



# Data Efficiency Appliances and Flash Storage Arrays

Louis Imershein

VP Product

Permabit Technology Corporation





For over 50 years, storage was dominated by spinning disk.

- In the HDD world of 2005:
  - Capacity was cheap
  - Performance was scarce
- In flash world of 2015:
  - Capacity costs 8x as much
  - Performance is plentiful



## Data Reduction Changes the Flash Equation

- Vendors are seeing typical 6:1 data reduction rates across a wide range block storage applications
- The benefits of data reduction techniques are multiplicative:

**(10:1 dedupe) \* (5:1 compression) = 50:1 savings**



# Integrated Data Reduction

- **Pros:**
  - No additional latency introduced from additional hops
  - Single point of management
- **Cons:**
  - Data reduction is competing for resources
    - **Memory:** large (>4K) chunk sizes result in significantly lower deduplication rates over time
    - **CPU:** contention encourages deferred processing which complicates capacity planning, interferes with ongoing data management
  - “Always-on” implementations



# Dedicated Reduction Appliance

- **Pros:**

- No impact on storage array resources
- Don't have to buy a new flash array to use it today
- Flexibility/modularity means you can apply where it makes sense and not use where it doesn't

- **Cons:**

- Latency introduced by gateway, cables, HBAs, switch
- Higher cost to equipment: motherboards, adapters, et al.
- LUN configuration managed separately



# Data Reduction Appliance Examples



Product	Example: IBM® SVC
Data Reduction	<ul style="list-style-type: none"><li>• IBM Real-Time Compression™</li><li>• Thin Provisioning</li></ul>

Goes beyond data reduction to incorporate many enterprise storage virtualization features including: tiering, snapshots, clones, replication

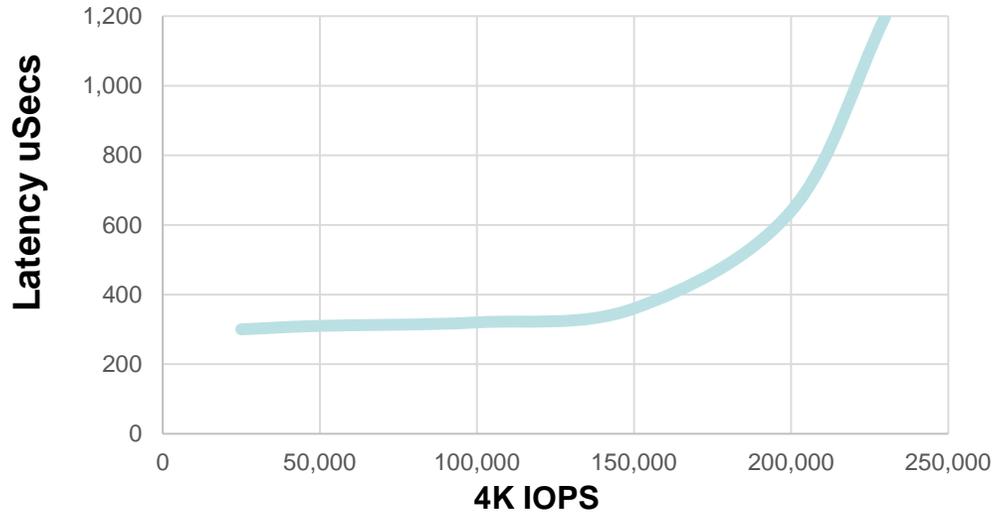
Product	Example: Permabit SANblox™
Data Reduction	<ul style="list-style-type: none"><li>• Permabit Albireo™ deduplication</li><li>• HIOPS compression™</li><li>• Thin Provisioning</li></ul>

Focused 100% on data reduction

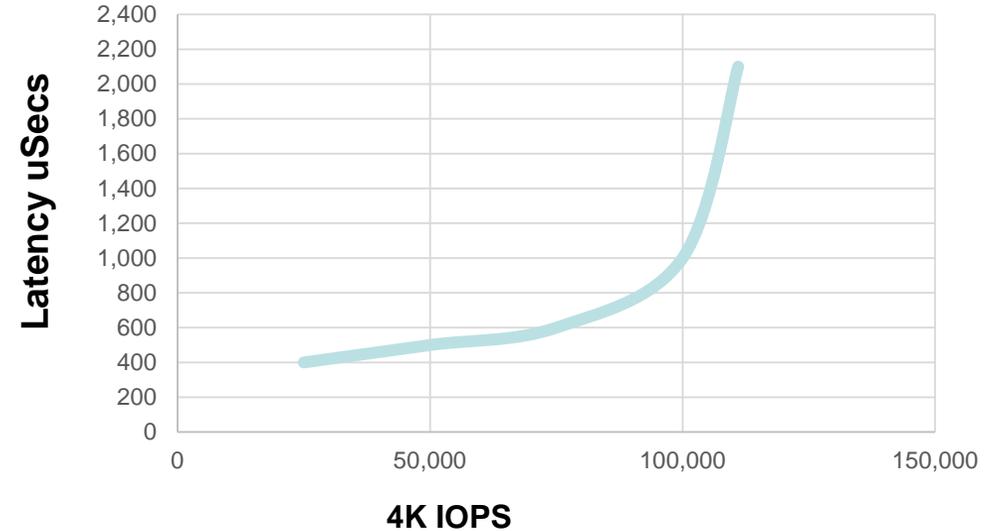


# Appliance Performance

Random Read Performance



Random Write Performance



## Performance per node\*

Random IO (4K IOPS)	Read: 230,000	Write: 111,000	Mixed RW70: 180,000
Min Latency (uSecs)	Read: 300	Write: 400	-
Latency at peak load (avg)	Read: 1,200	Write: 2,100	-

\* Performance measured using fio w/100% unique data



# Comparison of Approaches

	Data Reduction Appliance + AFA	Integrated Data Reduction AFA*
Read Latency @ peak load	1.2 ms	4 ms
Write Latency @ peak load	2.1 ms	5 ms
Mixed Latency @ peak load	3.0 ms	7 ms
4K IOPS Mixed RW70	170,000	150,000

*\* Testing performed across two leading AFAs, both with integrated data reduction*

**Source:** Permabit Labs

**Tools:** fio, vmbench, LoadDynamix



# Appliance Scalability

- Compression is a *performance* challenge
- Deduplication is a *scale* challenge
- For deduplication, current appliances address up to 256 TB of provisioned storage in a single pool
- Admins can start off with 25 TB of provisioned storage, presented as 2.5 PB of logical storage, then grow the backend capacity as needed
- Multiple appliances can be used to address larger capacities

## What about cost?

- Data reduction appliances do introduce extra cost of a motherboard and HBAs
- However, they can utilize CPU and RAM resources more efficiently as a dedicated resource
- Street price per effective GB\* comes down as low as \$0.89/GB for high-end AFAs that ship with data reduction appliances

**That's 1/3<sup>rd</sup> the street price of the leading AFA\***



# Conclusions

- Greater flexibility in deployment doesn't have to come with a higher performance penalty.
- When compared to integrated data reduction solutions, traditional AFA's and data reduction appliances with dedicated resources deliver:
  - Superior performance
  - Lower latency
  - Greater scalability
  - Lower cost