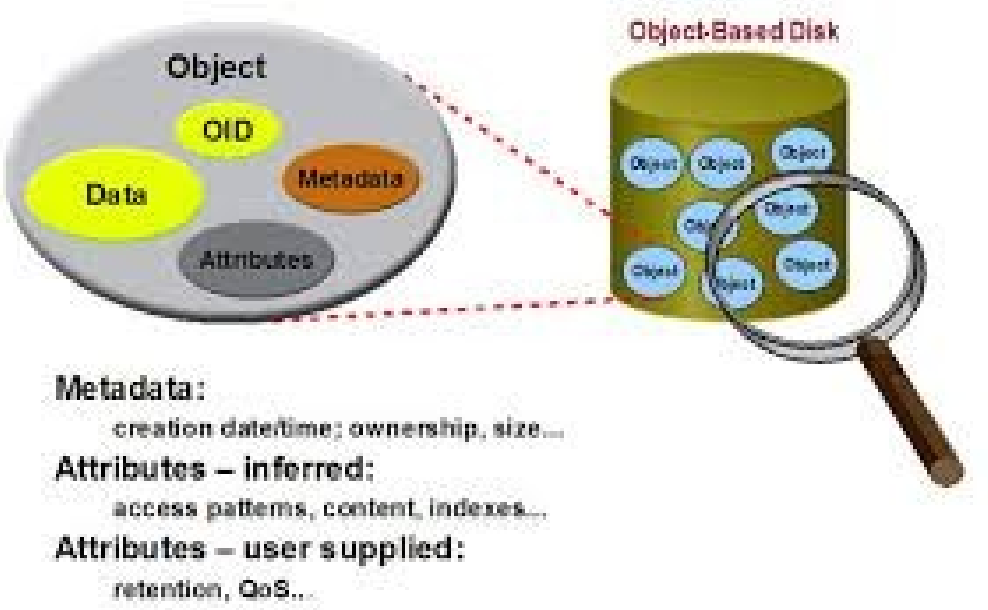


Figure 10-7: NAS device before and after file-level virtualization

13. Describe about Object-based StorageWith figure explain Object-based storage Device (OSD)?

Ans:

Object-based storage is a way to store file data in the form of objects based on its content and other attributes rather than the name and location. Due to varied application requirements, organizations have been deploying storage area networks (SANs), NAS, and object-based storage devices (OSDs) in their data centers. Deploying these disparate storage solutions adds management complexity, cost and environmental overhead. An ideal solution would be to have an integrated storage solution that supports block, file, and object access. Unified storage has emerged as a solution that consolidates block, file, and object-based access within one unifi ed platform. It supports multiple protocols for data access and can be managed using a single management interface. An OSD is a device that organizes and stores unstructured data, such as movies, office documents, and graphics, as objects. Object-based storage provides a scalable, self-managed, protected, and shared storage option. OSD stores data in the form of objects. OSD uses flat address space to store data.

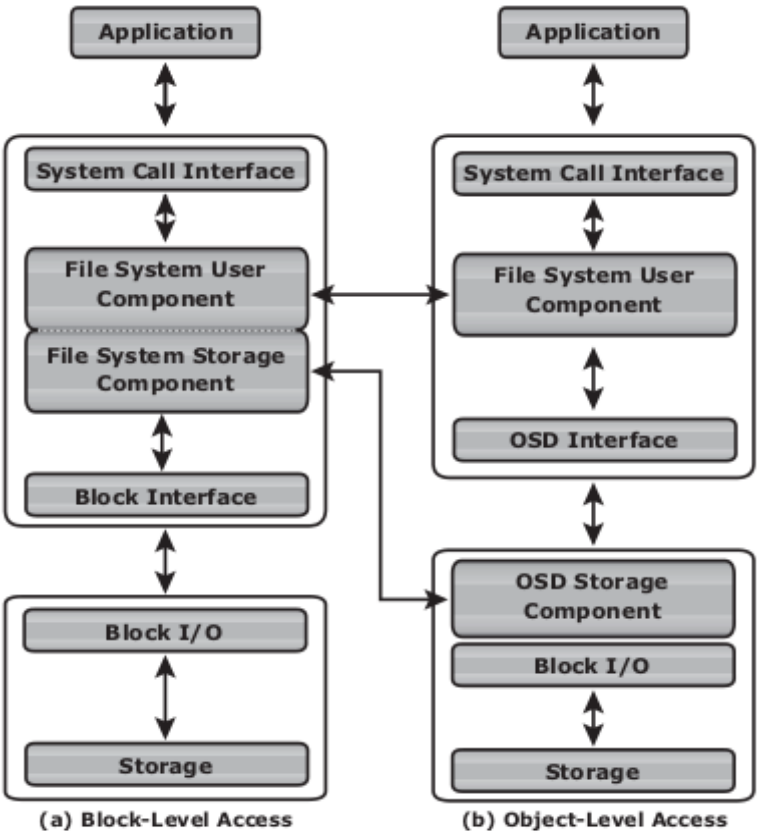


14. Compare I/O access in Traditional Vs. Object-based Storage Model. Draw figure.?

Ans:

An I/O in the traditional block access method passes through various layers in the I/O path. The I/O generated by an application passes through the

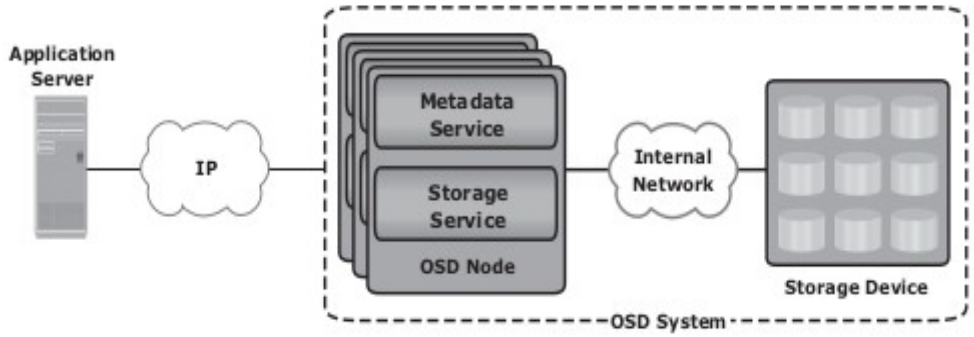
file system, the channel, or network and reaches the disk drive. When the file system receives the I/O from an application, the fi le system maps the incoming I/O to the disk blocks. The block interface is used for sending the I/O over the channel or network to the storage device. The I/O is then written to the block allocated on the disk drive. illustrates the block-level access.



15. Draw and explain Key Components of Object-based Storage Device.?

Ans:

The OSD system is typically composed of three key components: nodes, private network, and storageThe OSD system is composed of one or more nodes. A node is a server that runs the OSD operating environment and provides services to store, retrieve, and manage data in the system.The OSD node has two key services: metadata service and storage service. The metadata service is responsible for generating the object ID from the contents (and can also include other attributes of data) of a file. It also maintains the mapping of the object Ids and the file system namespace. The storage service manages a set of disks on which the user data is stored. The OSD nodes connect to the storage via an internal network.OSD typically uses low-cost and high-density disk drives to store the objects. As more capacity is required, more disk drives can be added to the system.



16. Explain about Storing and retrieval of Object in OSD?

Ans:

The data storage process in an OSD system is as follows:

- The application server presents the fi le to be stored to the OSD node.
- The OSD node divides the fi le into two parts: user data and metadata.
- The OSD node generates the object ID using a specialized algorithm. The algorithm is executed against the contents of the user data to derive an ID unique to this data.
- For future access, the OSD node stores the metadata and object ID using the metadata service.
- The OSD node stores the user data (objects) in the storage device using the storage service.
- An acknowledgment is sent to the application server stating that the object is stored.

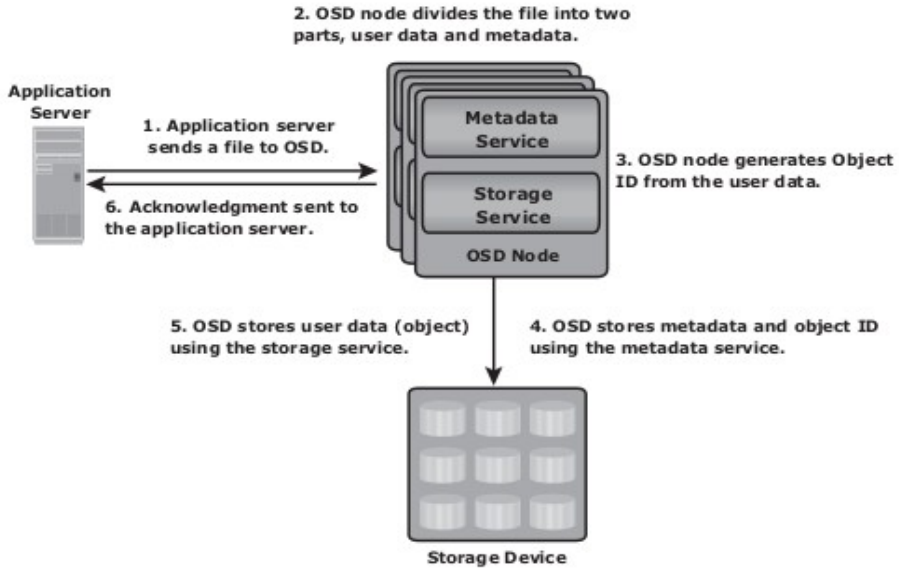


Figure 8-5: Storing objects on OSD

The process of data retrieval from OSD is as follows:

- The application server sends a read request to the OSD system.
- The metadata service retrieves the object ID for the requested fi le.
- The metadata service sends the object ID to the application server.
- The application server sends the object ID to the OSD storage service for object retrieval.
- The OSD storage service retrieves the object from the storage device.
- The OSD storage service sends the fi le to the application server.

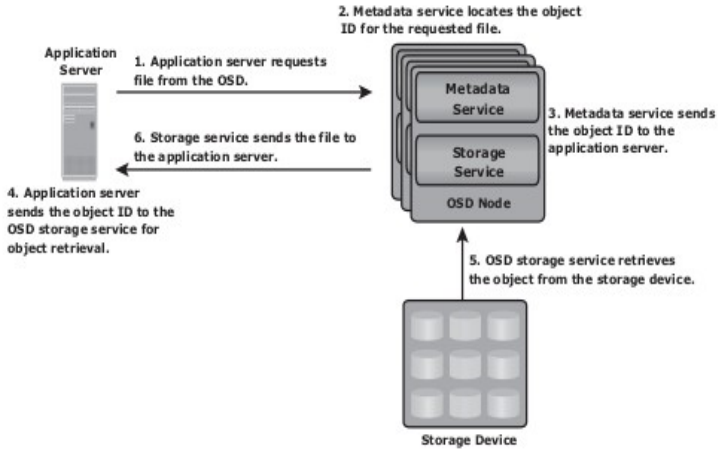


Figure 8-6: Object retrieval from an OSD system

17. Key Benefits of Object-based Storage

Ans:

- **Security and reliability**: Data integrity and content authenticity are the key features of object-based storage devices. OSD uses specialized algorithms to create objects that provide strong data encryption capability.
- **Platform independence**: Objects are abstract containers of data, including metadata and attributes. This feature allows objects to be shared across heterogeneous platforms locally or remotely. This platform-independence capability makes object-based storage the best candidate for cloud computing environments.
- **Scalability**: Due to the use of fl at address space, object-based storage can handle large amounts of data without impacting performance. Both storage and OSD nodes can be scaled independently in terms of performance and capacity.
- **Manageability**: Object-based storage has an inherent intelligence to manage and protect objects. It uses self-healing capability to protect and replicate objects. Policy-based management capability helps OSD to handle routine jobs automatically.

18. Example/Use case of OSD: Cloud-based Storage

Ans:

OSD is cloud-based storage uses a web interface to access storage resources. OSD provides inherent security, scalability, and automated data management. It also enables data sharing across heterogeneous platforms or tenants while ensuring integrity of data. These capabilities make OSD a strong option for cloud-based storage. Cloud service providers can leverage OSD to offer storage-as-a-service.

OSD supports web service access via representational state transfer (REST) and simple object access protocol (SOAP). REST and SOAP APIs can be easily integrated with business applications that access OSD over the web.

REST is an architectural style developed for modern web applications. REST provides lightweight web services to access resources (for example, documents, blogs, and so on) on which a few basic operations can be performed, such as retrieving, modifying, creating, and deleting resources.SOAP is an XML-based protocol that enables communication between the web applications running on different OSes and based on different programming languages.

19. Write a note on Data archival?

Ans :

- Data archieving is the process of moving data that is no longer used to separate storage device for long term retention. Archive data consist of older data that remains important to the organization or must me retained for future reference or regulatory compilance reasons.
- Archieve data important to organisation ,that used in future purposes.
- Archieve data stored on lower cost.
- Some archive systems treat the data as read only to prevent from modification.

20. Content Addressed Storage (CAS) (Example/Use case of OSD)

Ans:

- CAS is an object-based storage device designed for secure online storage and retrieval of fixed content. CAS stores user data and its attributes as an object. The stored object is assigned a globally unique address, known as a content address (CA).
- This address is derived from the object's binary representation. CAS provides an optimized and centrally managed storage solution. Data access in CAS differs from other OSD devices

21. What are the Key Features of CAS?

Ans:

The key features of CAS are:

- **Content authenticity:** It assures the genuineness of stored content. This is achieved by generating a unique content address for each object and validating the content address for stored objects at regular intervals.
- **Content integrity:** It provides assurance that the stored content has not been altered. CAS uses a hashing algorithm for content authenticity and integrity.

- **Location independence:** CAS uses a unique content address, rather than directory path names or URLs, to retrieve data.
- **Single-instance storage (SIS):** CAS uses a unique content address to guarantee the storage of only a single instance of an object.
- **Retention enforcement:** Protecting and retaining objects is a core require-ment of an archive storage system.
- **Data protection:** CAS ensures that the content stored on the CAS system is available even if a disk or a node fails.

22. CAS Use Case (Example): Healthcare Solution Hospital?

Ans:

- Large healthcare centers examine hundreds of patients every day and generate large volumes of medical records.Each record might be composed of one or more images that range in size from approximately 15 MB for a standard digital X-ray to more than 1 GB for oncology studies.
- Even if a patient's record is no longer needed, compliance requirements might stipulate that the records be kept in the original format for several years.
- Medical image solution providers offer hospitals the capability to view medical records, such as X-ray images, with acceptable response times and resolution to enable rapid assessments of patients
- Patients' records are retained on the primary storage for 60 days after which they are moved to the CAS system. CAS facilitates long-term storage and at the same time, provides immediate access to data, when needed.

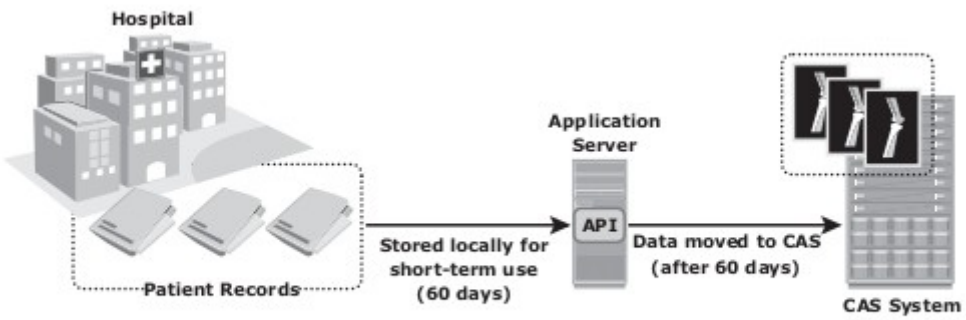


Figure 8-7: Storing patient studies on a CAS system

23. Write a note on Unified Storage?

Ans:

- Unified storage also called network unified storage or multiprotocol storage is a storage system that makes it possible to run and manage files and applications from a single device. A unified storage system simultaneously enables storage of file data and handles the block-based input/output (I/O) of enterprise applications. To this end, a multiprotocol storage system consolidates file- and block-based access in a single storage platform compared to traditional arrays that contain either one or the other.
- A unified storage architecture uses file protocols such as Server Message Block (SMB) and Network File System (NFS), and block-based protocols such as Fibre Channel (FC) and Internet Small Computer System Interface (iSCSI) so users can access consolidated applications and storage.

24. Describe about Components of Unified Storage?

Ans:

A unified storage system consists of the following key components: storagecontroller, NAS head, OSD node, and storage.

- The storage controller provides block-level access to application servers through iSCSI, FC, or FCoE protocols. It contains iSCSI, FC, and FCoE front-end ports for direct block access.
- The storage controller is also responsible for managing the back-end storage pool in the storage system. The controller confi gures LUNs and presents them to application servers, NAS heads, and OSD nodes.
- The LUNs presented to the application server appear as local physical disks. A file system is configured on these LUNs and is made available to applications for storing data.

- A NAS head is a dedicated fi le server that provides fi le access to NAS clients.The NAS head is connected to the storage via the storage controller typically using a FC or FCoE connection.
- The system typically has two or more NAS heads for redundancy. The LUNs presented to the NAS head appear as physical disks.
- The NAS head configures the file systems on these disks, creates a NFS, CIFS, or mixed share, and exports the share to the NAS clients.

25. Explain about the Data Access from Unified Storage.?

Ans:

In a unified storage system, block, file, and object requests to the storage travel through different I/O paths.

- **Block I/O request**: The application servers are connected to an FC, iSCSI, or FCoE port on the storage controller. The server sends a block request over an FC, iSCSI, or FCoE connection. The storage processor (SP) pro-cesses the I/O and responds to the application server.
- **File I/O request**: The NAS clients (where the NAS share is mounted or mapped) send a fi le request to the NAS head using the NFS or CIFS protocol. The NAS head receives the request, converts it into a block request, and forwards it to the storage controller. Upon receiving the block data from the storage controller, the NAS head again converts the block request back to the fi le request and sends it to the clients.
- **Object I/O request:** The web application servers send an object request, typically using REST or SOAP protocols, to the OSD node. The OSD node receives the request, converts it into a block request, and sends it to the disk through the storage controller.