CRITIQUE-1
Lecture 01

In this lecture of computer storage systems is about brushing up the concepts of the operating systems and introduction to the computer storage system. It laid a foundation for the upcoming seminars by briefly going over the basics of Operating and storage systems and clearing up any doubts the listener may have. The lecture / discussion was divided into three parts covering three different topics yet related to storage systems. First is the basics of operating system, how OS makes the hardware resources like memory and disk available to the user, how the data is organized in the disk using file system. Second is the Volume management, RAID technology, and last is the appliance fundamentals such as client interface, memory etc.

In first part of the lecture - Basics of operating systems. I learnt how the operating system makes the hardware resources like main memory, disk etc available to the user or application to use. The OS is also responsible for fair sharing the resources between different applications i.e time slicing resources. The Operating system is responsible for all housekeeping management and coordination of all of the peripheral devices. Operating system also virtualizes the main memory to the applications. This virtualization makes every application to assume that the entire RAM is available for itself to run the program. But infact the OS is just swapping the pages(regions in memory) in and out of the RAM to the disk(also called as swap area/ fake RAM) on demand to satisfy the needs of the application.

File systems manages the storing and retrieval of persistent data on the disk. Without file system the disk would just be a blob of data without any metadata i.e not knowing where chunks of data starts or ends. Multiple file system can be present however there can be only one /root and all other file systems has to be mounted underneath the root folder. Virtual file system was a new concept that I was not aware of. The virtual file system is an abstraction layer on top of the more concrete file systems like NTFS, HFS+, ZFS etc. This abstraction layer allows the application to access different file systems in a uniform way. One can think of this VFS as an interface between the operating system(kernel) and concrete file system. It would have been nice if at least one of the concrete file system and VFS was explained in more detail. I hope this would be covered in the future seminars. VFS operations can be split into two those that operate on objects(like files, directories etc) are called VNOPS. And those that operate on file system called VFSOPS. VNOPS(Virtual node operations) defines interfaces for creating, destroying objects. VFSOPS(virtual file system operation) defines interfaces that can operate on file system level like create, destory FS(file system), mount and unmount FS. Allocating disk space for the writes could be done either with pre-allocated disk space or “just in time” allocation. Traditional disk overwrite can cause data corruption if the write fails, this is can be avoided by using COW(copy on write). COW ensures that the old data is not corrupted or overwritten by not mapping new data block until the writes are completed.

The second part of the lecture was Volume management and volume manager. Volume management provides a higher level view of the disk storage on computer system than the

conventional disk and partitions. This would allow the administrators to dynamically allocate the storage resources to the application or users. Volume management increases the performance of the file system by distributing the concurrent IO across multiple drives.

RAID which stands for Redundant array of Inexpensive(Independent) drives is a volume manager. There are different levels of RAID 0,1,2...6, 10 etc. RAID 0 is not ideal for mission critical system but provides an increased performance in read/writes by striping the data across all drives. RAID 1 is used for backup. It's a mirroring technique. No gain in performance but provides reliability. RAID 2 minimum of 3 drives. Data is striped on a bit level and a parity is calculated and written to a separate disk. RAID 3 is same as RAID 2 but data is striped on byte level. RAID 4 works similar to RAID 2 and 3 but the data is striped on a block level. Though RAID 2,3, 4 all provide a data correction or rebuilding the data on a disk failure it uses most of the disk space for storing the parity. But if the parity disks were to fail along with data drive then there's on recovery.RAID 5 works most like RAID 4, however rather than storing the parity on a dedicated disk drive it's stripes the parity across all available disks. Thereby eliminating the problem of parity disk failure and all disks are used for storing the data and parity. RAID 10 is a hybrid level that combines the techniques of RAID 0 (striping) and RAID 1(mirroring) to improve the performance of IO's and make the volume more reliable. If there is any temporary or permanent drive failure then "Resilvering" or "Rebuilding" techniques are used to get the data back on the storage.

In the last part of the lecture appliance fundamentals where hardware as a platform to serve data was discussed. The important aspects of this hardware consists of CPU, memory, client interfaces and disk. CPU cycles are not consumed when IO are issued however if the data needs to be protected like encryption,decryption, compression etc does take CPU cycles. Performance gain could be obtained by adding various levels of caching. High speed cables like ethernet, fiber channel, infiniband etc can be considered to avoid bottlenecks during IO on the client interfaces. Even though data is encrypted and stores there are still various ways we can lose the data on the drive such as drive failure, degradation of data over time(bit rot), phantom writes (writes that doesn't make it to the disk), read or write happening to or from a wrong location on the disk (misdirected read/write) etc.

To achieve data protection we have to go with data duplication. We can use any of the RAID techniques are discussed previously for data backup. I learnt two new terms hardware RAID and software RAID. Software RAID is a RAID task that runs on a CPU where as Hardware RAID uses it's own processor and memory to run the RAID application. Some solutions include a mix of software and hardware RAID implementations called Hybrid RAID.

References:

http://www.adaptec.com/nr/rdonlyres/14b2fd84-f7a0-4ac5-a07a-214123ea3dd6/0/4423_sw_hwraid_10.pdf
https://www.youtube.com/watch?v=wTcxRObq738

References:

\url(http://www.adaptec.com/nr/rdonlyres/14b2fd84-f7a0-4ac5-a07a-214123ea3dd6/0/4423_sw_hwraid_10.pdf)

https://www.youtube.com/watch?v=wTcxRObq738