A State Machine Architecture for High-Efficiency, Low-Latency SS Media Performance Extraction

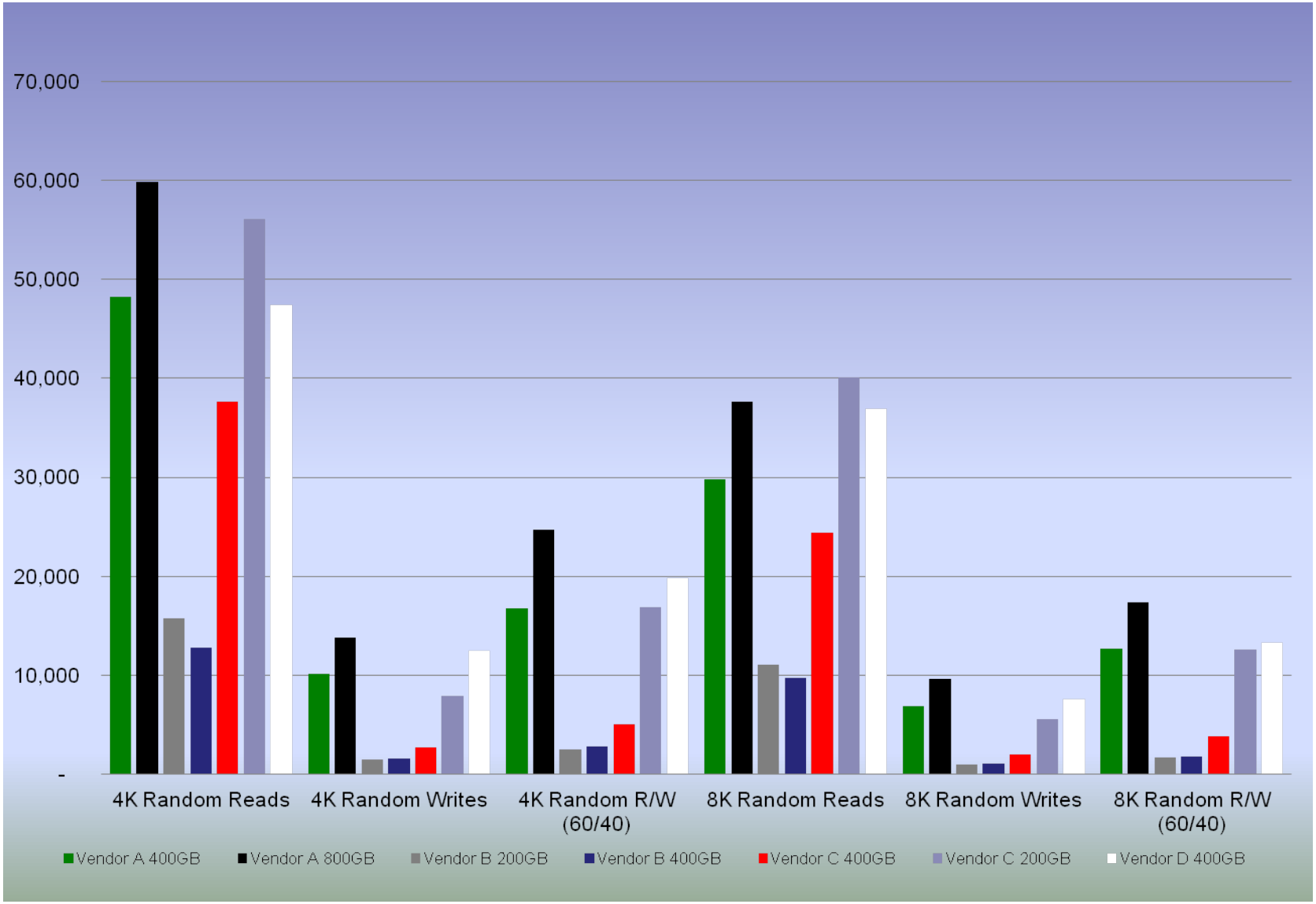
Bret S. Weber

DataDirect Networks

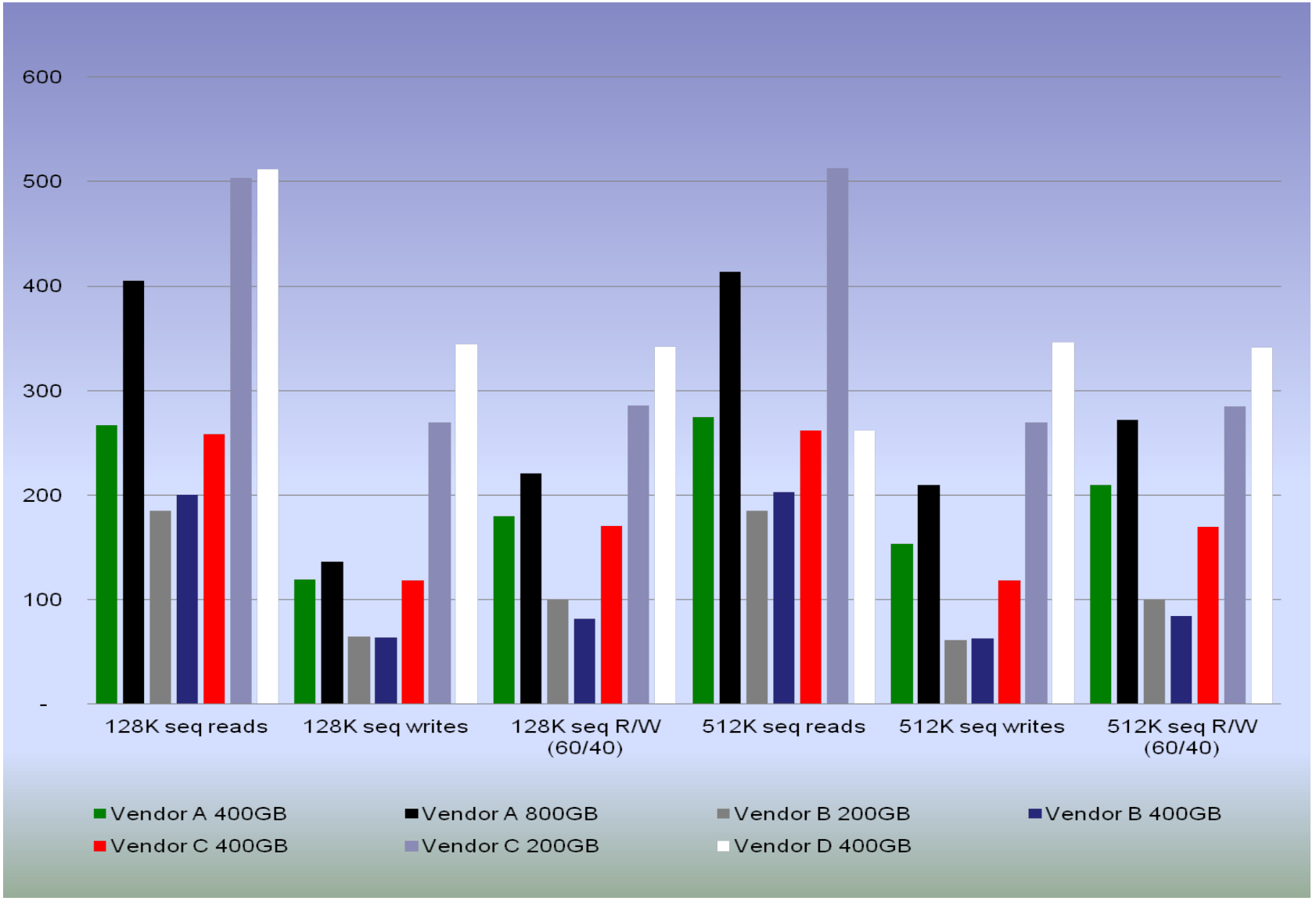
The Real Issue

- Want SSD Performance & Latency
- Need Enterprise Redundancy
- Use Commodity Flash Products
  - Non Proprietary
  - Customer Replaceable
  - No Technology “Lock In”
  - Lowest $/TB
- Allow Seamless Levels of Storage
  - Low $/TB
  - Low $/IOP

SSD Small Block IOPs



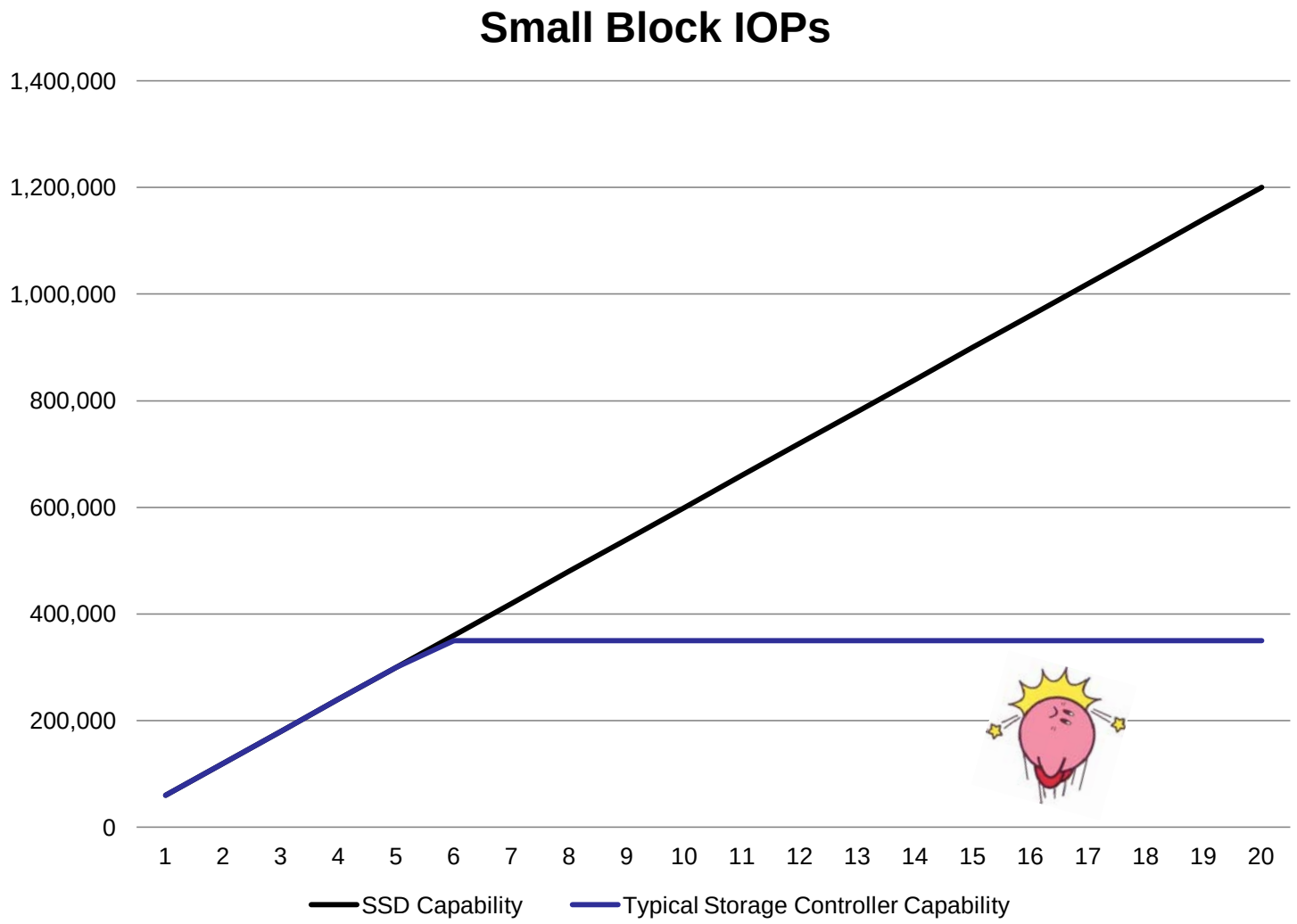
SSD Large Sequential



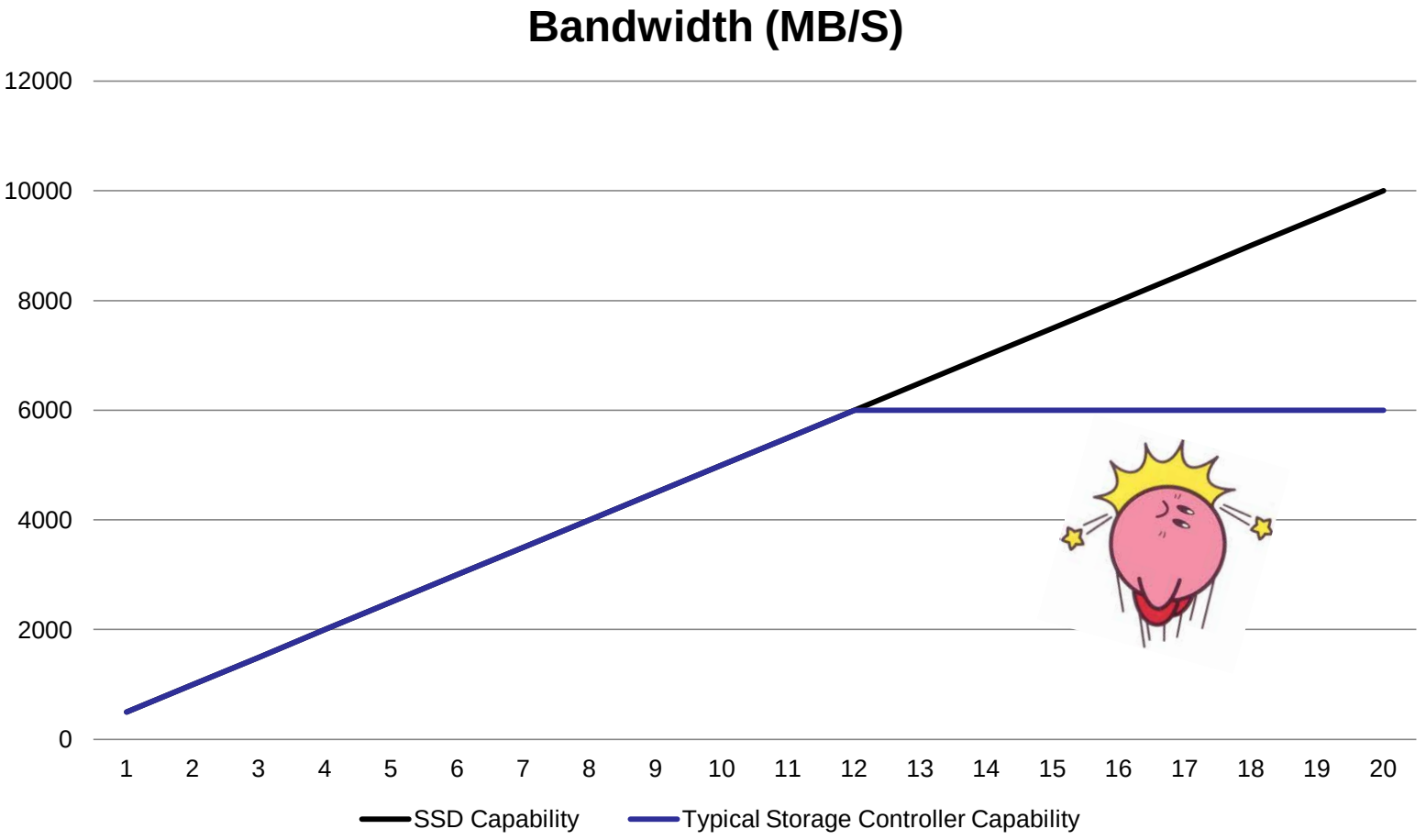
Flash Performance Takeaways

- Flash Performance is All over the Map
- The Technology Changes Fast
- Tradeoffs between Cost, Performance and Life
- Investment Protection

Typical Controller IOPs Throttling



Typical Controller BW Throttling



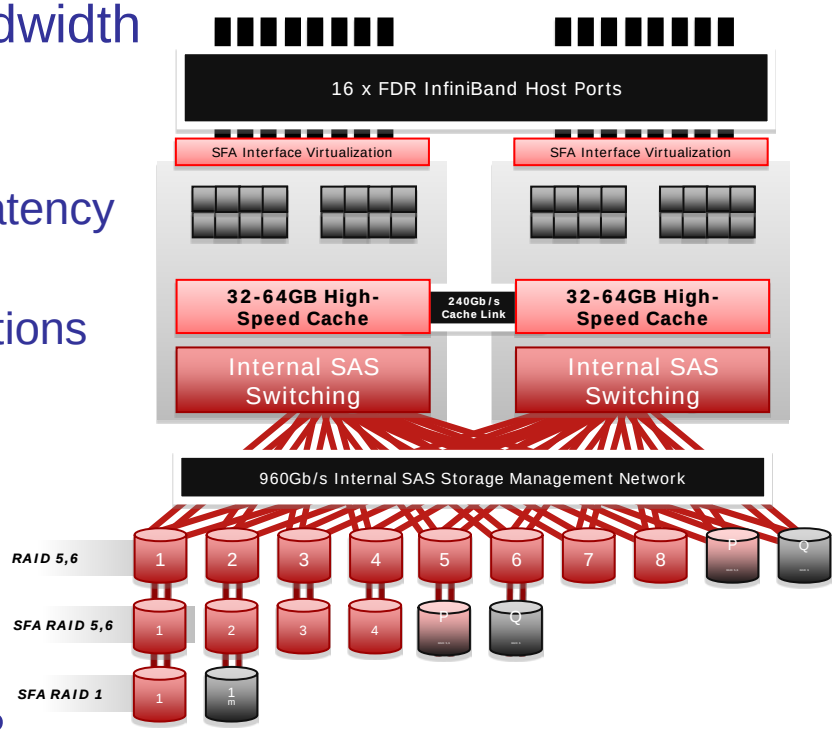
Storage Fusion Architecture History

- Previous Experience with Silicon-Based RAID-3 Arrays
  - High Bandwidth
  - Very consistent Quality of Service (Bandwidth and Latency)
- Ground Up “Blank Sheet of Paper” Architecture in 2007
  - Typical Five Year Maturity Cycle
- Address Emerging Applications
  - HPC
  - Data-Intensive Cloud Applications
  - Big Data Analytics
- Address the Emerging Technologies
  - Multi-Core Processors
  - High IOP Non Volatile Memory Technologies
  - Server Virtualization
  - Scale Out

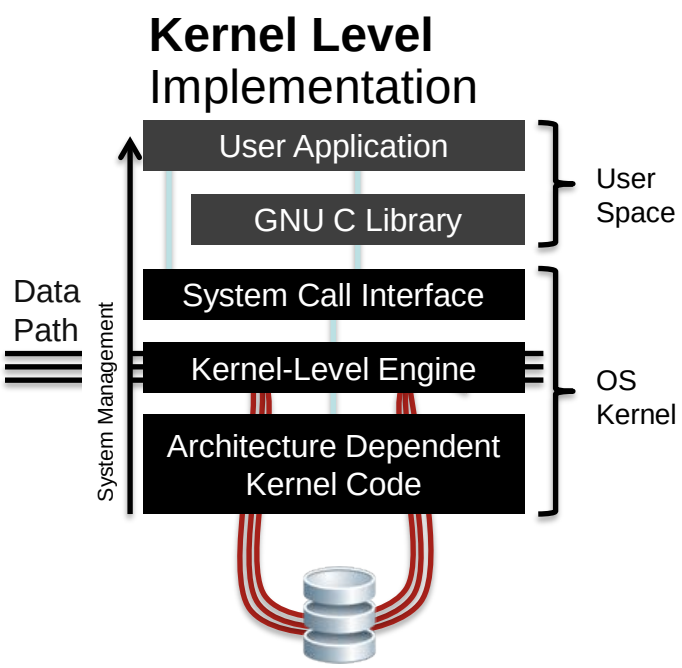
Ground-Up Design
Big Data Optimized

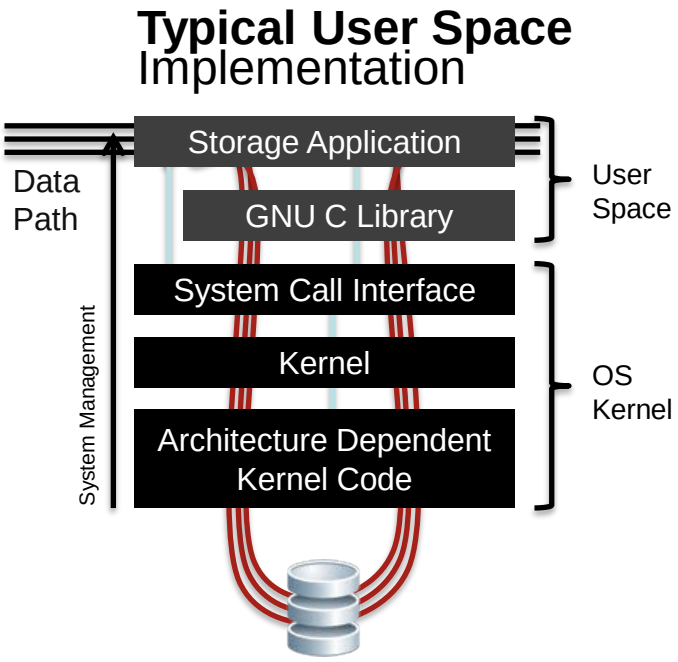
SFA – Details

- Ground Up Design Exploiting *Emerging* Technologies
- Optimized for Ultra-High IOPs & Bandwidth
  - Application Space Code
  - Kernel Bypass for all IO Operations
  - Real-Time NUMA Scheduler Minimizes Latency
  - Built to Exploit All Available CPU Cores
  - BW, IOPs and Mixed Workload Configurations
- Fully Parallel I/O Execution Engine
  - No Lock Architecture
- Portable Code
  - Rapid Time To Market
  - Optimizations around Linux architectures
- Highest levels of Performance & QoS
- DDN device drivers accelerate I/O
- Optional Server Virtualization Brings Big Data Closest to Processing



Typical Approaches

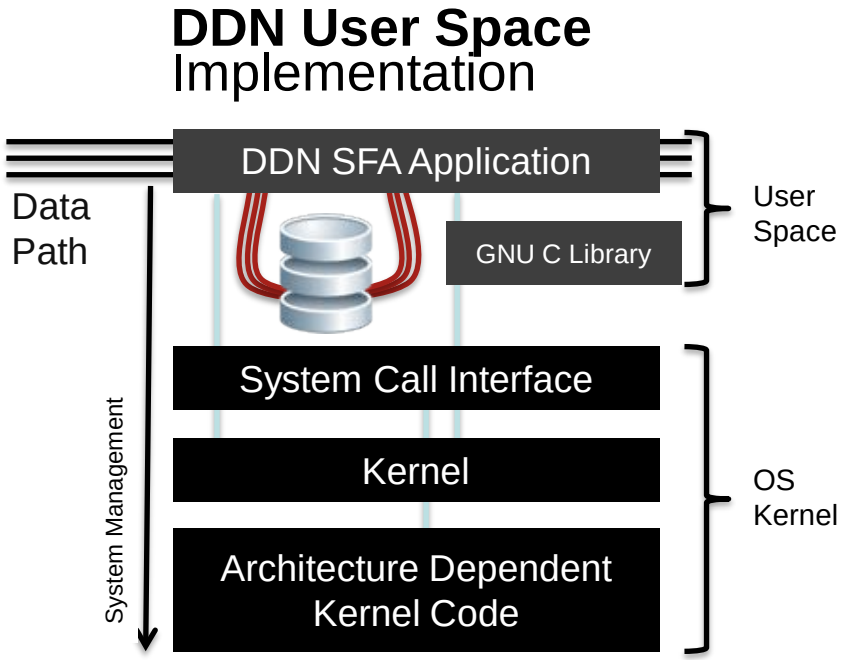


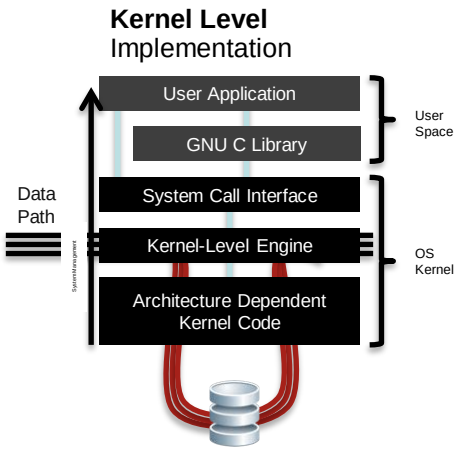


In-Kernel Dependencies Prohibit Easy Portability

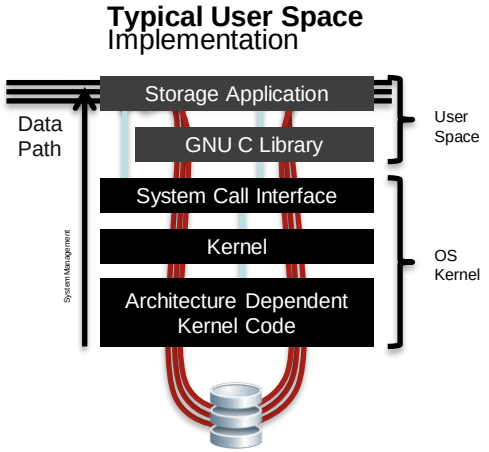
Kernel Context Switches Add Significant Latency

Low Latency – No Context Switching

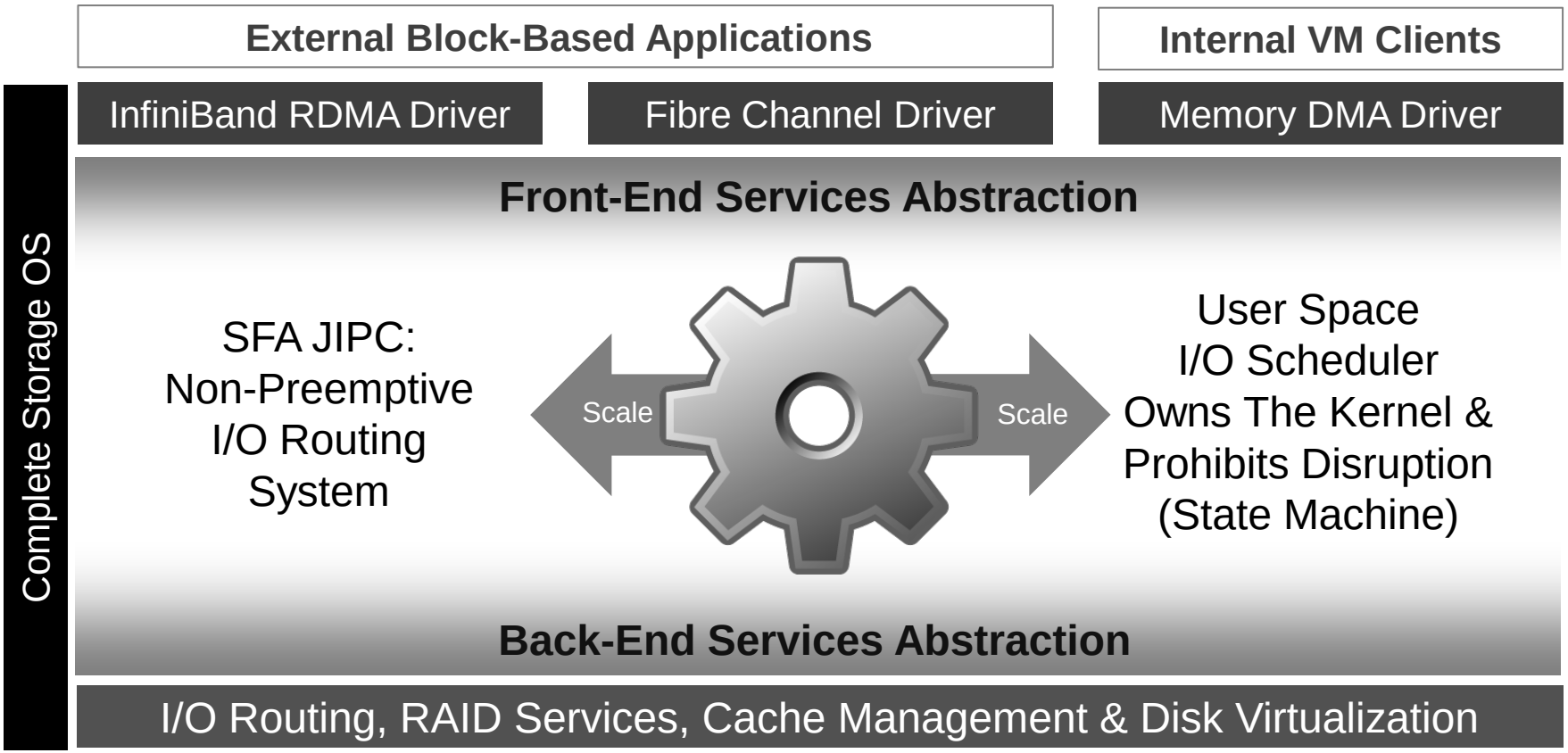




- No Context Switching
- Low Latency IO
- Predictable Performance
- High Speed Routing Architecture

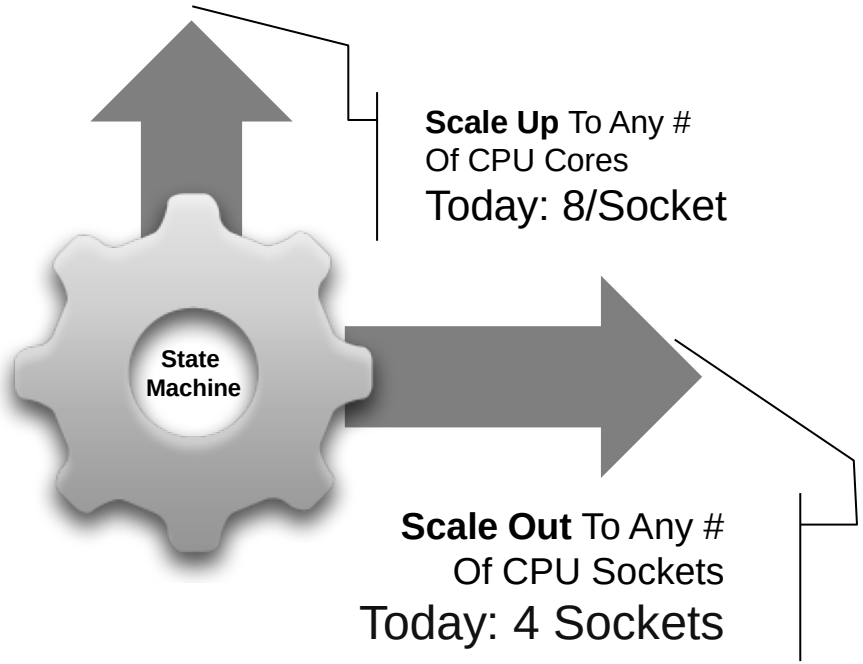


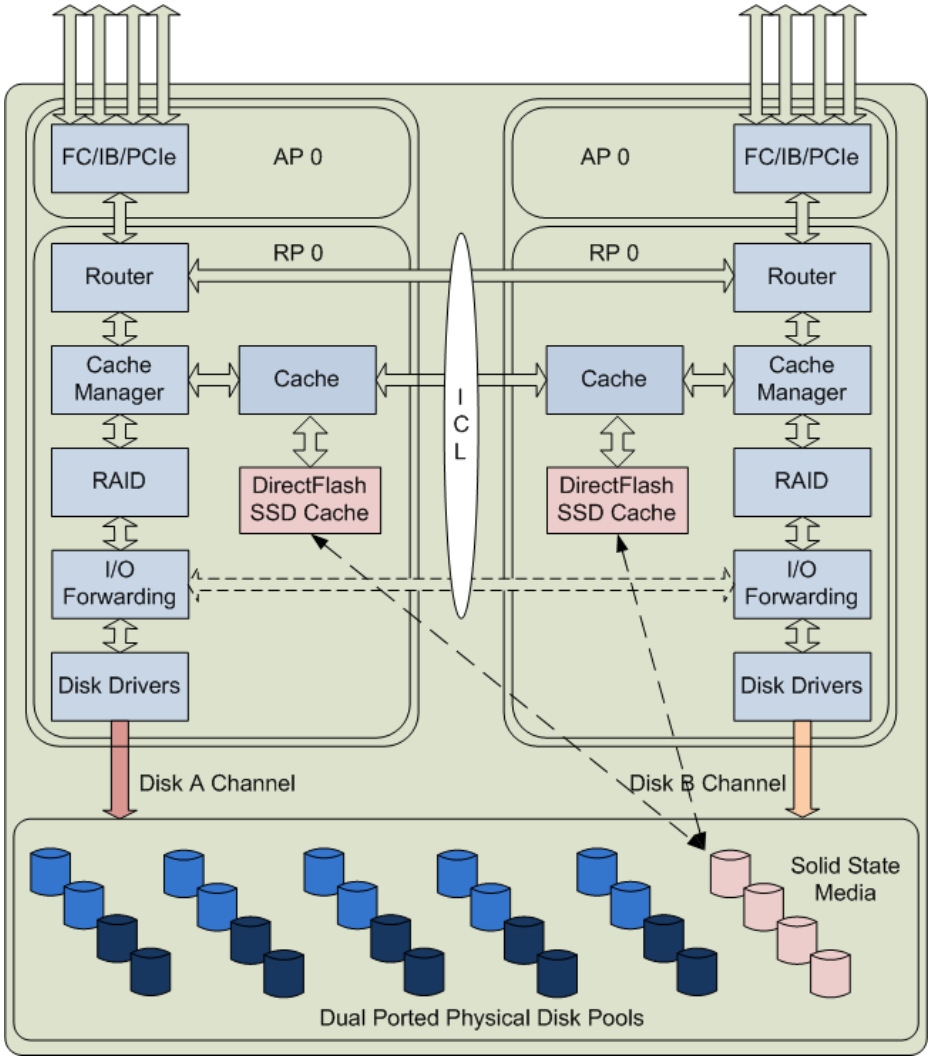
A Scalable & Abstractable I/O Delivery Architecture



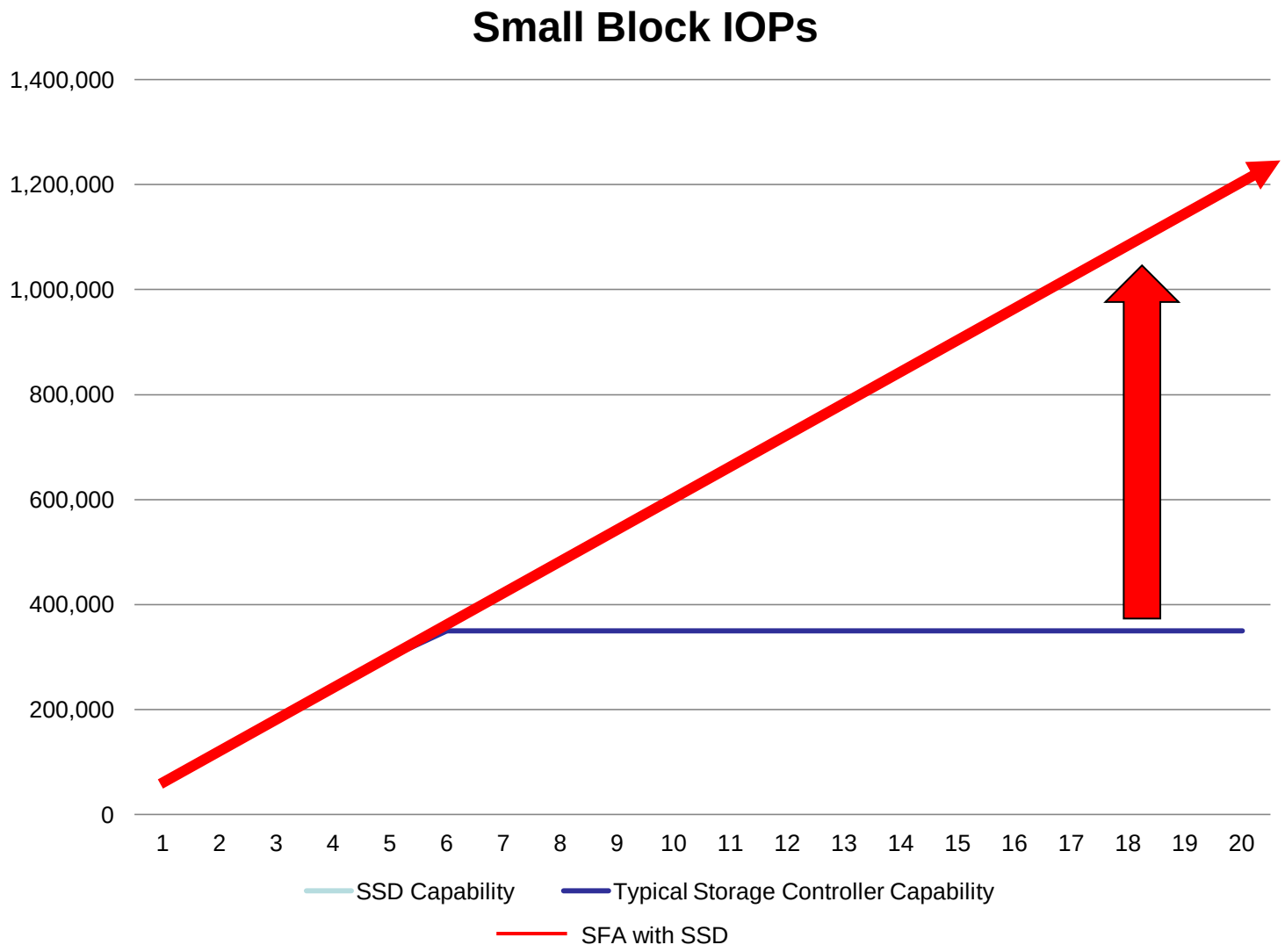
SFA12K Architecture

- 40GB/S Throughput
- 1.4M IOPs to SSD
- DirectFlash Caching Acceleration

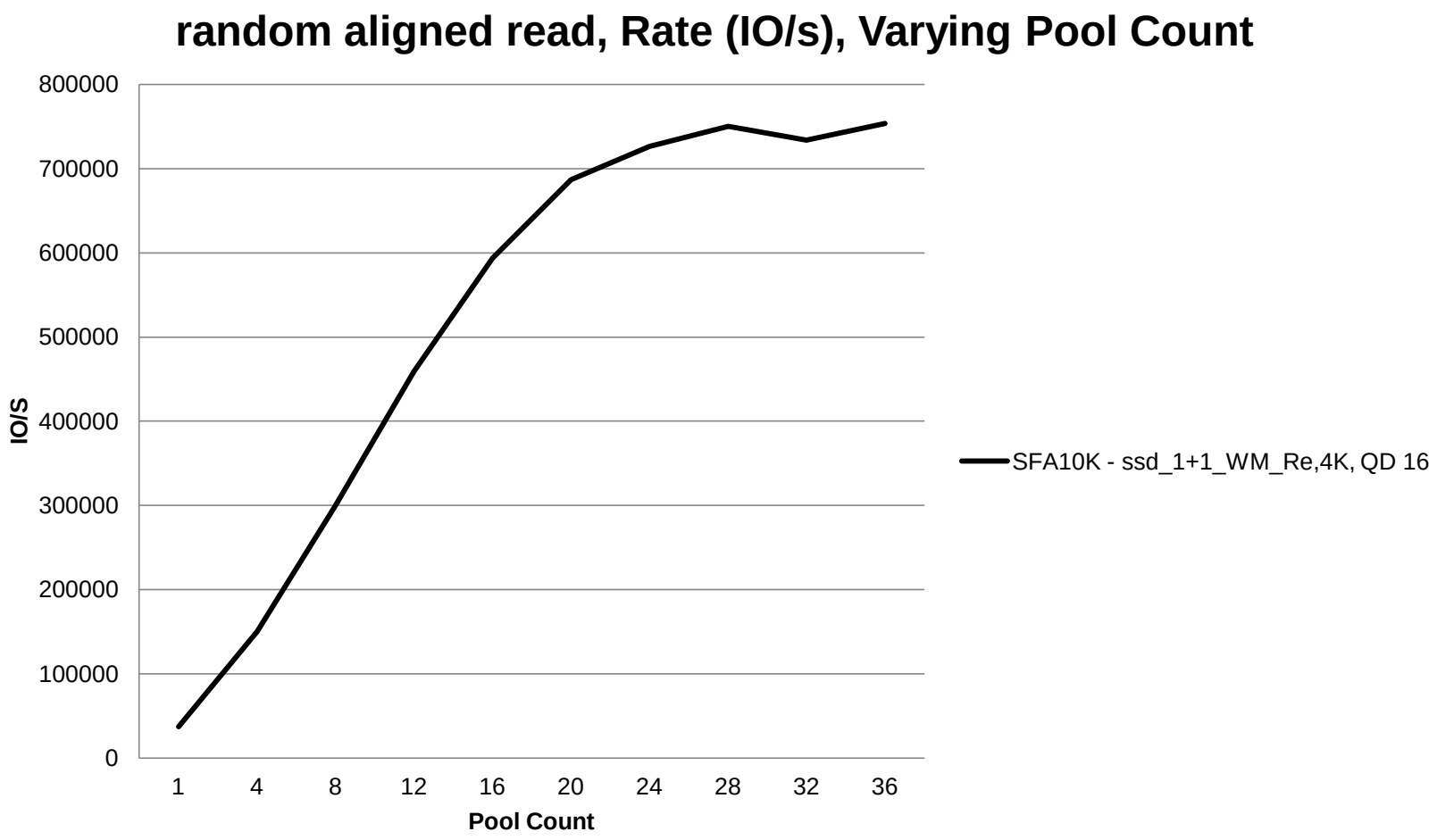




SFA Controller IOPs – No Throttling

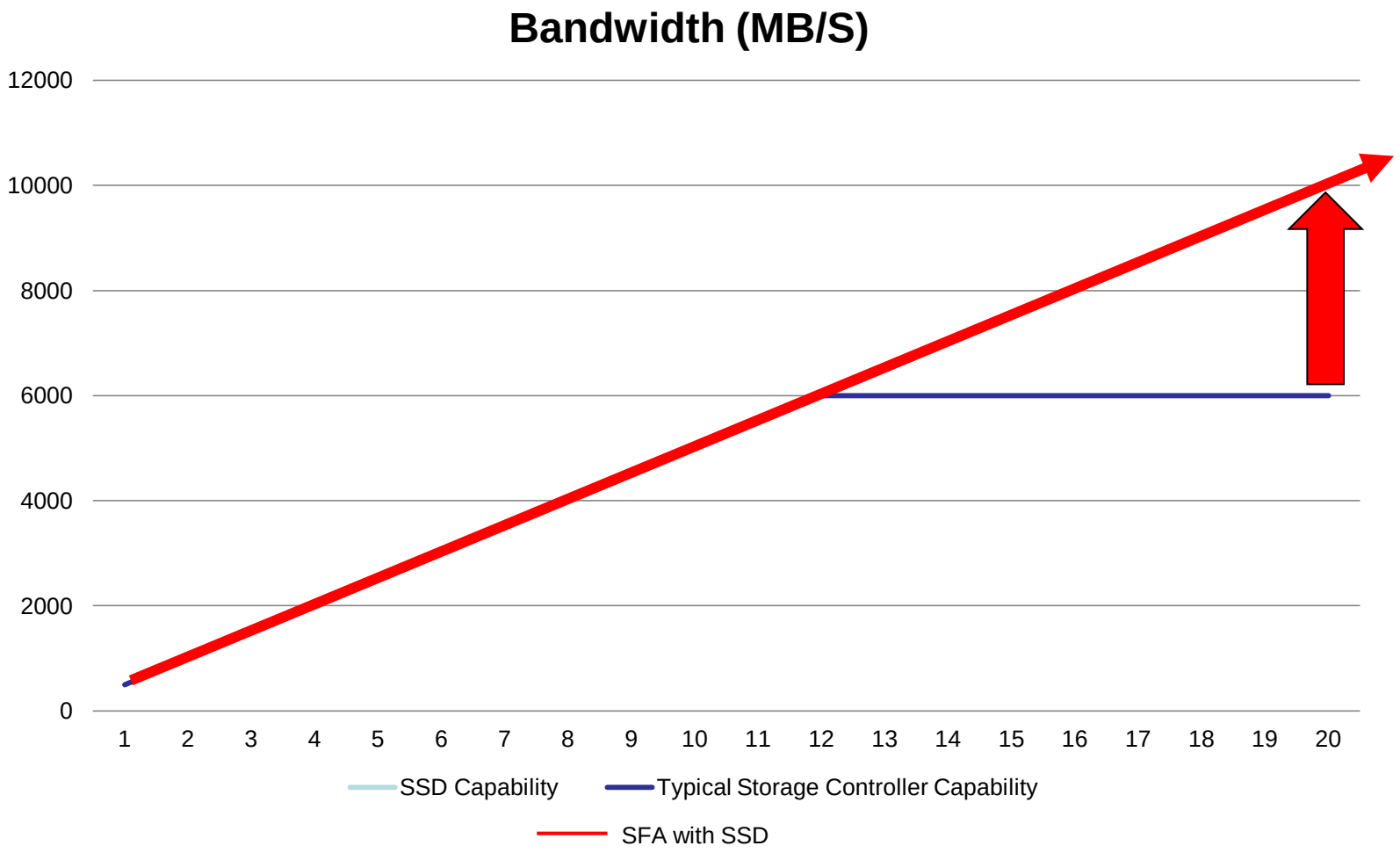


SFA Linear Scaling (SFA10K)

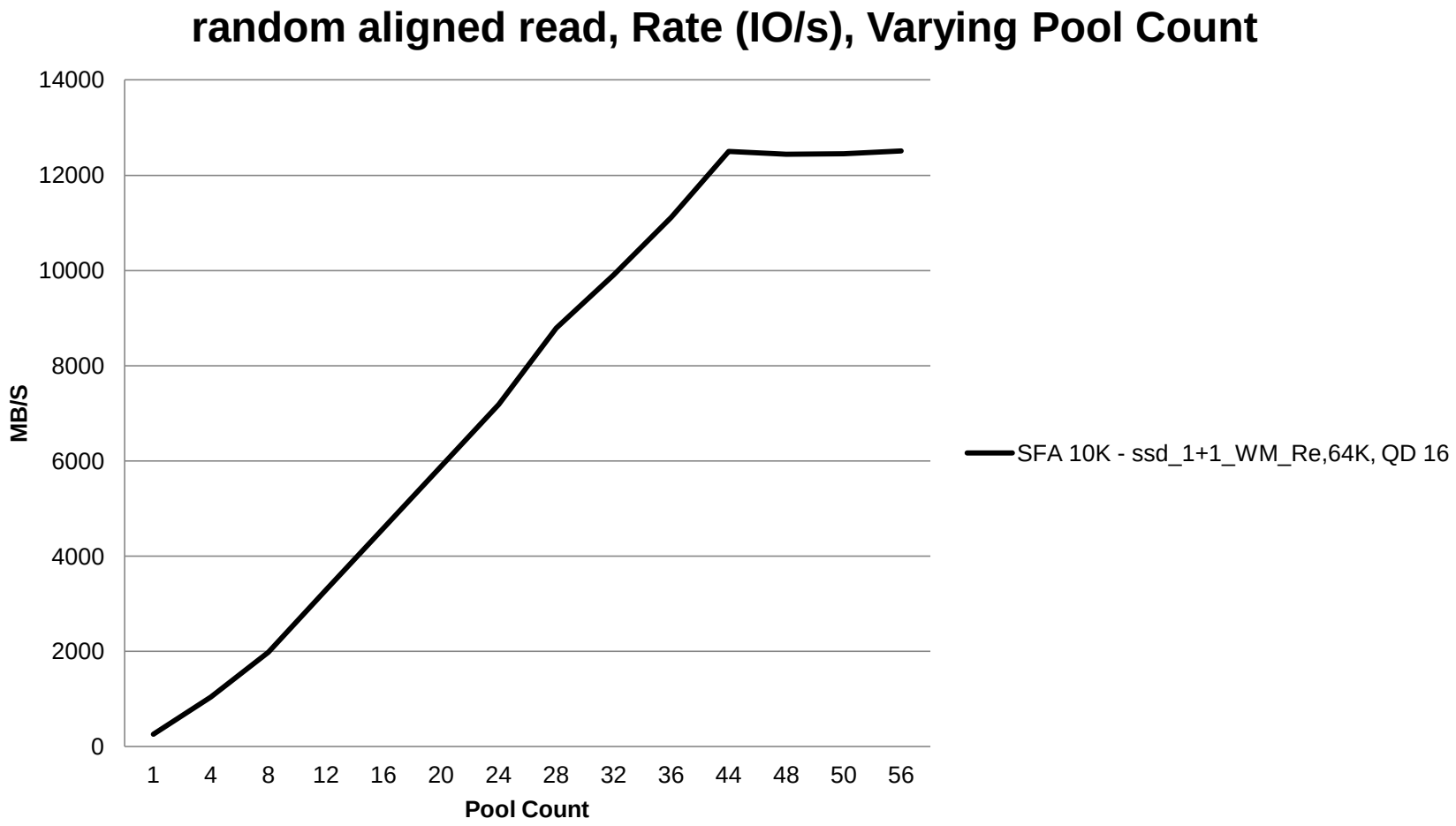


Previous Generation Product

SFA Controller – No BW Throttling



SFA Linear Scaling (SFA10K)



Previous Generation Product

Our Surprise: 4KB I/O (SFA10K) Latency That Defies Convention

Random Write

	1 x 8+2: Pliant 800GB	1 x 4+1: Pliant 800GB
QD	Avg Latency (µs)	Avg Latency (µs)
1	75.87	79.15
2	62.77	69.78
4	72.58	76.89
8	78.21	84.97
16	86.80	90.98
32	95.67	96.79
64	102.17	107.89
128	99.88	105.70

Random Read

	1 x 8+2: Pliant 800GB	1 x 4+1: Pliant 800GB
QD	Avg Latency (µs)	Avg Latency (µs)
1	324.78	323.02
2	156.92	156.07
4	72.91	73.11
8	37.86	38.15
16	19.97	20.34
32	10.38	11.16
64	6.51	7.58
128	6.89	8.23

Actual Case Studies

Leading U.S. Cloud-Based Applications Provider

SFA10K-M Systems
QDR InfiniBand Connected
114 x 800GB SSDs

- Rich Imagery Processing Application
- GPFS File System
- Mix of Small & Large File Requests (mostly small)
- SSD System Bake-Off
- DDN Selected Due To:
  - Multi-Dimensional Performance
  - Low Latency & Wall Clock
  - Lowest Data Center Footprint

Conclusions

- Get Performance *and* Redundancy
- SSDs Can Compete With The Right Architecture
- Lots Of Options: Don’t Lock Yourself In
- Don’t artificially limit your performance
- Big Data Requires All Varieties of Performance
  - We don’t know what we don’t know…

Questions ???