Dharmi Gujarati

CWID - 20018001

Assignment - 5

Answer: 1

- Servers have different form factors than desktop computers. They come in the form of either rack-mounted or blade servers. These forms are optimized to reduce their physical footprint and interconnection complexity (cabling spaghetti). Such optimization is necessary in the face of an ever-increasing number of servers that need to be put in the constrained space of a data center.
- A rack-mounted server is inserted horizontally into a rack. It is denoted by its height, which varies discretely in the rack unit, of 1.75 inches (known as RU or simply U). Namely, a 1U server is 1U high, a 2U server is 2U high, and so on. Most single and dual-socket servers are available as 1U servers. A rack housing rack-mounted servers may be a simple metal enclosure or it can be a complex piece of equipment armed with power distribution, air or liquid cooling, and a keyboard/video/mouse switch that allows a single keyboard, video, and mouse to be shared among servers.

- A blade server (or simply a blade) is even more compact than a rack-mounted server. The smaller form factor is achieved by eliminating pieces that are not specific to computing, such as cooling. As a result, a blade may amount to nothing more than a computer circuit board that has a processor, memory, I/O, and an auxiliary interface. Such a blade certainly cannot function on its own. It is operational only when inserted into a chassis that incorporates the missing modules. The chassis accommodates multiple blades. It also provides a switch through which the servers within connect to the external network.
- A given rack space can house more blade servers than rack-mounted servers. The chassis–blade arrangement offers other benefits as well: reduced power consumption, simpler cabling, lower cost, and so on. This makes blade servers more attractive in the Cloud Computing environment.

- **Answer: 2**

- Ethernet technology is particularly important to data centers because of its potential to eliminate employing separate transport mechanisms (e.g., FC) for storage and interprocessor traffic.
- Prior to the introduction of Ethernet, data centers had to rely on a variety of proprietary networking protocols. Due to this, it was challenging and expensive to integrate devices from many suppliers into the same network.
- Data centers can utilize Ethernet, a widely used and supported networking protocol, in their infrastructure. To lower the cost of networking hardware, this makes it simpler to deploy, operate, and extend the infrastructure.
- High-speed data transfer rates are also provided by Ethernet technology, which is crucial for managing the high-speed amounts of data handled in data centers. Depending on the particular architecture, Ethernet networks may handle data rates of up to 100 Gbps.

- **Answer 3:**

- Storage classified as Direct Attached Storage (DAS), Network-Attached Storage (NAS), and Storage Area Network (SAN). NAS and SAN reside across a network. This network is purpose-built for, and dedicated to, storage traffic in the case of SAN. The NAS units are files or objects, while the SAN units are disk blocks. SAN relies on specialized transport, FC, which is optimized for storage traffic. NAS does not require anything special apart from the IP network. DAS is directly attached to a processor through a point-to point link. It is the most common storage arrangement, is dedicated to a single host.
- NAS and SAN are readily applicable to Cloud Computing but DAS has a limitation. An essential feature of Cloud Computing is flexible allocation of virtual machines based on, among other factors, resource availability and geographical location. In the DAS case, when a virtual machine moves to a new physical host, the associated storage needs to move to the same host, too, which is likely to result in consuming both much bandwidth and much time.
- **DAS** has **drawbacks** such restricted scalability, low redundancy, and no centralized administration tools. The maximum number of physical drives that can be attached to a server is usually the limit for DAS, and increasing storage involves increasing the number of servers, which may be expensive and difficult. A disk failure might result in data loss with DAS since it lacks the redundancy capabilities seen in NAS and SAN.

- **Local data** like boot images or swap space are ideal for archiving on **DAS**. This is so that system-level processes that demand low latency and high performance can benefit from DAS' quick and direct access to data. The system's performance can be enhanced and network traffic can be decreased by keeping local data on DAS.

Answer 4:

- Phy layer deals with line coding, out-of-band signals, and other preparations (e.g., speed negotiation) necessary for serial transmission. The name of the layer reflects the logical construct phy that represents a transceiver (consisting of a transmitter and a receiver) on a device. A phy has an 8-bit identifier that is unique within a device. The identifier is assigned by a management function. Its value is an integer equal to or greater than zero and less than the number of phys on the device.
- The physical layer, which is a conceptual layer that specifies the transmission of data through a physical media, is distinct from the PHY layer in the SAS architecture. While the PHY layer in the SAS architecture deals with the physical interface between SAS devices, the physical layer in the OSI model deals with the physical transmission of bits across a physical media.
- In SAS architecture, the PHY layer is responsible for implementing the SAS physical interface. This includes physical connectors, cables and signalling protocols. The PHY layer also provides functions such as data serialization and de-serialization, error detection, and signal amplification to ensure reliable data transmission over long distances and high speeds.

Answer 5:

- Some of the genetic file-related system calls:
  1. **open** − for opening the file for reading, writing.
  2. **close** − to close the opened file.
  3. **read** − for reading the data from the file into the buffer.
  4. **write** − for writing the data from the buffer into file.

- In NFS, the close system call does not result in an RPC invocation. There are two reasons for this. First, the NFS protocol does not have the close routine because of the original stateless design of servers (which do not keep track of past requests) to facilitate crash discovery. Second, in this case there is no file modification.

- A remote file operation, even if it has an RPC counterpart, does not necessarily result in an RPC invocation. No such invocation is needed when the information is stored in the client cache, which reduces the number of remote procedure calls and improves performance. Nevertheless, caching makes it difficult to maintain file consistency. For example, a write operation to a file at one site may not be visible at other sites that have this file open for reading.
- Additionally, some operations such as fstat() and fcntl() may not require an RPC call if the file descriptor information is already available on the client side.

Answer 6:

- FC-2M multiplexing is concerned with end-to-end connectivity, addressing, and path selection. Three types of connection are supported: point-to-point, fabric, and arbitrated loop.
  1. **Point-to-point**
     The point-to-point topology is the simplest, with a direct link between two ports (which are analogous to the SAS ports discussed earlier). It has the same effect as DAS, while supporting longer distances and working at a higher speed.
  2. **The fabric topology**
     The fabric topology is most flexible. It involves a set of ports attached to a network of interconnecting FC switches through separate physical links. The fabric routes frames individually based on the destination port address in each frame header.
  3. **The arbitrated loop**
     The arbitrated loop topology allows three or more ports to interconnect without a fabric. On the loop, only two ports can communicate with each other at any given time through arbitration.

- The fabric topology is most flexible. It involves a set of ports attached to a network of interconnecting FC switches through separate physical links. The switching network (or fabric) has a 24-bit address space structured hierarchically, according to domains and areas. An attached port is assigned a unique address during the fabric login procedure (which we will discuss later). The exact address typically depends on the physical port of attachment on the fabric (or switch, to be precise). The fabric routes frames individually based on the destination port address in each frame header. Switch fabric topologies also allow advanced features such as zoning, which allows devices on the same switch to be isolated from each other, and virtualization, which allows the creation of multiple virtual storage area networks (SANs) on the same physical switch.

➢ Answer 7:

- FCoE-aware entities are classified into **FCoE Nodes** (ENodes) and **FCoE Forwarders** (FCFs). The FCoE Initialization Protocol (FIP) messages are carried in Ethernet frames. A special Ethertype value distinguishes these frames from the FCoE frames. FIP addresses FCoE entity discovery, virtual link instantiation, and virtual link maintenance. The entity discovery procedure is typically hinged on FCFs sending, periodically, multicast discovery advertisements to a known multicast address.
- An ENode selects a compatible FCF based on the advertisement and sends a discovery solicitation at which the capability negotiation starts. Upon receiving the solicitation, the FCF responds to the ENode with a solicited discovery advertisement, confirming the negotiated capabilities. Once receiving the solicited discovery advertisement, the ENode can proceed with setting up a virtual link to the FCF.
- The procedure here is similar to the fabric login procedure in FC. Successful completion of the login procedure results in creation of a virtual port on the ENode, a virtual port on the FCF, and a virtual link between them. The MAC address of the virtual port on the ENode is typically assigned by the FCF, although it may be assigned by the ENode. A MAC address in the former case is known as a Fabric-Provided MAC Address (FPMA).

➢ Answer 8:

A. TCP is leveraged for the features that are essential to SCSI operations: reliable in-order delivery, automatic retransmission of unacknowledged packets, and congestion control.

1. **Reliable transport**

   TCP gives reliable, in-order delivery of packets, which is fundamental for SCSI operations to guarantee that information is delivered accurately and within the rectify order.

2. **retransmission**

TCP gives flow control components to avoid a sender from overpowering a receiver with data, which is vital in SCSI operations where huge amounts of information may be exchanged.

3. **Congestion control**

   TCP gives congestion control components to prevent network clog and guarantee that data is conveyed effectively, which is vital in SCSI operations to guarantee that information is exchanged in a convenient way without affecting other activity on the organize.

- The Stream Control Transmission Protocol (SCTP) is similar to TCP in its support for the features essential to SCSI operations. At the time of standardization of iSCSI, however, the SCTP was considered, too new to be relied on.

B. These features are fundamental to SCSI operations because SCSI could be a block-based protocol that requires dependable, in-order delivery of information. The information being exchanged may be basic for the operation of a system or application, and any loss or corruption of information might result in data corruption or system failure. Flow control and congestion control are too imperative in SCSI operations since large amounts of information may be exchanged, and it is imperative to anticipate the sender from overpowering the collector or causing arrange clog.

C. A SCTP (Stream Control Transmission Protocol) isn't utilized in iSCSI because it isn't broadly supported by operating systems and network devices. TCP may be a broadly supported and well-established transport protocol that's well-suited to the necessities of SCSI operations. Although SCTP offers a few advantages over TCP, such as progressed execution under certain network conditions, it isn't yet broadly acknowledged or supported.

D. iSCSI has to be deployed over an IPsec tunnel when its path traverses an untrusted network because all native iSCSI communication is in the clear, subject to eavesdropping and active attacks. , iSCSI itself does not provide any mechanisms to protect a connection or a session. By sending iSCSI over an IPsec tunnel, the information is secured from attackers and can be securely transmitted over untrusted systems.

- **Answer 9:**

  - To avoid this complexity, iSCSI employs a scheme known as connection allegiance. With this scheme, the initiator can use any connection to issue a command but must stick to the same connection for all ensuing communications.
  - The iSCSI sessions need to be managed. A big part of session management is handled by the iSCSI login procedure. Successful completion of the login procedure results in a new session or adding a connection to an existing session.
  - A prerequisite for the procedure is that the initiator knows the name and address of the storage device (i.e., the target) to use. One approach is to have such information pre-configured in the initiator. Then any change will require reconfiguration.
  - An alternative approach is based on the Service Location Protocol. It allows the initiator to dynamically discover available targets. To start the login procedure, the initiator first sets up a connection to a known TCP port on the target. Once the connection is established, the initiator performs the login steps through Login Request and Login Response PDUs. When everything is in order, the target sends a Login Response PDU with an indication that the login procedure has completed successfully. Only then can the new connection associated with the session be used for SCSI communication.
  - Effective distribution of loads across multiple TCP connections and recovery from errors are also part of session management. iSCSI supports a three-level hierarchy for error recovery, with ascending increase in complexity. At the bottom is session recovery, which rebuilds a defunct session all over again. It involves cleaning up all the associated artifacts (such as closing all TCP connections and aborting all pending SCSI commands with error indications) and then re-establishing a new set of TCP connections.
  - Next is the digest failure recovery, which, in addition to session recovery, allows the receiver of a PDU with a mismatched data digest to request that the PDU be resent. Finally, the connection recovery includes the digest failure recovery and also allows a pending command on a broken connection to be transferred to another connection (which may need to be created).
  - Each recovery procedure is suitable for a specific environment. For example, in a LAN where errors of any kind are rare, it would be sufficient to just have session recovery. Overall, an iSCSI session can remain active for as long as it is possible to have a connection between the initiator and the target. The session terminates when the last connection closes. To make multiple connections

appear as a single SCSI interconnect between the initiator and the target, iSCSI employs sequence numbers and tags.

- In iSCSI, the identifier of a session consists of an initiator part and a target part. The former (the initiator session ID) is explicitly assigned by the initiator at the session establishment; the latter is implied by the initiator's selection of the TCP endpoint at connection establishment. To ensure that the initiator session ID is unique for every session that an initiator has with a given target (especially when the initiator is distributed), a hierarchical namespace controlled by a registration authority is prescribed.
- We must emphasize that the mutual authentication step that may be part of the login procedure is only a one-time affair. It has no bearing on whether the ensuing communication is still between the authenticated nodes. Moreover, iSCSI itself does not provide any mechanisms to protect a connection or a session. All native iSCSI communication is in the clear, subject to eavesdropping and active attacks. In an untrusted environment, iSCSI should be used along with IPsec.

➢ <u>Answer 10:</u>

- A credential, as defined by ANSI INCITS 458-2011, may be an information structure containing information about entities such as a user or Device and is used to confirm and authorize access to assets. The basic requirements of a proof of interest. At a minimum, it should be verifiable, tamper-proof, hard to forge, and safe against unauthorized use.
- A credential meets all but the last requirement; there is no in-built mechanism to bind it to the acquiring client or to the communication channel between the client and the storage device. This is clearly not good, especially if the credential is subject to eavesdropping over an improperly protected storage transport. Thus, another proof scheme is in order. The standardized scheme derives a proof based on the capability key. The proof is a quantity computed with the capability key over selective request components according to the negotiated security method.
- A verification derived from the capability key, which may be a portion of the credential, can serve as a proof for access control. The capability key could be a cryptographic key that's utilized to create capabilities, which are information structures that contain authorization data for getting to particular resources.

➢ Answer 11:

- There are three approaches to block-level virtualization depending on where virtualization is done: the host, the network, or the storage device.
  1. In the **host-based approach**, virtualization is handled by a volume manager, which could be part of the operating system. The volume manager is responsible for mapping native blocks into logical volumes, while keeping track of the overall storage utilization. Ideally the mapping should provide a capability to be adjusted dynamically to allow the capacity of virtual storage to grow or shrink according to the latest need of a particular application. A major drawback of the approach is that per-host control is not favourable to optimal storage utilization in a multi-host environment, not to mention that the operational overhead of the volume manager is multiplied.
  2. In the **network-based approach**, virtualization is handled by a special function in a storage network, which may be part of a switch. The approach is transparent to hosts and storage systems as long as they support the appropriate storage network protocols (such as FC, FCoE, or iSCSI). Depending on how control traffic and application traffic are handled, it can be further classified as in-band (symmetric) or out-of-band (asymmetric).
  3. In the **storage device-based approach**, virtualization is handled by the controller of a storage system. Because of the close proximity of the controller to physical storage, this approach tends to result in good performance. Nevertheless, it has the drawback of being vendor-dependent and difficult to work across heterogeneous storage systems.
- The network-based approach is most suitable for Cloud Computing, given its relative transparency and flexibility in storage pooling. With this approach, storage can be assigned to VM hosts, which, in turn, can allocate the assigned virtual storage to VMs through their own virtualization facilities.
- The difference between in-band and out-of-band mechanisms of the network-based approach:
  1. **The in-band approach**, where the virtualization function for mapping and I/O redirection is always in the path of both the control and application traffic. Naturally the virtualization function could become a bottleneck and a single point of failure. Caching and clustering are common techniques to mitigate these problems.
  2. **The out-of-band approach**, where the virtualization function is in the path of the control traffic but not the application traffic. The virtualization function directs the application traffic. In comparison with the in-band

approach, the approach results in better performance since the application traffic can go straight to the destination without incurring any processing delay in the virtualization function.

- Advantages of in-band mechanisms include simplicity, high performance, and ease of deployment, while disadvantages include potential network congestion and latency issues.
- Advantages of out-of-band mechanisms include improved performance and reliability, but also disadvantages of additional hardware and configuration requirements.

➢ **Answer 12:**

- NOR and NAND flash memory both are types of non-volatile memory that are commonly used in solid state drives (SSDs). While, they share some similarities, there are also significant differences in their capabilities.
- The first type is known as NOR flash because its basic construct has properties resembling those of a NOR gate. NOR flash also typically has faster read speeds than NAND flash, but slower write speeds and it can be randomly addressed to a given byte. Its storage density is limited however..
- The later type of flash memory removes this limitation (while also reducing the cost). It is called NAND flash because its basic construct has properties similar to those of a NAND gate. NAND flash, however, allows random access only in units that are larger than a byte. The NAND flash has made a splash in consumer electronics [29], and it is used much more widely than NOR flash—in digital cameras, portable music players, and smart phones.

➢ **Answer 13:**

- The three limitations that are stand in the way of deploying the NAND flash solid state drives in the Cloud:
  1. A write operation over the existing content requires that this content be erased first. (This makes write operations much slower than read operations.)
  2. Erase operations are done on a block basis, while write operations on a page basis
  3. Memory cells wear out after a limited number of write–erase cycles.

- Given the limitations, directly updating the contents of a page in place will cause high latency because of the need to read, erase, and reprogram the entire block.

Answer 14:

- Memcached implementations usually employ variants of consistent hashing to minimize the updates required as the server pool changes and maximize the chance of having the same server for a given key.
- The basic algorithm of consistent hashing can be outlined as follows:
  1. Map the range of a hash function to a circle, with the largest value wrapping around to the smallest value in a clockwise fashion.
  2. Assign a value (i.e., a point on the circle) to each server in the pool as its identifier.
  3. To cache a data item of key k, select the server whose identifier is equal to or larger than H(k).
- An immediate result of consistent hashing is that a departure or an arrival of a server only affects its immediate neighbours.
- Consistent hashing is a mechanism used in distributed caching systems such as Memcached to ensure that data is evenly distributed across cache nodes, even as nodes are added or removed from the cluster. By mapping nodes to a virtual ring and using hash values to determine which node is responsible for a particular key-value pair, consistent hashing allows for efficient load balancing and minimizes the impact of node failures.