SKA Workshop
Feb 2015

Feb, 2015

DDN Australia

DDN | About Us

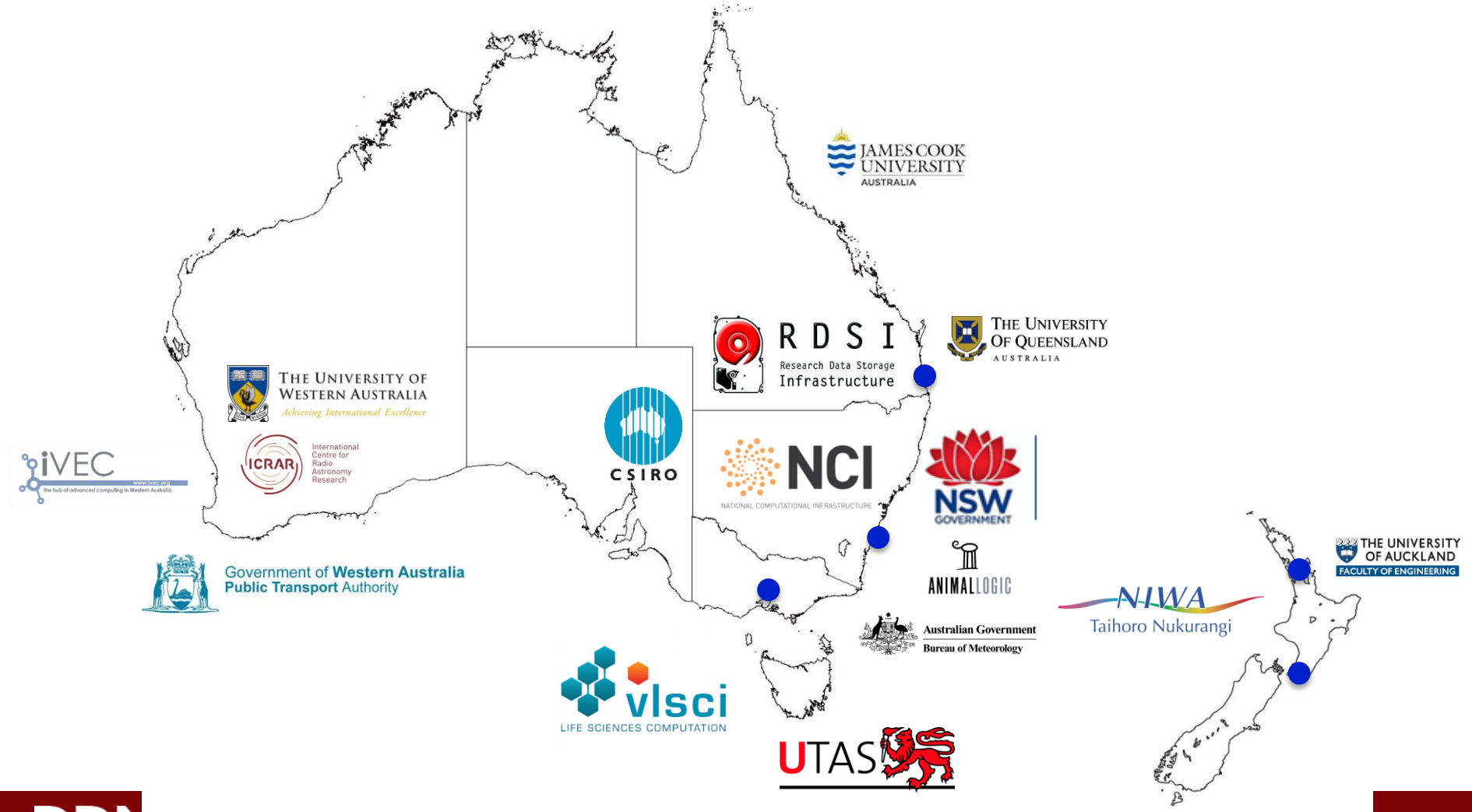
DDN is a Leader in Massively Scalable Platforms and Solutions for Big Data and Cloud Applications

- Established: 1998
- Revenue: $250M+ – Profitable, Fast Growth
- Main Office: Sunnyvale, California, USA
- Worldwide Presence: 20 Countries
- Installed Base: 1,000+ End Customers; 50+ Countries
- Go To Market: Global Partners, Resellers, Direct

World-Renowned & Award-Winning

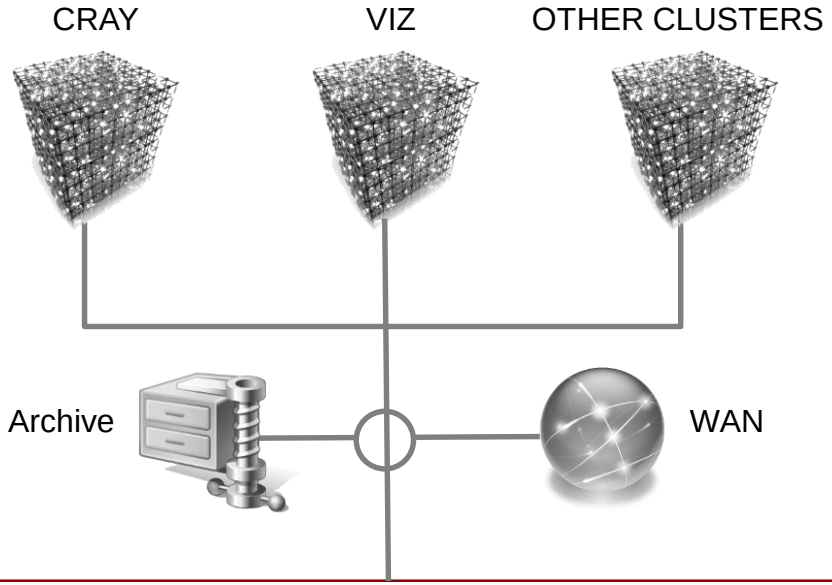
IDC Analyze the Future · Inc. · Gartner · the 451 group · HPCwire · STORAGE · Federal Computer Week

Sample Customers Australia & New Zealand



Oak Ridge National Laboratory

Case Study: Building The World's Fastest File System



ORNL Selected DDN SFA12K Technology To Power The World's Fastest Storage

DDN was selected because:

- Sustained Quality of Service @ Scale
- Best Price/Performance
- Leadership-Class Data Center Density
- Open-Platform For Parallel File I/O
- Deep Expertise in Scaling File Storage

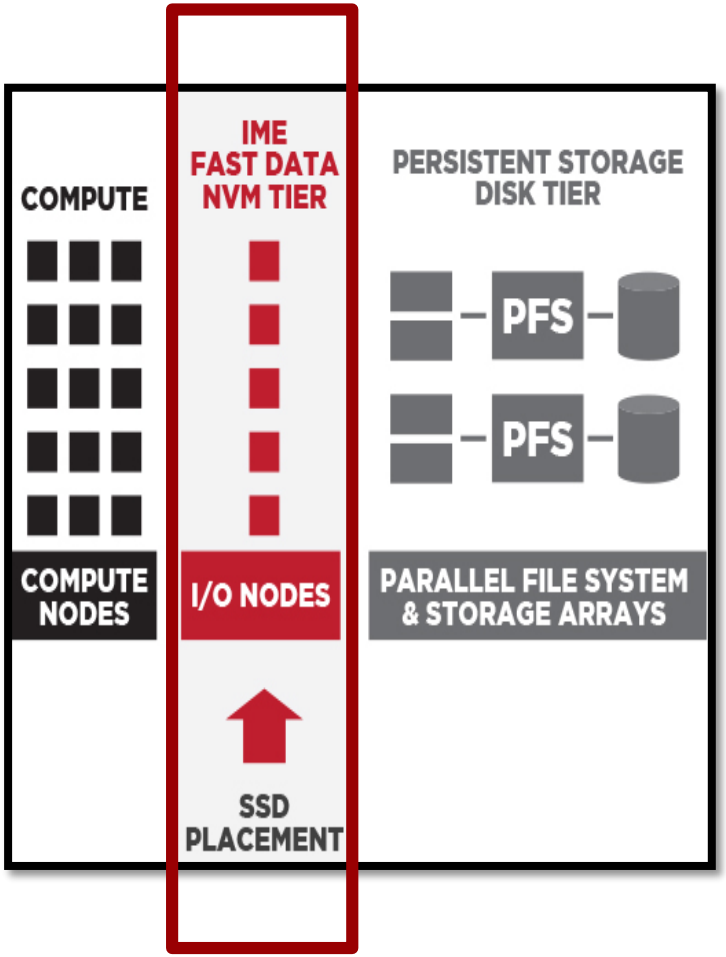
36 x DDN SFA12K-40

File System Performance: 1TB/s+
Capacity: 40.3PB (raw)
File System: Lustre®
I/O Platform: 36 x DDN SFA12K-40
Media: 20,160 HDDs

What is IME? A Tier of Non-volatile Memory

Residing Between Compute and Persistent Storage

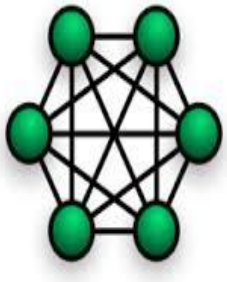
IME creates a new application-aware fast data tier that resides right between compute and the parallel file system to accelerate I/O, reduce latency and provide greater operational and economic efficiency



DDN IME Ecosystem – Client IO Interfaces

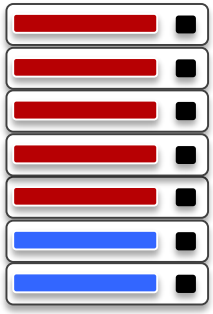
- **Three primary interfaces for IME**
  - IME FUSE
    - Provides POSIX IO
    - Captures IO requests through the Linux VFS
    - Target Use Case: General purpose applications that use POSIX
  - IME ROMIO
    - Provides MPI-IO support
    - Captures IO requests through the MPI runtime in user space
    - Target Use Case: Parallel applications
  - IME Native Library
    - Low-level programming interface
    - FUSE and ROMIO layers implemented on this interface
    - Target Use Case: Highly-optimized customer applications that may not map cleanly onto POSIX or MPI-IO

The IME Advantages



Designed for Scalability
Patented DDN Algorithms

Fully POSIX & HPC Compatible
No Application Modifications



Scale-Out Data Protection
Distributed Erasure Coding

Intelligent, Adaptive System
On-the-Fly Data Placement

Integrated With File Systems
Designed to Accelerate Lustre*, GPFS
No Code Modification Needed

Writes Fast; Read Fast Too
No other system offers both at scale.

The IME Advantages

1000X Application Acceleration
Run More Complex Simulations Faster With Less Hardware

50% Less Latency Than All Flash Arrays
Optimizing Workload Performance to reduce time to insight and discovery

Scales Memory to 100s of TB
To Move Large Datasets Out of storage & into memory extremely fast, without storage latency

80% Lower Cost
Infinite Scalability With the Highest Efficiency To provision I/O Performance with the Highest Efficiency

Early Access Testbeds Deployed Globally

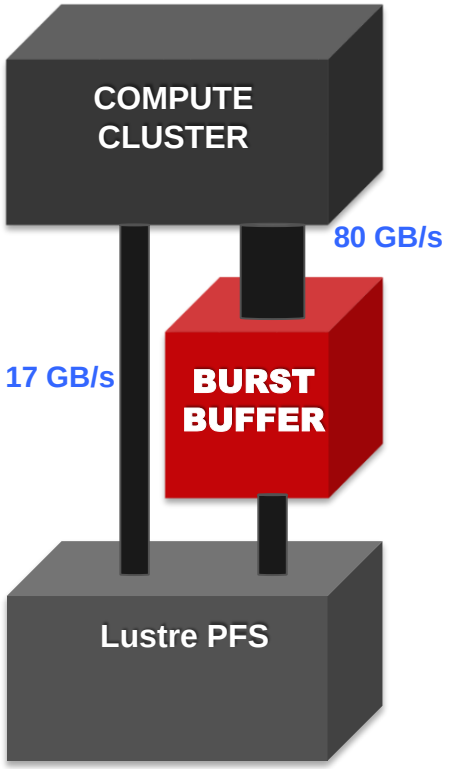
At customer sites and regional benchmark centers since June

- DDN Maryland
- DDN Japan
- DDN Germany
- DDN Paris
- TACC
- NERSC
- ICHEC
- DESY
- FJZ Juelich
- CSCS / EPFL
- NCSA

HACC_IO @ TACC (from Hardware/Hybrid Accelerated Cosmology Code)

Cosmology Kernel

Particles per Process	Num. Clients	IME Writes (GB/s)	IME Reads (GB/s)	PFS Writes (GB/s)	PFS Reads (GB/s)
34M	128	62.8	63.7	2.2	9.8
34M	256	68.9	71.2	4.6	6.5
34M	512	73.2	71.4	9.1	7.5
34M	1024	63.2	70.8	17.3	8.2
IME Acceleration		**3.7x-28x**	**6.5x-11x**		

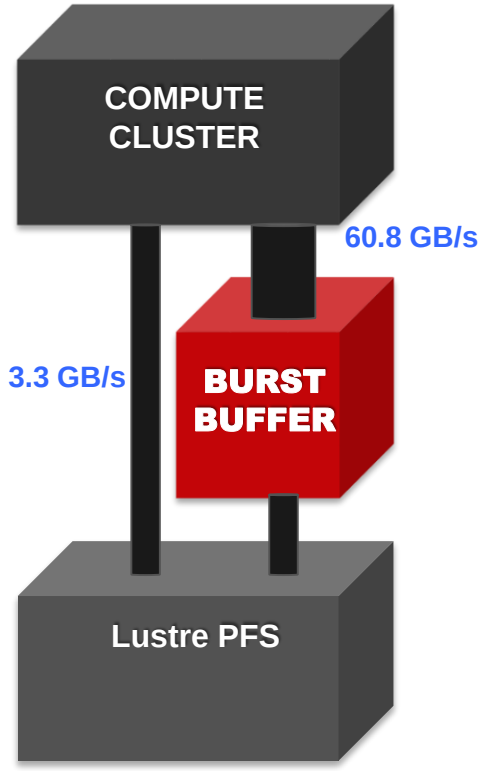


HACC_IO Cosmology

S3D @ TACC

Turbulent Combustion Kernel

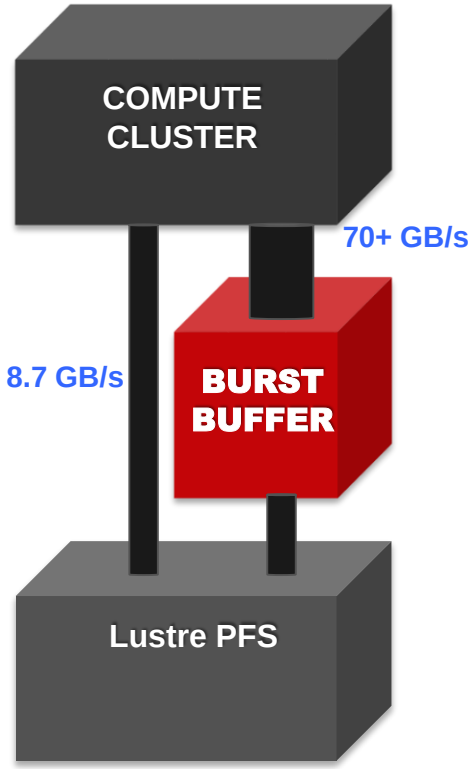
Processes	X	Y	Z	IME Write (GB/s)	PFS Write (GB/s)	Acceleration
16	1024	1024	128	8.2	1.2	**6.8x**
32	1024	2048	128	14.0	1.5	**9.3x**
64	1024	4096	128	22.3	1.5	**14.9x**
128	1024	8192	128	31.8	3.0	**10.6x**
256	1024	16384	128	44.7	2.6	**17.2x**
512	1024	32768	128	53.5	2.4	**22.3x**
1024	1024	65536	128	60.8	3.3	**18.4x**



S3D Turbulent Combustion

MADBench @ TACC

Phase	IME Read (GB/s)	IME Write (GB/s)	PFS Read (GB/s)	PFS Write (GB/s)
S		71.9		7.1
W	74.6	75.5	7.8	8.7
C	74.7		11.9	
IME Accel.	**6.2x-9.6x**	**8.7x-10.1x**		



Application Configuration: NP = 3136, #Bins=8, #pix = 265K

IME Test Nodes (Minimum of 4 nodes)

- 2 E5-2650v2 8 cores CPUs with HT enabled
- 128 GB RAM (8 x 16GB DDR3-1866 ECC REG)
- 1 dual port InfiniBand FDR HCA, OFED 2.2, IPoIB configured
- Centos 6.5, kernel 2.6.32-431.23.3
- THP enabled
- 24 240GB SSD drives
- 2 SAS2308 PCI-Express Fusion-MPT SAS-2

Approx 10GB/sec per node

IME Product Offerings

Ideally Suited for Commercial Customers, DIY Customers & DDN OEMs

Software Only

Your Compute Nodes

DDN IME Client Software

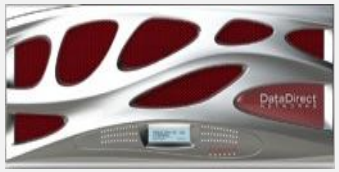
Your I/O Server

DDN IME Server Software

DDN Appliance

Your Compute Nodes

DDN IME Client Software



DDN IME I/O Server Appliance

DDN IME Server Software

Example: PFS vs. IME+PFS

More peak bandwidth, same persistent capacity, lower cost and HIGHER VALUE

PFS Only

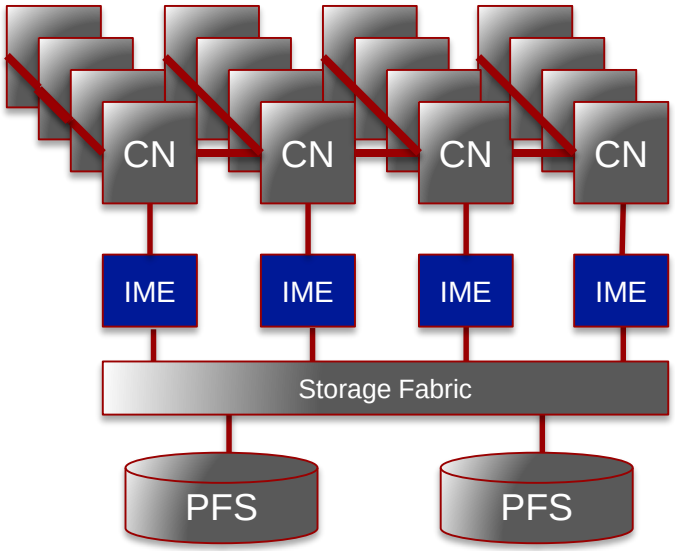
Cluster Memory: 400 TB

Cluster I/O BW: 540 GB/s

Storage Fabric: 540 GB/s

#OSS: 112
#SFA: 14
#HDD per SFA (5*80)= 400

#HDD Total: 5,600



IME + PFS

Cluster Memory: 400 TB

Cluster I/O BW: 756 GB/s

Storage Fabric: 270 GB/s

#OSS: 56
#SFA: 7
#HDD per SFA (10*80)= 800

#HDD Total: 5,600

IME Value Proposition

- ✓ **40% more bandwidth to the cluster**
- ✓ Faster job turn-around, more jobs in same period, fewer nodes needed to complete same amount of work
- ✓ **Fewer OSS and SFAs**
- ✓ **Reduced power, space and operational cost**
- ✓ **Similar persistent capacity**
- ✓ **Lower overall capital cost**

16

Common Capacity Configs

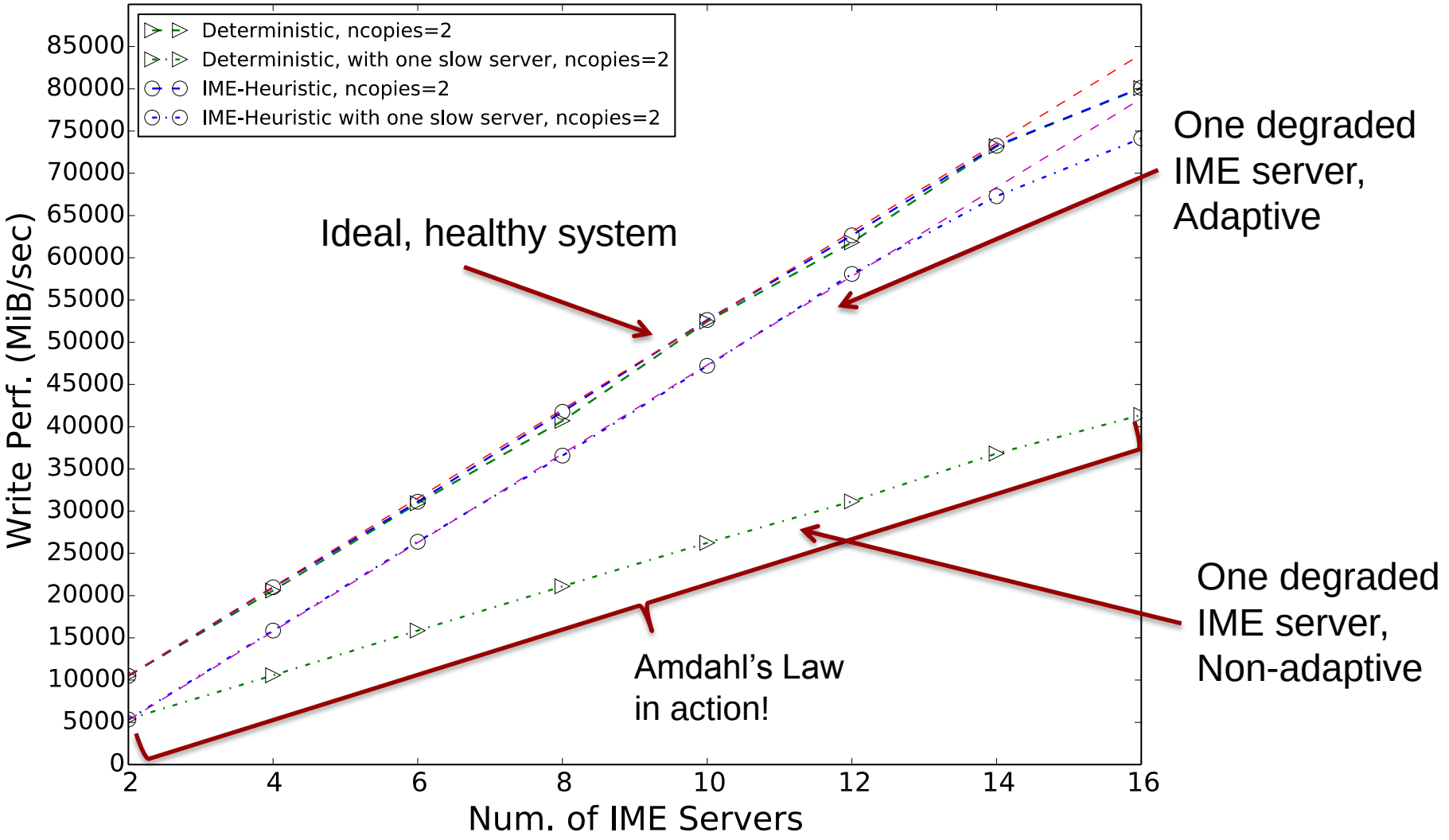
9 to 73 TB (usable) per 50 GB/s bandwidth. Guidance based on numerous RFP responses

- **Our basic IME Appliance is intended to provide >50 GB/s bandwidth, and configured with 24 to 48 NVMe SSDs. The NVMe SSDs are expected to have 480GB, 960GB, or 1.92TB raw capacity**
- **To account for data protection overheads, we assume an 0.86 usable capacity factor**
- **Basic IME Appliance configuration:**
  - Between 9 and 73 TB of usable capacity per 50 GB/s
  - Other capacities and bandwidths are possible, and when using 72 SAS SSDs per IME Appliance, the capacities can go higher than 150 TB per 50 GB/s

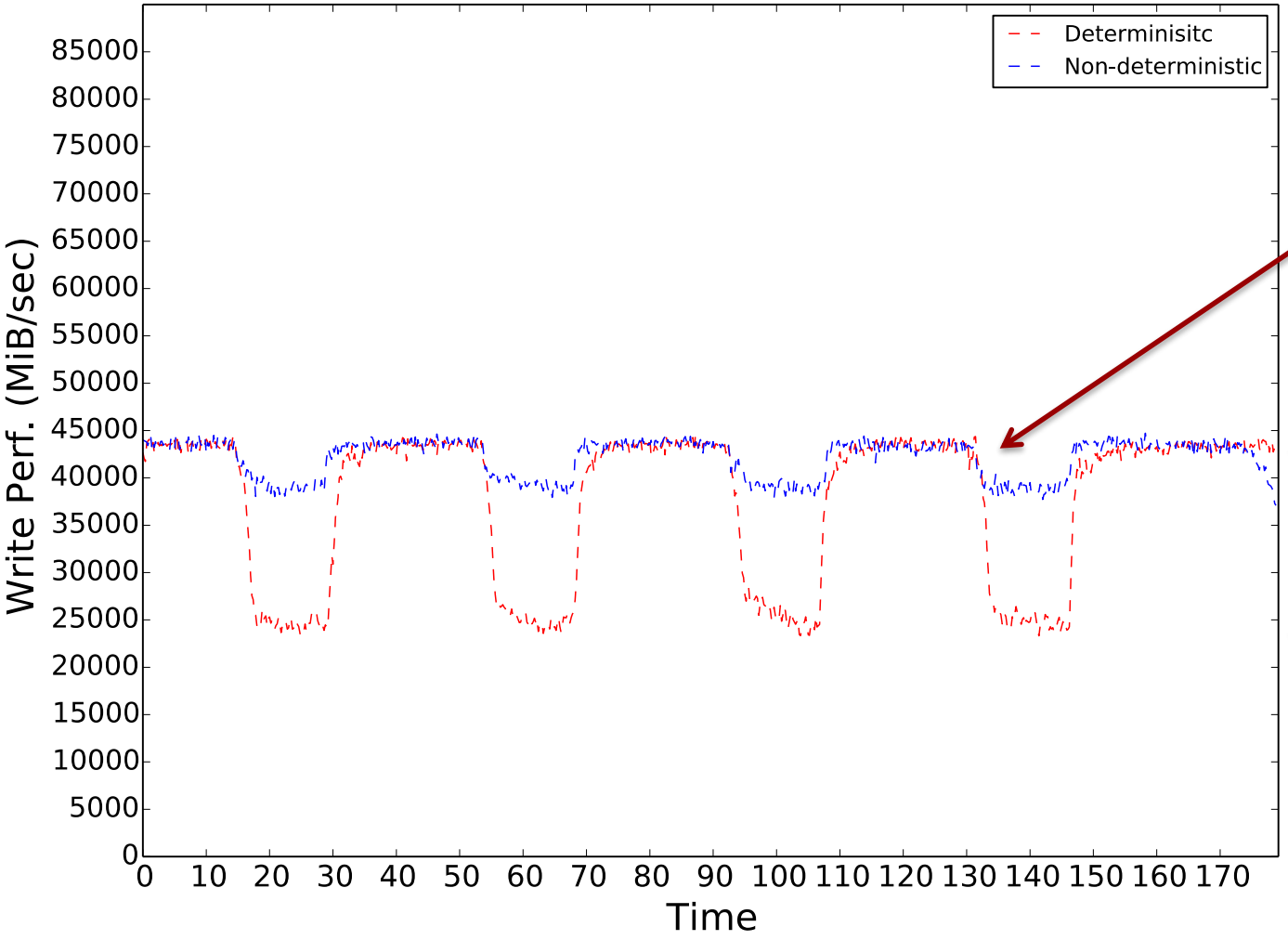
DataDirect™
NETWORKS

Thank you

Aggregate IME Adaptive vs. Non-Adaptive WRITE Performance



Real-Time IME Adaptive vs. Non-adaptive WRITE Performance



Adaptive heuristic learns "quickly"

4x Performance Lost with Non-adaptive