VirtuCache Case & VirtuStor Study

orange™

USE CASE
- Primary Storage
- Video surveillance Storage for 400 cameras over 40 VMs

LOCATION
- Europe

CHALLENGES
- Reduce the cost of storage for video surveillance without losing performance.

SOLUTION
- VirtuStor replaced their existing storage product for long term storage of video footage using low cost, high capacity HDD.
- Super-fast access to metadata on NVMe, PCIe, SSD for reads delivers exceptional performance

The VirtuCache Difference

Using VirtuStor and VirtuCache, Orange was able to transform video surveillance – by enabling low cost with high performance.

High Performance, Low Cost for Video Surveillance Storage

Here are unique requirements of video storage, some are obvious and others not so much, that inspired us to put together a different architecture than the conventional storage OEM design.

1. **Large amounts** (TB/PBs) of storage are required even for relatively small organizations with small budgets.
2. **Right sized storage.** Since video storage needs to be inexpensive, it needs to be right sized, without the customer overcommitting capacity or performance upfront. Consequently, after the initial rollout, capacity needs to scale up as needed and in smaller increments.
3. **Cheap.** Video storage, especially for surveillance video, is considered less important than mission critical data and so it is allocated less money than the organization's regular production storage. As a result, it demands operational simplicity. For this reason, some organizations are moving towards storing video data in the cloud. Cloud subscription costs over the long term prove more expensive than on-premises storage.
4. **High availability.** Despite such low-cost requirements, video storage still needs enterprise grade high availability. The liability for lost video because of hardware or software failure is just too high, especially in the case of law enforcement video.
5. **Deficiencies of RAID.** Related to the above point, is the fact that if the volume of data is large and if RAID is being used to protect against drive failure, the time it takes to rebuild a failed drive will be high, especially if large capacity drives are used. For 12TB hard drives, RAID5 rebuild times are in days and RAID6 rebuild time could be two weeks. So an alternative to RAID to protect against drive failure is required.
6. **Storage appliance dedupe and compression are useless for video.** Most video is captured and managed by a video management system that typically has built-in codec (encoder-decoder) software. The video encoding functionality within these systems ensures that the video is deduped and compressed. As a result, dedupe and compression features in expensive major brand OEM storage appliances are of little value to video.
7. **IO path traversal optimized.** Video comprises of two interconnected data, the recording and its associated metadata. The video recording is large and written directly to the file system, whereas the associated metadata is small and written to a database.
   - **Regarding writes** – Video capture process is write intensive. For instance in-store video and police dash cam/body cam video have a pattern of intense bursty writes and long periods of relative quiet. Conversely, in cases where the encoder for some reason doesn't compress or dedupe well, or if the video is action packed, the writes will be continuous and sequential. The storage functionality should be such that if large amounts of sequential writes start to happen, those should be written to hard drives, bypassing the SSDs, since hard drives process sequential data well.

VirtuCache & VirtuStor Case Study

- Conversely in case of bursty random writes the writes should be first written to SSDs and then flushed to hard drives, since the SSDs will act as a 'shock absorber' and will allow large amounts of writes to be processed quickly.
- **Regarding reads** – In case of video surveillance/law enforcement video, video retrieval is infrequent as it happens only when there is a case related search request. If all the metadata is on SSDs and video data reside on hard drives, the searches will be fast.

Ultimately, storage does not need to be high speed all of the time. It needs to service the reads fast, in the rare case of a search request, and it needs to be able to write the video stream to storage without undue queuing. Since the metadata even for large amounts (~ PBs) of recorded video is small and can easily fit on a single SSD, the storage system should strive to keep metadata on SSDs all the time. And of course all these storage decisions about how the IO path traverses various media depending on workload characteristics needs to be automatically and mathematically done. Yes, you can have an all-SSD storage to keep even the video files on SSDs, making the storage software logic simpler, but the uptick in cost just doesn't justify all-flash storage for video. Also since large video files are retrieved and written sequentially, SSDs don't offer a significant performance benefits versus hard drives.

Video surveillance infrastructure at Orange
Orange wanted to offer a video surveillance system to its customers. Each surveillance 'pod' comprises of video surveillance software running in VMs that captured, searched, and managed video. Two VMware hosts hosted the video management VMs. Each VMware host had Virtucache caching to 9TB SSDs in each host.

This VMware cluster was connected over iSCSI to a 3-server 36TB usable VirtuStor storage cluster. Each VirtuStor server contained 3x 8TB HDDs for storage, 300GB SSD for journaling, 4 core Xeon processors, and 32GB RAM. Replication factor was 2x, thus usable capacity was halved and at 36TB (raw 72TB).

Here is how this system addressed each of the requirements for video storage.
1. **3-server VirtuStor storage cluster.** 36TB of storage was the minimum required, with the ability to scale up to 200TB by simply adding additional SSDs and hard drives to the same 3 servers in the VirtuStor cluster.
2. **Operational simplicity**
   - **Simple GUI.** The VirtuStor cluster has only 3 easy to navigate pages to configure and monitor storage.
   - **Any make or model.** In case of hard drive, SSD, or server failure, a replacement can be added hot. Also, any failed hardware can be replaced with similar part of any make or model, so long as it has a driver for Linux, which virtually all server parts do.
3. **$90K for 200TB storage and 18TB SSDs.** The entire compute + storage cluster cost less than $50,000 and as it scales out to 200TB, the incremental cost would be another $40,000. So for a total of $90,000, Orange would get 200TB iSCSI storage with VirtuStor, and 18TB SSDs with VirtuCache on the VMware hosts.
4. **Replication, not RAID.** VirtuStor doesn't use RAID, it uses replication instead. With VirtuStor data is copied in place to two different servers within the VirtuStor cluster (one hard drive in each server). Because we don't use RAID, there are no parity bits that need to be searched and so rebuild times are faster than with RAID. A failed drive can be rebuilt at the speed at which data can be copied back to the replacement drive. Also because rebuild times are smaller with VirtuStor, high capacity (12$^+$TB) drives can be used.
5. **Hot add/replace.** The failed drive or server can be replaced hot. Once the drives or server is replaced, VirtuStor copies data back to the drives as fast as the network allows.
6. **Low cost processors as no dedupe and compression is needed.** Since video codec takes care of dedupe and compression for video, conventional storage controller based dedupe and compression are of little use. Thus, customers don't need expensive processors that are typically required by these two features. For instance, the video surveillance infrastructure described in this article supported 400 cameras streaming video through 40 VMs and writing video at 50MBps using a single 4 core Xeon processor in each VirtuStor node. The codec, which is CPU intensive, was in the video management software that was running in VMs and as is the case with VMware compute nodes, those had beefier processors that were now put to good use by the codec.

VirtuCache & VirtuStor Case Study

7. **IO Path traversal logic between host side SSDs and backend hard drives.** To understand how data flows through the system, you have to understand how VirtuCache and VirtuStor work in tandem.
   - VirtuCache caches frequently and recently used random reads and writes to in-host SSDs.
   - If the data pattern is sequential, VirtuCache doesn't cache it and hence it is read from or written to hard drives on backend VirtuStor iSCSI SAN.
   - Now all metadata in both recording and retrieval operations is random and so it will always be cached to in-host SSDs by VirtuCache.
   - Typically metadata is 1/100th of the actual video recording and so even for petabyte scale video storage, this entire metadata will be cached to in-host SSDs. This results in quick searches.
   - If the video recording is bursty like it most likely is, then it too will get cached to in-host SSDs, thus keeping pace with the write bursts. If the video recording or searches are sequential, the videos will be written to and read from the VirtuStor hard drives, which is a good thing, since hard drives perform well for sequential data.
   - It is only for random and bursty reads & writes that hard drives completely break down in performance, and in these cases in-host SSDs act as a 'shock absorber' to absorb the bursty random reads and writes.

Summary

By deploying SSDs closer to the host CPUs (VirtuCache) and high capacity hard drives in backend storage (VirtuStor); and going up against conventional storage architecture wisdom to not use dedupe and compression (and thus reduce CPU costs); and not use RAID (to be able to use 12$^+$TB hard drives and reduce drive rebuild time in case of drive failure); we are able to offer low cost per TB for capacity and low cost per MBps for throughput, ideally suited for video storage.