CS 524 Introduction to Cloud Computing

Dharmit Viradia

Homework 5

Prof. Igor Faynberg

1) Explain the motivation behind the two forms of server placement (rack-mounted servers and blade servers). What is sacrificed to make a blade server more compact than a rack-mounted server?

Sol: Rack-mounted Servers: A rack-mounted server, also called a rack server, is a computer dedicated to use as a server and designed to be installed in a framework called a rack. The rack contains multiple mounting slots called bays, each designed to hold a hardware unit secured in place with screws. A rack server has a low-profile enclosure, in contrast to a tower server, which is built into an upright, standalone cabinet.

Blade Servers: A blade server is a server chassis housing multiple thin, modular electronic circuit boards, known as server blades. Each blade is a server in its own right, often dedicated to a single application. The blades are literally servers on a card, containing processors, memory, integrated network controllers, an optional Fiber Channel host bus adaptor (HBA) and other input/output (IO) ports.

The motivation behind the two forms of server placement is to reduce the complexity in connection, space occupancy, and to provide flexibility.

- Space: A blade server (or simply a blade) is even more compact than a rack-mounted server. They are optimized to reduce their physical foot print and interconnection complexity. 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.

- Computing Power: 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.

- Flexibility: A blade server provides a switch through which the servers within connect to the external network. Worth noting here is that the chassis also fits into a rack much like a rack-mounted server.

(Reference: http://whatis.techtarget.com/definition/rack-server-rack-mounted-server,
http://searchdatacenter.techtarget.com/definition/blade-server,
https://en.wikipedia.org/wiki/Blade_server)

2) Why is the use of the Ethernet technology particularly important to the data centers? [Hint: What need does the use of the Ethernet effectively eliminate?]

Sol: The servers of a data centre need to be interconnected and they need to connect to the outside world as well. As the number of services increases, more cables have to fit into a given space. Top-of-Rack (TOR) and End-of-Row (EOR) are two approaches to connectivity resulting in different cabling options. Both TOR and EOR switches are implemented using the Ethernet Technology. Ethernet Technology is particularly important to data centres because of its potential to eliminate employing separate transport mechanisms (e.g. FC) for storage and inter processor traffic.

(Reference: https://en.wikipedia.org/wiki/Data_center_bridging)

3) Explain why NAS and SAN but not DAS are readily applicable to Cloud Computing. What are the limitations of DAS? Why is DAS suitable for keeping local data (such as boot image or swap space)?

Sol: DAS (Direct-Attached Storage), directly attaches to the processor through a point to point link. While NAS (Network Attached Storage) and SAN (Storage Area Network) reside across a network. This network is built for and dedicated to storage traffic in the case of SAN. The difference in NAS and SAN lies in the semantics of the interface. NAS units are files or objects while SAN units are disk blocks. Another difference is that SAN relies on specialized transport, FC which is optimized for storage traffic while NAS does not require anything special apart from IP network. NAS and SAN are readily applicable to Cloud Computing but DAS has a limitation. 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.

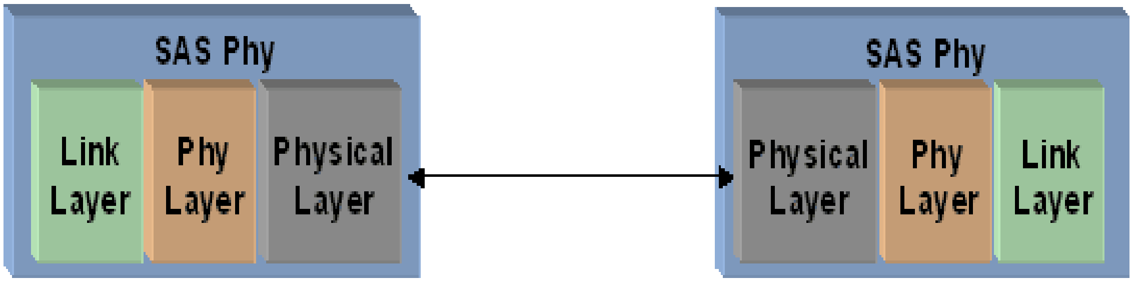
As DAS is not subject to network delay, it is suitable for keeping local data such as boot image and swap space. Depending on the location of the storage device with respect to the host, DAS may be internal as well as external.

(Reference: http://www.computerweekly.com/tip/DAS-vs-NAS-vs-SAN-Which-is-best-for-virtual-storage)

4) Why is there a need for the Phy layer in the SAS architecture? How is it different from the physical layer?

Sol: A phy layer deals with line coding out of band signals and other preparations necessary for serial transmission. It has an 8 bit identifier that is unique within a device. The identifier is assigned by a management function.

A phy, as defined in SAS, is a combination of the physical layer, phy layer and link layer functions. A minimum of two phys (one at the initiator, the other at the target) is required to complete a SAS physical connection pathway, as shown in Figure.



The physical layer deals with the physical and electrical characteristics of cables, connectors, and transceivers.

(Reference:https://www.snia.org/sites/default/education/tutorials/2007/spring/networking/SAS-Overview.pdf)

5) List the generic file-related system calls. Why in the NFS there is no RPC invocation for the close <file> system call? Under which circumstances other file operations may not result in an RPC invocation?

Sol: Generic file-related system calls are open, close, read, and write.

The reasons because there is no RPC invocation for the close <file> system calls are:

I. 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.

II. In this case there is no file modification.

A remote file operation, even if it has a 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 efficiency. Nevertheless caching makes it difficult to maintain file consistency.

(Reference: http://www2.cs.uregina.ca/~hamilton/courses/330/notes/unix/filesyscalls.html, https://en.wikipedia.org/wiki/Remote_procedure_call)

6) What types of connection topologies are supported in FC-2M? Which of them is the most flexible? Why?

Sol: In FC-2M, there are three types of connection topologies are supported:

I. Point to point

II. Fabric

III. Arbitrated Loop

Fabric connection 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 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. The exact address typically depends on the physical port of attachment on the fabric. The fabric routes frames individually based on the destination port address in each frame header.

(Reference: https://en.wikipedia.org/wiki/Switched_fabric, https://en.wikipedia.org/wiki/Arbitrated_loop , https://en.wikipedia.org/wiki/Point-to-point_(telecommunications))

7) How does the FCF respond to a discovery solicitation from the ENode?

Sol: 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 a creation of virtual port on the ENode, a virtual port on the FCF and a virtual link between them.

(Reference: Cloud Computing: Business Trends and Technologies)

8) Please answer the following four questions:
   a) What features of TCP are leveraged in iSCSI?
   b) Explain why these features are essential to SCSI operations.
   c) Why is not SCTP used in iSCSI?
   d) Why does iSCSI has to be deployed over an IPsec tunnel when its path traverses an untrusted network?

Sol: a. The features of TCP those are leveraged in iSCSI , as multiple iSCSI nodes may be reachable at the same address, and the same iSCSI node can be reached at multiple address. As a result, it is possible to use multiple TCP connections for a communication session between a pair of iSCSI nodes to achieve a higher throughput.

b. These features are essential to SCSI operations because of reliable in-order delivery, automatic re-transmission of unacknowledged packets, and congestion control.

c. The Stream Control Transmission Protocol or the SCTP is similar to the TCP in its support for the features essential to the SCSI operations. However, at the time of standardization of iSCSI, the SCTP was considered, too new to be relied on.

d. 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. Inan untrusted environment, iSCSI should be used along with IPsec

(Reference: Cloud Computing: Business Trends and Technologies)

9) What is connection allegiance? Explain how iSCSI sessions are managed.

Sol: To avoid this complexity, the 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 the session management is managed by the login procedure. Successful completion of the login procedure results in a new session or adding a connection to the existing session.

(Reference: Cloud Computing: Business Trends and Technologies)

10) Why the credential (as defined in ANSI INCITS 458-2011) itself cannot serve as a proof for access control? Give one example of a proof derived from the capability key.

Sol: A credential is essentially a cryptographically protected tamper proof capability, involving the keyed-Hash Message Authentication Code (HMAC) of a capability with a shared key. More specifically a credential is structure :

<Capability, Object Storage Identifier, Capability Key>

Where, Capability Key = HMAC (Secret Key, Capability || Object Storage Identifier)

Example: 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.

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.

(Reference: Cloud Computing: Business Trends and Technologies)

11) Describe the three approaches to the block-level virtualization. Which approach is most suitable to the needs of Cloud Computing? What are the differences between the in-band and out-of-band mechanisms of the network-based approach along with their advantages and disadvantages.

Sol: There are three approaches to block-level virtualization depending on where virtualization is done: the host, the network, or the storage device.

- Host-based: In this approach, virtualization is handled by the 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. 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 the operational overhead of the volume manager is multiplied.

- Storage Device-based: In this 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.

- Network-based: In this 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. Depending on how control traffic and application traffic are handled, it can be further classified as in-band (symmetric) or out-of-band (asymmetric).

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.

Advantages	Disadvantages
I. On the positive side, the central point of control afforded by the in-band approach simplifies administration and support for advanced storage features such as snapshots, replication and migration.	I. Naturally the virtualization function could become a bottleneck and a single point of failure.
II. The snapshot feature is of particular relevance to Cloud Computing. It can be applied to capture the state of a virtual machine at a certain point in time, reflecting the run-time conditions of its components (e.g., memory, disks, and network interface cards).	II. There is a trade-off as in this case the performance of other virtual machines on the same host may suffer when the snapshot of a virtual machine is being taken

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.

Advantages	Disadvantages
I. 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.	I. This approach does not lend itself to supporting advanced storage features. More important, it imposes an additional requirement on the host to distinguish the control and application traffic and route the traffic appropriately. As a result, the host needs to add a virtualization adaptor, which, incidentally, may also support caching of both metadata and application data to improve performance..
	II. Per-host caching, however, faces the challenging problem of keeping the distributed cache consistent

I think **Network based approach is the most suited 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.

(Reference: Cloud Computing: Business Trends and Technologies)

12) Explain the difference (in terms of their capabilities) between the NOR flash- and NAND flash solid state drives.

Sol:

NOR flash	NAND flash
I. Its basic construct has properties resembling those of a NOR gate.	I. Its basic construct has properties similar to those of a NAND gate.
II. It is fast (at least faster than hard disk), and it can be randomly addressed to a given byte.	II. It allows random access only in units that are larger than a byte.
III. Its storage density is limited	III. It has made a splash in consumer electronics
	IV. It is more widely than NOR flash – in digital cameras, portable music players, and smart phones.

13) What are the three limitations that stand in the way of deploying the NAND flash solid state drives in the Cloud?

Sol: In order to be deployed , in the cloud, the solid state drives must overcome three limitations inherent to NAND Flash:

I. A write operation over the existing content requires that this content be erased first. (This makes Write operations much slower than Read operations).

II. Erase operations are done on a block basis, while write operations on a page basis.

III. Memory cells erase out after a limited number or write-erase cycles.

(Reference: Cloud Computing: Business Trends and Technologies)

14) Explain the mechanism of consistent hashing used in Memcached servers.

Sol: Depending on the size of DRAM available on a server, caching the workload data may need more than one server. In this case, the hash table is distributed across multiple servers, which form a cluster with aggregated DRAM. Memcached servers, by design, are neither aware of one another nor coordinated centrally. It is the job of a client to select what server to use, and the client (armed with the knowledge of the servers in use) does so based on the key of the data item to be cached.

How should the hash table be distributed so that the same server is selected for the same key? A naïve scheme might be as follows:

$$s = H(k) \bmod n$$

where H(k) is a hashing function, k the key, n the number of server, and s the server label, which is assigned the remainder of the division of H(k) over n. The scheme works as long as n is constant, but it will most likely yield a different server when the number of servers grows or shrinks dynamically – as is typically the case in Cloud Computing. As a result, cache misses abound, application performance degrades, and all servers in the latest cluster have to be updated.

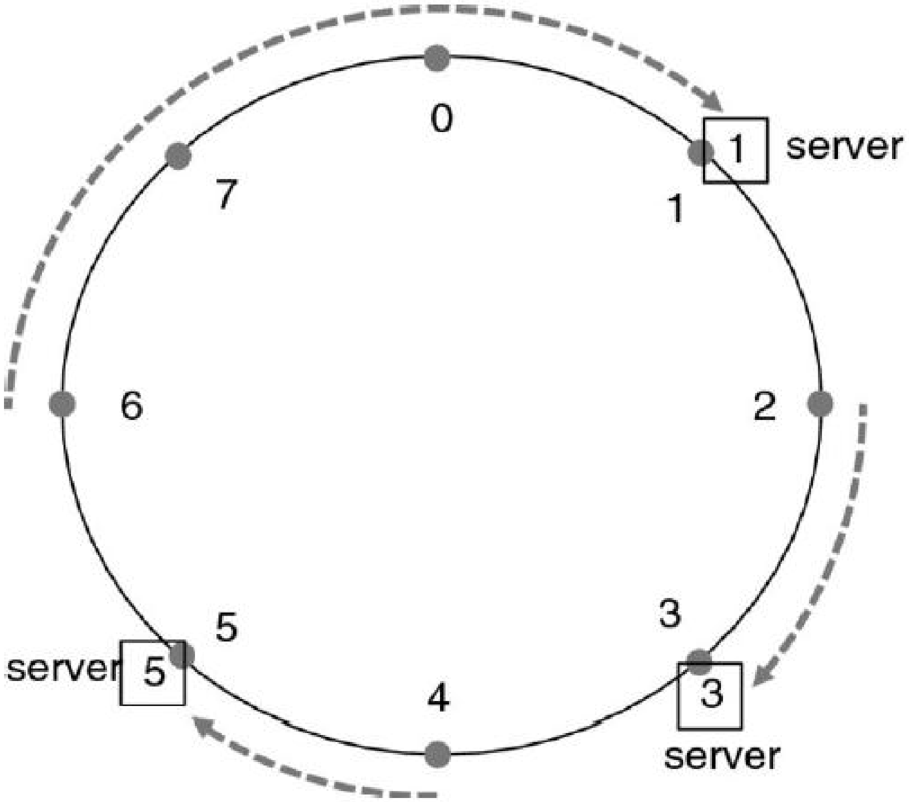
Obviously this is undesirable, and so another scheme is in order. To this end, mem cached implementations usually employ variants of consistent hashing to minimize the updates required as the server pool changes and maximize the chance of having the same server for a given key. The basic algorithm of consistent hashing can be outlined as follows:

I. Map the range of a hash function to a circle, with the largest value wrapping around to the smallest value in a clockwise fashion;

II. Assign a value (i.e., a point on the circle) to each server in the pool as its identifier, and

III. To cache a data item of key k, select the server whose identifier is equal to or larger than H(k).

The server selected for key k is called k's successor, which is responsible for the arc between k and the identifier of the previous server. As an example, below figure shows a circle of three servers, where server 1 is responsible for caching the associated data items for keys hashed to 6, 7, 0, and 1; server 3 for keys hashed to 2 and 3; and server 5 for keys hashed to 4 and 5.



An immediate result of consistent hashing is that a departure or an arrival of a server only affects its immediate neighbors. In other words, when a new server p joins the pool, certain keys that were previously assigned to the original p's successor will now be re-assigned to server p, while other servers are not affected. Similarly, when an old server p leaves the pool, the keys previously assigned to it will now be reassigned to p's successor while other servers are not affected. Adding a new server 7 would result in reassigning keys 6 and 7 to the new server; removing server 3 would result in reassigning keys 2 and 3 to server 5.

(Reference: Cloud Computing: Business Trends and Technologies)