# Deduplication and Compression

Introduction

The amount of data in the virtualized datacenters is continuously growing, making VMcentric storage solutions more and more expensive. This introduces a challenge for vendors to develop storage which can sustain constant data growth and maintain the required level of performance. With server virtualization not yet on its peak there are also VDI's starting to elevate. This means that industry is about to face a much bigger data wave, and the VM-centric storage challenge becomes even more relevant.

Problem

Existing space reduction (deduplication and compression) technologies available on the market are not tailored for the needs of flash based storage. Offline space reduction assumes that data is written to the array twice: first time as is, and second time when it's "dehydrated". This **causes premature flash wear out and doesn't preserve space**, which is critical for arrays with high $/IO and $/GB ratio. Starting the data optimization process after raw data is physically written to the array also requires the array to **maintain space reserved for storing raw data until it's optimized**. Implementing space reduction on commodity SAS and SATA storage arrays doesn't make much sense because space reduction process is stealing a vast amount of IOPS from an initially slow array and g**ives no significant improvement** for the cost per terabyte and cost per IO in return. With the modern high capacity SATA and SAS drives it is more cost effective to add an extra drive instead.
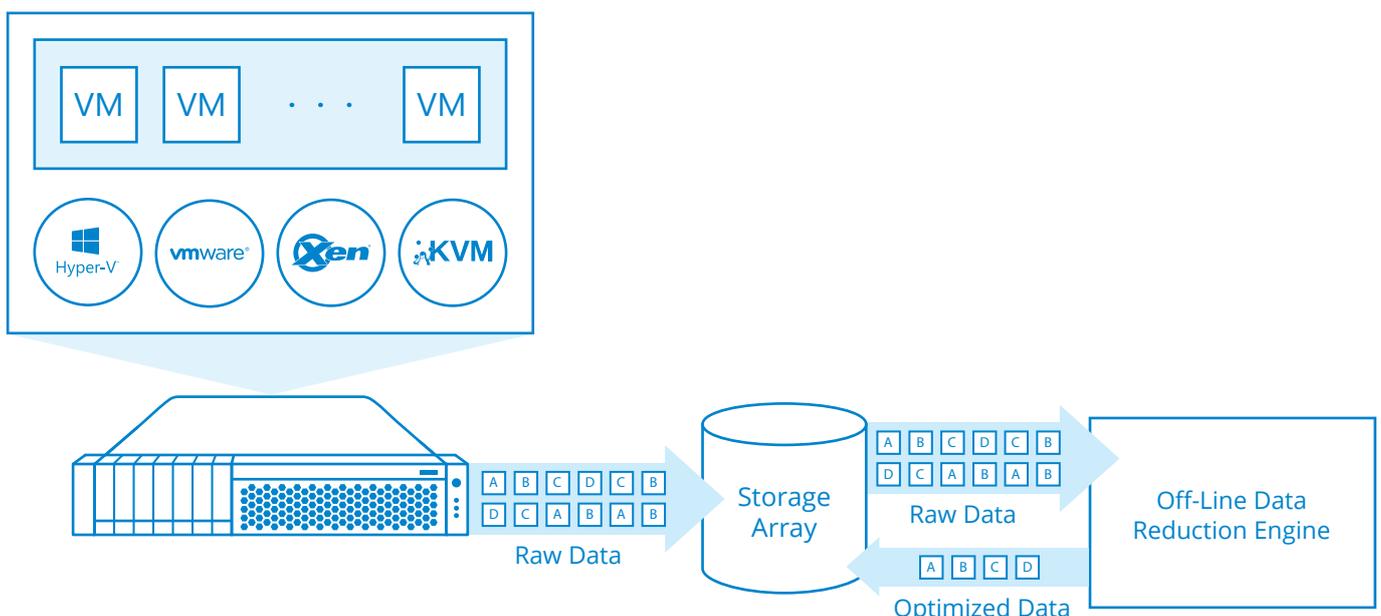

Traditional space reduction technologies are write-intensive

Solution

StarWind Virtual SAN implements in-line deduplication using industry standard 4 KB block for highest effectiveness and optimal deduplication ratio. Deduplication is then followed with optional compression of the written data blocks. Combined with Log Structuring this results in a **double profit**:

- **Amount of data physically written to the array is reduced** which gives more IOPS for the VMs and no IOPS stolen by the data optimization engines

- Less data written also means less erase and write cycles resulting in **much better flash storage utilization and flash cell lifespan**

- Log structuring allows to **get rid of flash drive "spot burns"**

With the VDI scenarios where the amount of overlapping data is close to 90% StarWind inline deduplication makes it possible to elevate the performance further by implementing in-memory computing. With this approach entire VM data is pinpointed in RAM cache resulting in uncompromised performance for virtual desktop infrastructures. A log structured copy of the data is stored on commodity spindle-based storage array to protect the volatile high performance storage layer.
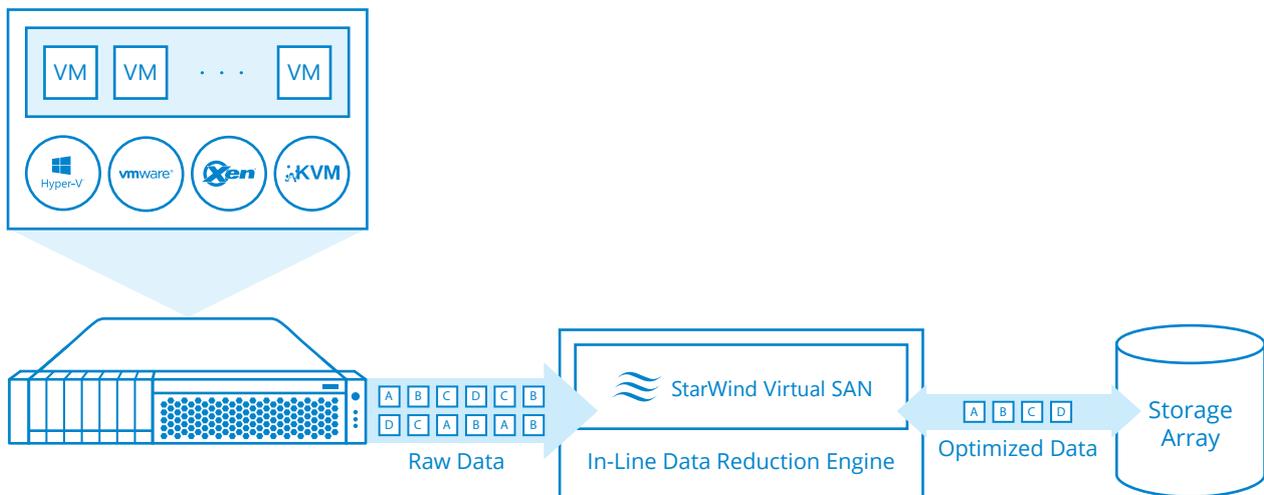
StarWind in-line deduplication writes physically less data to the array

Conclusion

StarWind Virtual SAN dramatically increases the usable space available on the flash based storage by deduplicating the data before it physically hits the storage array. Combined with log structuring, In-line deduplication engine does not cripple performance and steal IOPS from the tier 1 storage. In fact, it significantly increases the performance compared to a scenario where local flash-based storage is used.

In 2016, Gartner named StarWind "Cool Vendor for Compute Platforms".

To learn more, visit **www.starwind.com**