**BOSCH**

Invented for life

# Software Defined Video Surveillance Storage

Author:     Verhaeg Mario (BT-SC/PAS4-MKP)
Date:       18 September, 2019

# 1 Document information

| Project | BVMS |
|---|---|
| Reference | Software Defined Video Surveillance Storage with SUSE and HPE. |
| Version | 28 |
| Last modified | 📅 18 September 2019 |

## 1.1 Version history

| Version | Date | Author | Comment |
|---|---|---|---|
| 28 | 2019-18-09 | Verhaeg Mario (BT-SC/ PAS4-MKP) | |

# 2 Introduction

> Please consult a certified SUSE, HPE, or Bosch pre-sales engineer to assist in designing a tailored solution.

As video surveillance use grows in commercial, government and private use cases, the need for low-cost storage at scale is growing rapidly. BVMS, Bosch cameras, HPE hardware and SUSE Enterprise Storage provide a platform that is an ideal target for recording these streams.

There are numerous difficulties around storing unstructured video surveillance data at massive scale. Video surveillance data tends to be written only once or become stagnant over time. This stale data takes up valuable space on expensive block and file storage, and yet needs to be available in seconds. With this massive scale, the difficulty of keeping all the data safe and available is also growing. Many existing storage solutions are a challenge to manage and control at such scale. Management silos and user interface limitations make it harder to deploy new storage into business infrastructure.

The solution is software-defined storage (SDS). This is a storage system that delivers a full suite of persistent storage services via an autonomous software stack that can run on an industry standard, commodity hardware platform. Bosch, Hewlett Packard Enterprise (HPE) and SUSE have partnered to deliver the benefits of SDS to the video surveillance industry. Using SUSE Enterprise Storage™ on HPE ProLiant DL and Apollo servers in a Bosch video surveillance environment simplifies the management of today's volume of data, and provides the flexibility to scale for all enterprise storage needs.

# 3 Solution

## 3.1 Scalability

Video surveillance systems need to store large numbers of data. To enable security operators to effectively manage security incidents when they happen, this data also needs to be available at all times. Based on the Bosch Video Recording Manager, SUSE Enterprise Storage and the HPE line of density-optimised servers, a scalable storage solution on different levels can be achieved.

Scalability meets adaptability with HPE's line of density-optimised servers: Powerful computing with multiple systems on a chip (SoC), Speed of provisioning and the time to service are increased with built-in fibre and software for connectivity, scalability and adaptability. Make your move toward modular, hyper scale and ultra-converged infrastructure.



On top of the physical server infrastructure, SUSE Enterprise Storage allows for unlimited scalability with a distributed storage cluster designed to scale to thousands of nodes and multi-hundred petabyte environments and beyond to meet the growing data requirements.

Last, but not least, the Bosch Video Management System (BVMS) allows to scale up to 256 petabyte per system of 2000 cameras. Multiple BVMS systems can be combined into a massive BVMS Enterprise system, potentially spread out across the world. Existing BVMS systems can be easily expanded using a scale-out configuration: just add a new storage system to the Bosch video surveillance environment and the system will automatically take this into it's virtual pool of available storage capacity. There is no need to re-assign cameras to NVRs or re-configure NVRs to use a different storage system.

All of this data is put at the fingertips of a video surveillance security operator. With Forensic Search, the operator can browse recorded images based on movement, but also on criteria, like the size or colour of objects. So, with just a few clicks, huge data volumes can be searched for all objects recognized as, for example, a person or a yellow taxi.
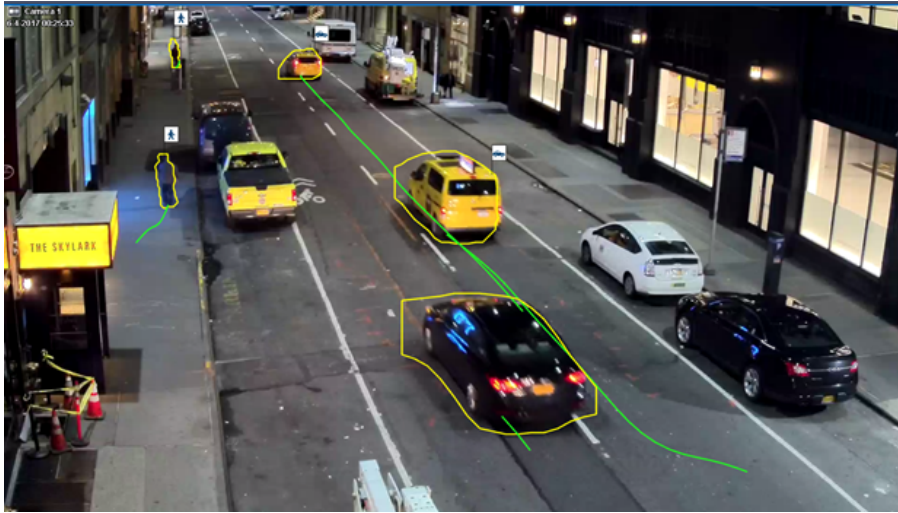
## 3.2 Performance

HPE's high-density server family delivers breakthrough performance with efficient rack-scale compute, storage, networking, power and cooling for your most demanding massive data analytics and object storage workloads.

The latest iteration of Ceph, available in SUSE Enterprise Storage, offers BlueStore, which doubles the write performance of previous releases and significantly reduces input and output latency. The ability to provide a write-back cache tier enables the system to also service high performance short-term streams where only a percentage of requests actually end up being served from the long-term archive.

The Bosch Video Recording Manager automatically balances the load across the available (SUSE Enterprise) storage targets, depending on a target specific performance configuration. This allows for very flexibly configurations, in which lower performance storage systems can be combined with high performance storage systems, while distributing the load on those systems accordingly.

## 3.3 Availability

HPE InfoSight gathers operational intelligence from an infrastructure by analysing millions of sensors across a globally-connected installed base, and using behavioural data provides trend insights, forecasting and recommendations, to predict and prevent problems. The result is higher efficiency and reliability, creating a smarter, easier-to-manage infrastructure for customers.

SUSE Enterprise Storage comes with erasure coding, which lets you define the settings for data protection. You can determine how many device failures your cluster can tolerate before considering the data compromised.

Compared to other video management systems BVMS allows for maximum resilience. This means continuous live and playbacks – no matter what happens. While other systems require you to add redundant components to create a resilient system, BVMS offers several levels of resilience out of the box. As a result, cameras keep recording and streaming live video even when multiple system components fail simultaneously.
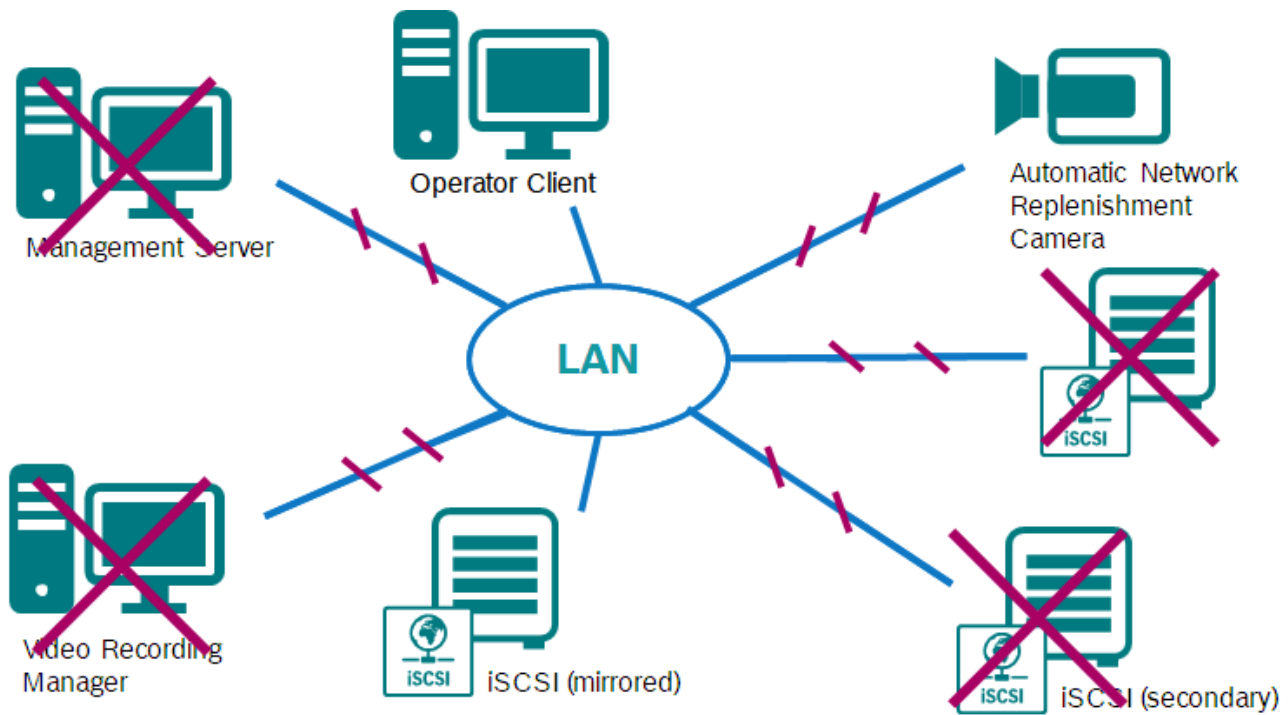
## 3.4 Data Security

HPE is the only vendor to provide silicon root of trust on Gen10 servers, which creates a digital fingerprint in the silicon and ensures that the server will never boot with compromised firmware. This root of trust and other unique security features have garnered HPE recognition for having the "world's most secure industry-standard servers. The SUSE Enterprise Storage operating system supports trusted boot scenarios to ensure that only signed kernels and drivers can be booted from and loaded. Because video data is often highly critical and sensitive, Bosch is driving a systematic approach to maximize data security by considering physical safety and cybersecurity simultaneously.

## 3.5 Proven configuration

The combination of SUSE Enterprise Storage, HPE and Bosch video surveillance components has been extensively tested. Not only the performance of the solution has been proved: several failure scenarios are considered to ensure the system continues to operate as expected.

# 4 Technology

## 4.1 SUSE Enterprise Storage - Powered by CEPH

### 4.1.1 Introduction

Ceph is the most popular OpenStack software-defined storage solution on the market today. It is extensively scalable from a storage appliance to a cost-effective cloud solution. It also provides industry-leading storage functionality such as Unified Block and Object, Thin Provisioning, Erasure Coding, and Cache Tiering. What's more, it is self-healing and self-managing.

### 4.1.2 Architecture
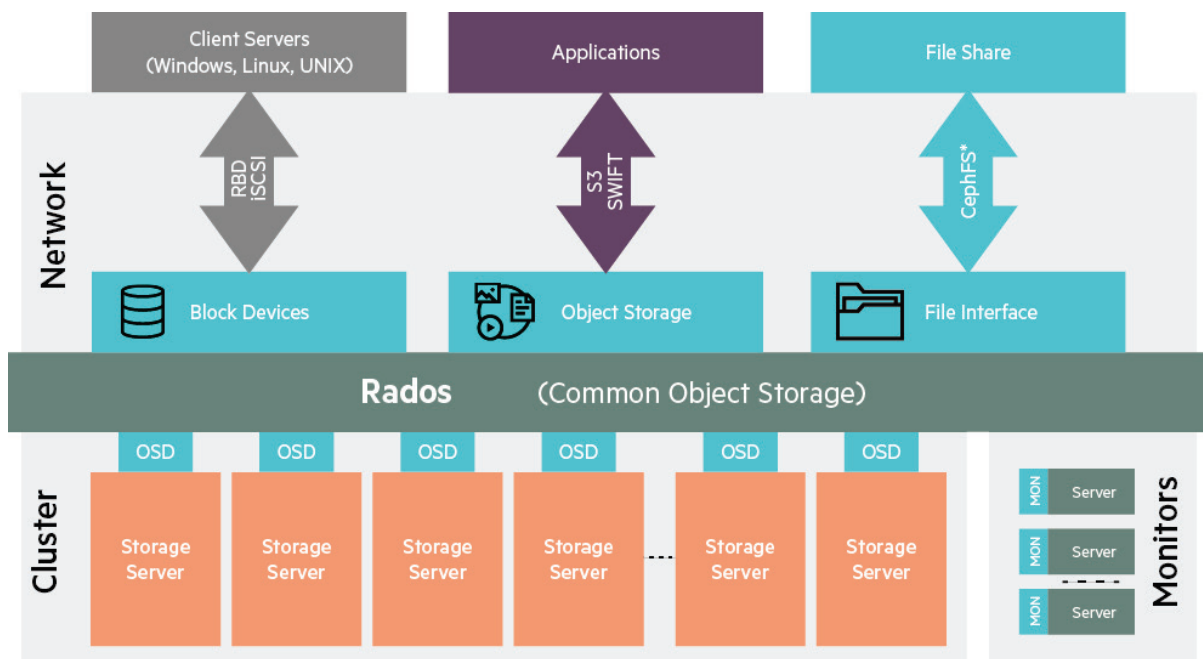


There are three primary roles in the SUSE Enterprise Storage cluster covered by this sample reference configuration:

OSD Host—Ceph server storing object data. Each OSD host runs several instances of the Ceph OSD Daemon process. Each process interacts with one Object Storage Disk (OSD), and for production clusters, there is a 1:1 mapping of OSD Daemon to logical volume. The default file system used on each logical volume is XFS, although Btrfs is also supported.

Monitor (MON): Maintains maps of the cluster state, including the monitor map, the OSD map, the Placement Group Map, and the CRUSH map. Ceph maintains a history (called an "epoch") of each state change in the Ceph Monitors, Ceph OSD Daemons, and Placement Groups (PGs). Monitors are expected to maintain quorum to keep an updated cluster state record.

Administrator: This is the self-master and hosts openATTIC, the central management system which supports the cluster.

RADOS Gateway (RGW)—Object storage interface to provide applications with a RESTful gateway to Ceph Storage Clusters. The RADOS Gateway supports two interfaces: S3 and Swift. These interfaces support a large subset of their respective APIs as implemented by Amazon and OpenStack Swift.

A minimum SES v4 cluster should contain:

- One administrator (typically a ProLiant DL360 server)
- Three or more MON nodes (typically ProLiant DL360 servers)
- Three or more OSD nodes (recommended Apollo 4000 servers)
- One or more RGW (typically ProLiant DL360 severs)
- Optional: iSCSI gateway (one or more ProLiant DL360 server)

Density-optimized Apollo 4000 servers are ideal for use as the bulk storage OSD nodes. Ceph supports mixing Apollo 4000 server types and generations, enabling seamless growth with current technologies.

## 4.1.3 Keeping data safe

SUSE Enterprise Storage brings Ceph's flexibility to bear by supporting data replication as well as erasure coding. Erasure coding mathematically encodes data into a number of chunks that can be reconstructed from partial data into the original object. This is more space efficient than replication on larger objects, but it adds latency and is more computationally intensive. The overhead of erasure coding makes it space inefficient for smaller objects, and block storage requires a replicated cache tier to utilize it. As such, erasure coding is recommended for capacity efficiency, whereas replication is most appropriate for lower capacity block storage and small objects.

## 4.1.4 Putting data on hardware

One of the key differentiating factors between different object storage systems is the method used to determine where data is placed on hardware. Ceph calculates data locations using a deterministic algorithm called Controlled Replication Under Scalable Hashing (CRUSH). CRUSH uses a set of configurable rules and placement groups (PGs) in this calculation. Placement groups tell data where it is allowed to be stored and are architected in such a way that data will be resilient to hardware failure.

# 4.2 HPE value for a Ceph storage environment

Hewlett Packard Enterprise (HPE) and SUSE have partnered to deliver the benefits of SDS on reliable hardware. SUSE Enterprise Storage™ on HPE ProLiant DL and Apollo servers can simplify management of today's volume of data— and provide the time to value, cost control and flexibility to scale for all of your enterprise storage needs.

HPE hardware gives you the flexibility to choose the configuration building blocks that are right for your business needs. The HPE Apollo 4000 Gen10 server systems are most suited for the task and allow you to find the right balance between performance, cost-per-gigabyte, building block size, and failure domain size.

## 4.2.1 Introduction

Software defined storage running on Linux servers can be deployed on a variety of hardware platforms. However, clusters built on a white-box server infrastructure work for business at small scale, but as they grow, the complexity and cost make them less compelling than enterprise hardware-based solutions. With white-box server infrastructure, IT has to standardize and integrate platforms as well as supported components themselves, and support escalation becomes more complicated. Without standardized toolsets to manage the hardware at scale, IT must chart their own way with platform management and automation. Often the result is the IT staff working harder and the businesses spending more to support a white-box hardware infrastructure than the one-time CAPEX savings realized in buying the white-box servers.

Using an HPE hardware and software solution provides advantages that reduce OPEX spending not available in an infrastructure built on white-box servers. Key OPEX savings from using an integrated HPE solution are:

- Platform management tools that scale across data centers
- Server components and form factors that are optimized for enterprise use cases
- Hardware platforms where component parts have been qualified together
- A proven, worldwide hardware support infrastructure

## 4.2.2 Disk encryption

In addition to the benefits above, all Apollo 4000 configurations include an HPE Smart Array card capable of secure encryption where enterprise-class encryption is needed. Encryption is FIPS-2—certified for security, has been tested as not affecting IOPS on spinning media for low-performance impact, and is transparent to the operating system for ease-of-use. This means any drive supported on the server can be used, giving much more cost/performance flexibility than encryption on drive solutions. Key management is simple and can be managed locally or via an enterprise key management system. hpe.com/servers/secureencryption

## 4.2.3 Multi-generational Ceph support

Ceph cluster support mixing multiple generations of x86 server storage nodes. Apollo 4510 Gen10 can be used to expand existing storage cluster based on Apollo Gen9 systems.

# 4.3 BVMS and Software Defined Video Surveillance Storage

## 4.3.1 Introduction

Video surveillance cameras generate more data as a result of higher resolutions and frame-rates. This has a direct impact on the storage costs of the entire video surveillance environment. The storage consumption can be reduced to use alarm recording instead of continuous recording. This means the system will only start the recording when an alarm is generated (for example, when motion is detected). However, in some projects continuous recording is still required. This entire concept is described in a separate whitepaper: BVMS - Policy Based Recording

## 4.3.2 Video Recording Manager

The Bosch Video Recording Manager (VRM) is at the heart of the recording system. Instead of acting like an NVR (which takes video streams from the cameras and stores it on a storage environment) it just tells the Bosch cameras where to record their video. This means that the VRM is not involved in the recording itself, which has a couple of benefits.
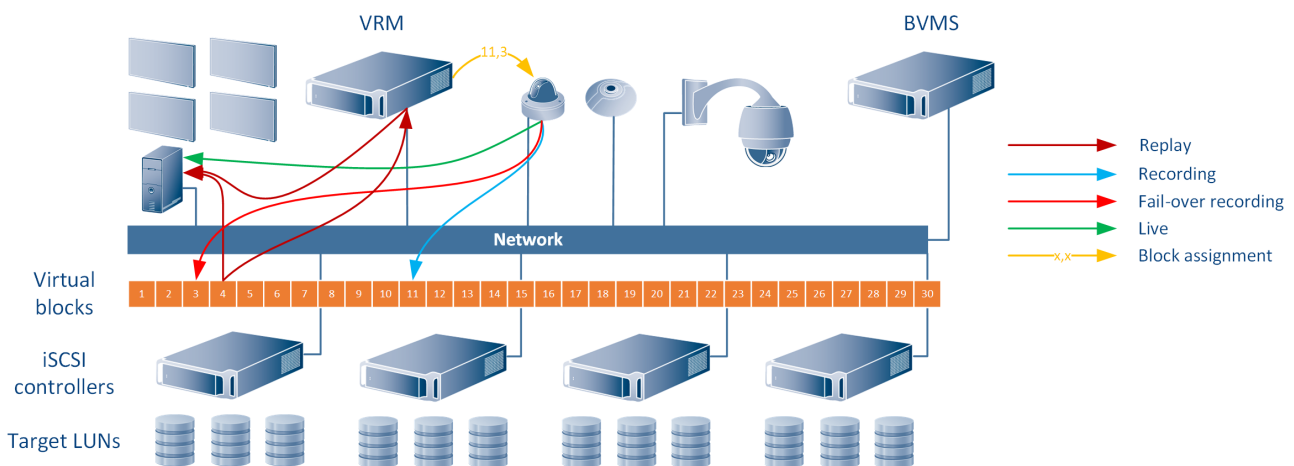


The first benefit relates to performance: one VRM is able to manage the recording for as much as 2000 cameras an 2 petabyte of storage, while a typical NVR struggles to handle 300 cameras. One BVMS system can handle up to 128 VRMs, which results in a maximum of 256 petabyte per BVMS system.

The second benefit relates to resilience: if the VRM fails, the cameras continue to record for a configurable amount of time. Depending on that configuration, the VRM ensures the cameras have a list of potential storage locations (or block) cached. The camera itself checks the availability of the location and, if a storage location is unavailable or goes off-line unexpectedly, it will automatically fail-over to the next location on the list. The camera even has a built-in cache which stores the last couple of seconds of video. This cache is flushed to the next storage location when a fail-over occurs, which results in a zero frame-loss fail-over. When the VRM fails, the camera offers as a fail-over replay path. Optionally a fail-over VRM can be added to the environment.

The third benefit relates to scalability: the VRM creates virtual blocks on the storage locations, which are reserved for a specific camera (depending on the configured retention time and failure options). If storage capacity is added to the system, the VRM re-calculates the available storage space and distributes the new virtual blocks to the cameras which are active in the system.

# 5 Test details

## 5.1 System configuration

### 5.1.1 Hardware

| Purpose | Amount | Type | RAM (GB) | CPU | Boot Storage | NVME | Ceph Storage | Mgt network interface | Network Interface |
|---|---|---|---|---|---|---|---|---|---|
| Mgt Server | 1 | xl170 Gen9 | 64 | 2x Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz | 2x 1TB RAID 0 (P40) | NA | NA | 1x GbE (inline) + ILO port | 100GbE PCIe card 1 port (public) |
| Monitor server | 3 | xl170 Gen9 | 32 | Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz | 2x 1TB RAID 0 (P40) | NA | NA | 1x GbE (inline) + ILO port | 100GbE PCIe card 1 port (public) |
| OSD Server | 4 | A4200 Gen9 | 256 | 2x Intel(R) Xeon(R) CPU E5-2650 v4 @ 2.20GHz | 1x 2TB HDD | 2x 800GB NVMe | 23x 2TB HDD | 1x GbE (inline) + ILO port | 100GbE PCIe card 1 port (public) |
| iSCSI servers | 2 | XL170 Gen9 | 64 | 2x Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz | 2x 1TB RAID 0 (P40) | NA | NA | 1x GbE (inline) + ILO port | 100GbE PCIe card 1 port (public) |
| Camera simulators | 11 | m710 | 32 | Intel(R) Xeon(R) CPU E3-1284L v3 @ 1.80GHz | 480GB SSD | NA | NA | 1x 10GbE link | 1x 10GbE link |

### 5.1.2 Software

Available on request at HPE, SUSE or Bosch.

# 5.2 Process

Describe test process.

# 5.3 Results

## 5.3.1 Recording

| Cameras | Total bitrate | Storage Write | Passed |
|---------|---------------|---------------|--------|
| 400 | 1100 Mbps | 275 MBps | **PASSED** |

## 5.3.2 Record and repla

| Cameras | Write performance | Replay sessions | Read performance | Network load client (IN/OUT) |
|---------|-------------------|-----------------|------------------|------------------------------|
| 400 | 1100 Mbps | 8 | 5 MBps | 10/110Mbps |

## 5.3.3 Link loss

The iSCSI gateway is disconnected from the network. The camera re-connect time (time it takes before recording is restarted) will be measured.

| Cameras | Seconds recording restart for all cameras |
|---------|-------------------------------------------|
| 400 | 10 |

## 5.3.4 Record, replay and rebuild

Disk is removed from the system

| Cameras | Write performance | Replay sessions | Read performance | Network load (IN/OUT) |
|---------|-------------------|-----------------|------------------|------------------------|
| 400 | 275 MBps | 8 | 5 MBps | 1100 Mbps / 110 Mbps |

## 5.3.5 Node loss

One OSD node is disconnected from the system.

| Cameras | Write performance |
| --- | --- |
| 400 | 275 MBps |