Next Generation Deduplication for Flash Memory

Louis Imershein
Sr. Director of Product Strategy
Permabit Technology Corporation

Todays mainstream data efficiency

- Dedupe and/or compression are integrated into most Enterprise-class flash arrays today
- Target application data types (eg. databases and VMs) frequently see 3X – 35X reduction
- Data efficiency challenges center around the cost of random write performance:

Metric	Mainstream 2014
Performance: 4K random write IOPS per server	150,000
DRAM overhead: GB of RAM / TB of storage	4

Assessing Performance

- Vendors have been less than forthcoming on flash performance with dedupe/compression
- Understanding which systems are fastest can’t be achieved by looking at vendor collateral
- Left to themselves, IT organizations will make their own measurements
- FIO is not the only tool for evaluation, but it’s a good start

Flexible IO Tester (FIO) General Recommendations

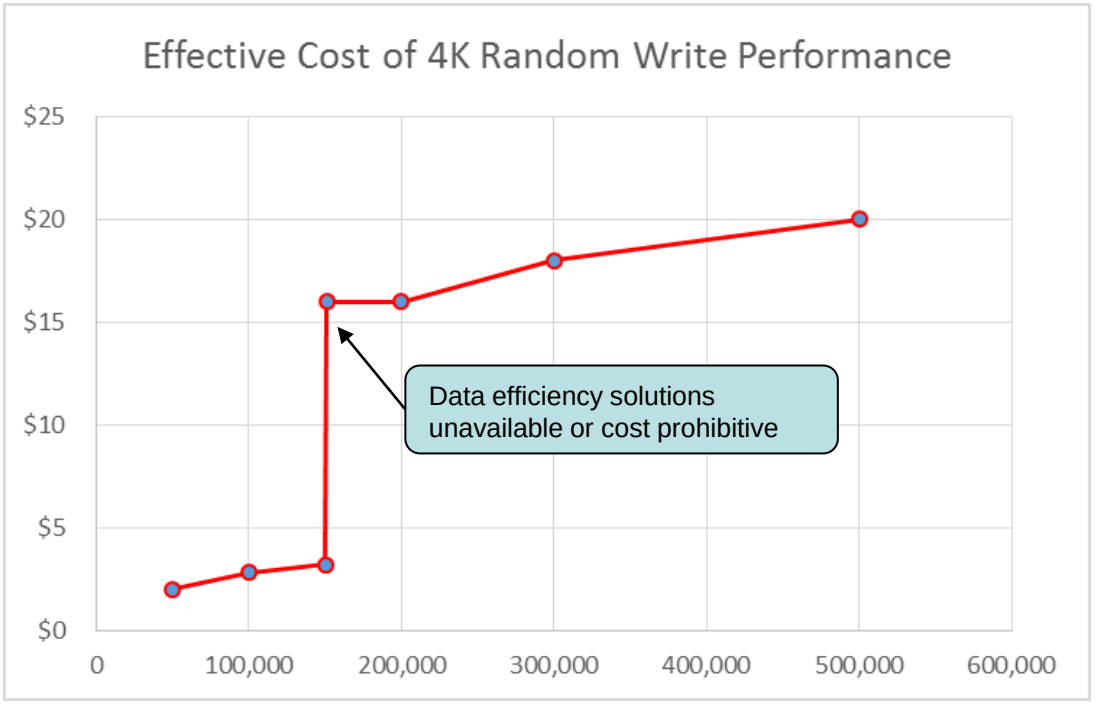
- Use the libaio engine (Linux AIO) and direct (non-buffered) IO
- Measure sequential and random reads/writes/mixed workloads
- Analyze workloads with a range of queue depths from 1 to 1024 and run the tool with varying numbers of simultaneous jobs
- Test IOPS across multiple IO request sizes from 4K to 128K

Flexible IO Tester (FIO) Dedupe/Compression Performance

- Start with random data (uncompressible, non-dedupeable)
- Fill the storage with data before commencing tests
- Run over long durations to develop the best understanding of how performance changes over time

The Price Performance High Jump

The jump in cost between midrange performance (w/ 5:1 mainstream data reduction) and high-end performance (w/out) is substantial!

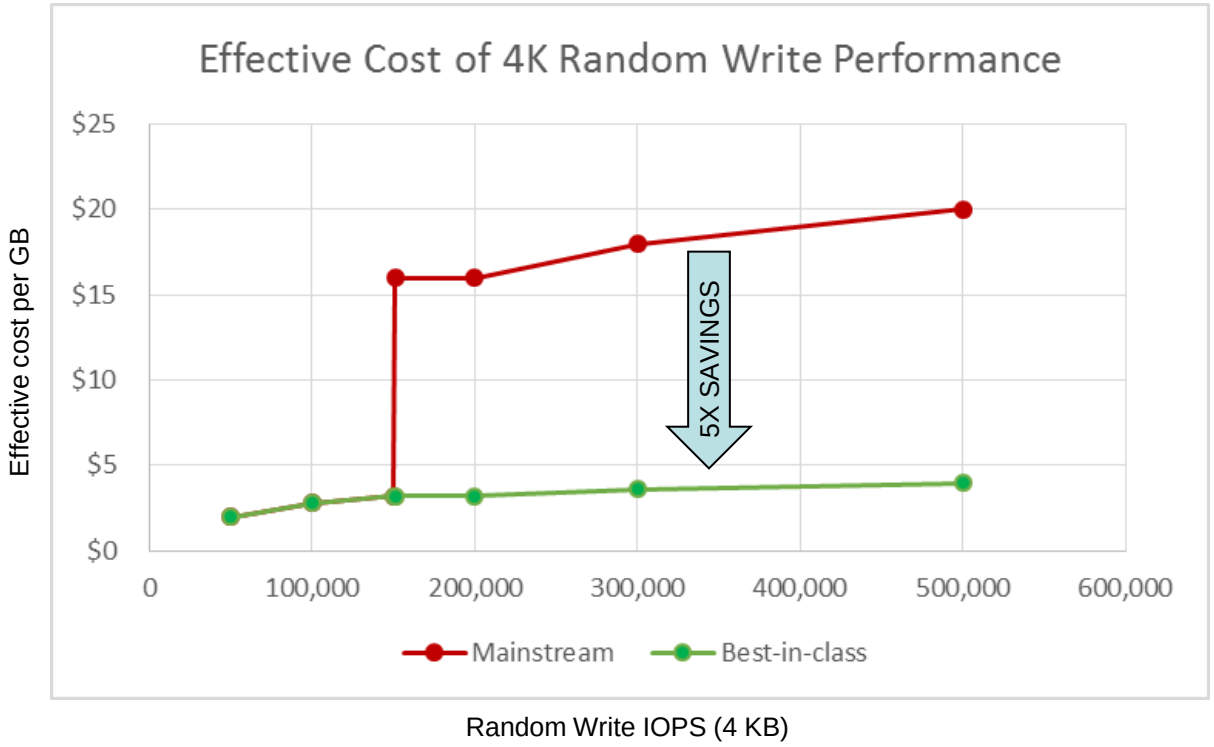


Eliminate the price jump for high performance

- Throw hardware at the problem; or
- Modify software (or license technology) to make it more efficient

Metric	Typical 2014	Best-in-Class SW	Improvement
Performance: 4K random write IOPS per server	150,000	500,000	3X
DRAM overhead: GB of RAM / TB of storage	4	0.3	13X

Best-in-class software dramatically reduces cost of performance



- Todays mainstream data efficiency doesn't meet the performance requirements for high-end flash performance
- Best-in-class software eliminates the cost performance penalty