MLC Flash for Enterprise Storage Performance

Erik de la Iglesia

Chief Architect

GridIron Systems
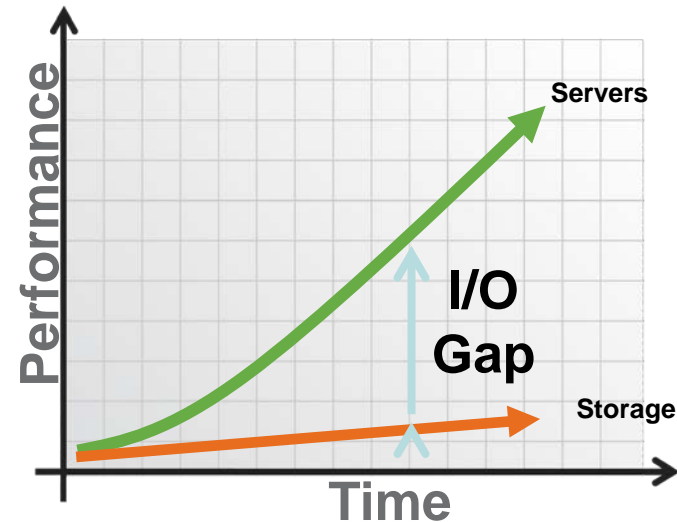
Flash Adoption in the Enterprise

- **I/O Gap Driving Adoption**
  - Accelerating business apps has high ROI
  - Cost per IOP is compelling (vs disk)
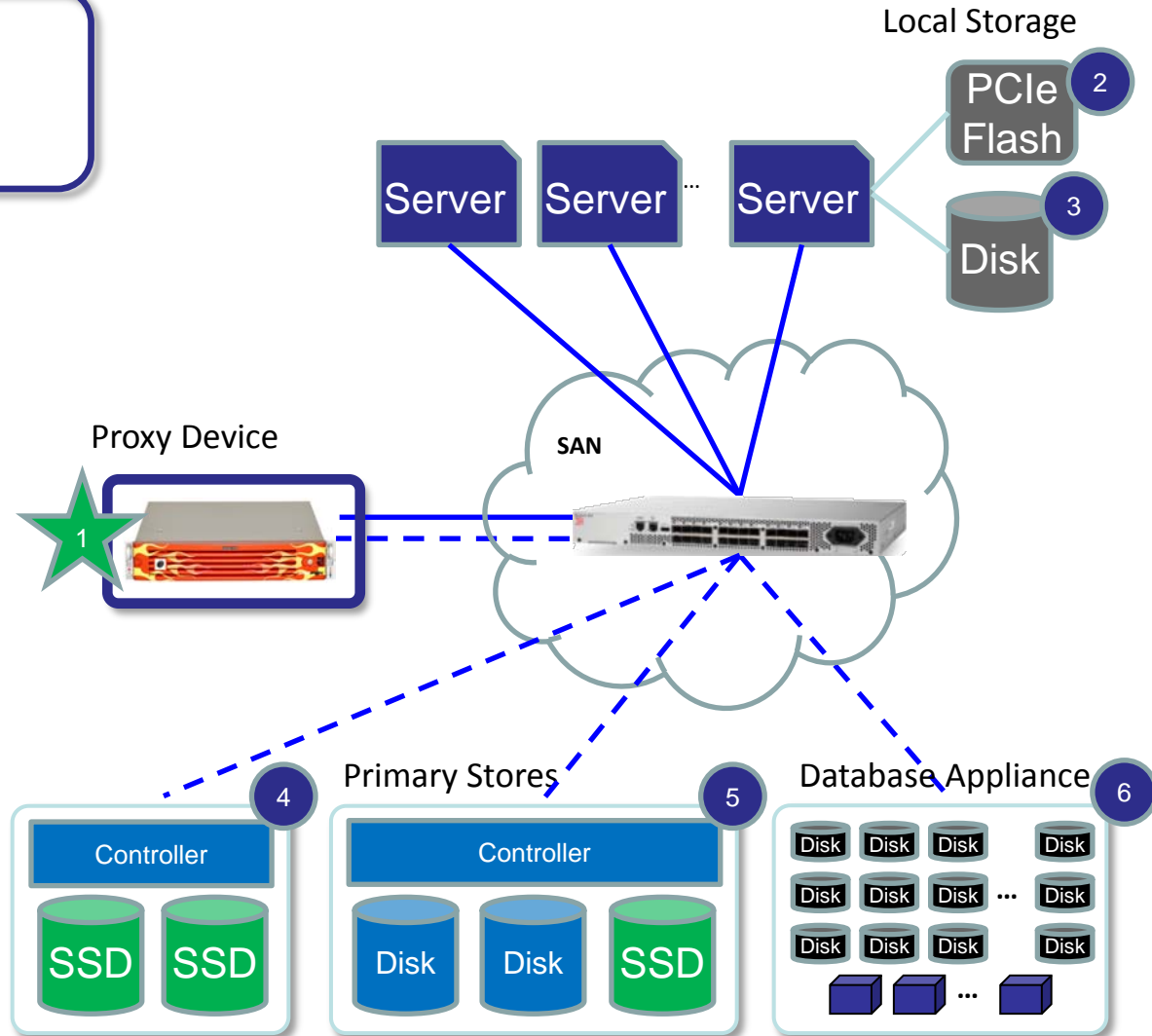 - Multi-core servers rapidly outclassing storage

- **Challenges abound**
  - Integration or forklift
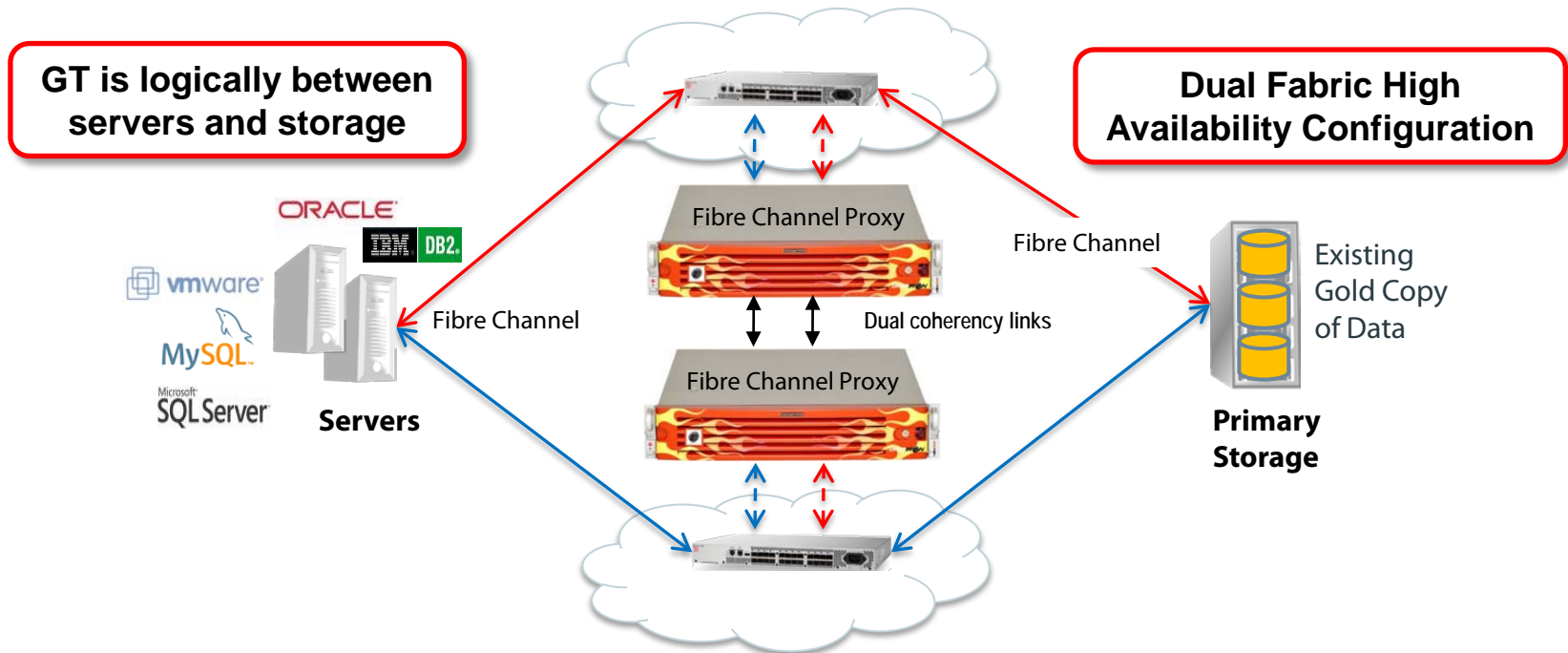 - Technology readiness
  - Cost effectiveness

Many Deployment Options

1. **SAN Proxy Device**
   Benefits all servers, storage and applications

2. **Flash PCIe Cards**
   Benefits only a single server, no virtualization

3. **Local Flash SSD**
   Data security, Degraded write performance over time

4. **Flash Primary Storage**
   Applies only to that storage, Wear issue for heavy writes

5. **Flash Disk Replacements**
   Controllers constrain SSD performance

6. **Specialized Processors**

Local Storage

PCIe Flash 2

Server Server ... Server 3 Disk

Proxy Device SAN

1

Primary Stores 4 5 Database Appliance 6

Controller Controller

Disk Disk Disk Disk
Disk Disk Disk ... Disk
Disk Disk Disk Disk

SSD SSD Disk Disk SSD

GT is logically between servers and storage

Dual Fabric High Availability Configuration

ORACLE

IBM DB2

vmware

MySQL

Microsoft SQL Server

Fibre Channel Proxy

Fibre Channel

Existing Gold Copy of Data

Fibre Channel

Dual coherency links

Fibre Channel Proxy

Servers

Primary Storage

Transparently accelerate data access in the SAN

Solid State Performance with <u>no change</u> to:
Software - Databases - Servers - Storage - Processes

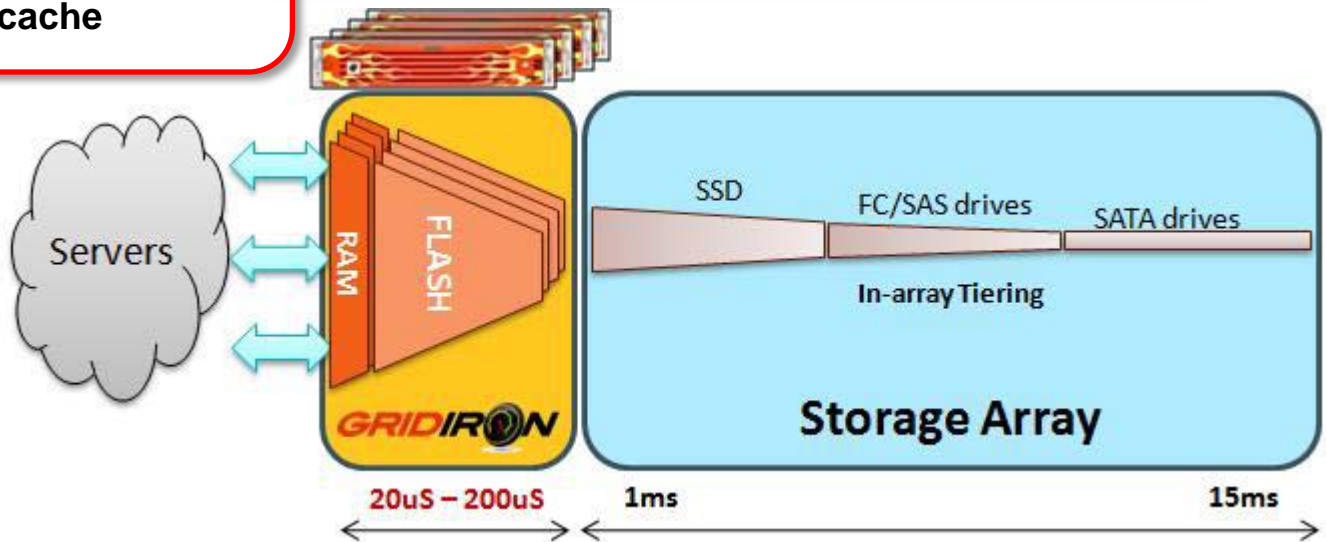# High Bandwidth Real-time Tiering

Cache only performance enhancing data

The GridIron Process
Learn data access graphs in real time
Use patterns to manage caching
Use feedback to continuously refine performance

GridIron provides
Higher concurrent IO bandwidth
Higher IOPS
Low latency multi-level cache

GridIron IP
HW Analytics
FC Proxy
Billion entry MMU
Flash Mgmt.

Servers

RAM FLASH

GRIDIRON

20uS – 200uS

SSD FC/SAS drives SATA drives

In-array Tiering

Storage Array

1ms 15ms

Results
- **Applications 2-10x faster**
- **Read latency 10-100x shorter**
- **Flash life a non-issue**

Real-World Deployments: High-Impact and Cost-Effective

Customer	Application / Business Problem	FastData	GridIron Performance Improvement	CapEx Savings With GridIron
shopzilla	**Oracle Data Warehouse** • "Real time" reports taking 6 hours • Lost revenue from delays • Over-provisioning storage for performance	• Bandwidth: >10GB/s • Concurrency: 25+ users • Data set size: 40TB DWH • Data turnover: continuous ETL	• Critical reports **6 hrs -> 30 mins.**	• $2M from storage and server consolidation
GILEAD	Large eDiscovery - **MS SQL under VMware** • Multi-hour query times affecting productivity • Need to support concurrent users • Serialized system impacting business	• Bandwidth: >2 GB/s • Concurrency: 4 users	• Query Times reduced **by >50%** • Increased query capacity **by 6x**	• Saved $775k on a storage upgrade (only 2x)
ACTIVISION HIGHMOON STUDIOS	Video game software builds under **VMware** • Builds taking 70 minutes to complete • Game quality impacted by long build time • Virtualized architecture not scalable	• IOPS: 40,000 Random • Concurrency: 24 users with parallel builds	• Build time reduced from **70 -> 8 mins.**	• $800K vs. alternatives
SONY	**Post production video editing** • Concurrent 2K streams needed • Support more users at lower cost • Complete system replacement is prohibitive	• Bandwidth: >3 GB/s • Concurrency: Five 2K video editing streams	• **5x improvement** in concurrent 2K video streams	• Avoided a complete system replacement

Handling Trends in SSDs

- **SLC / eMLC / heMLC / MLC / 3bMLC / ???**
  - Competing goals of endurance and economics
  - Lithography complicates choices
  - SSDs not the volume drivers for flash silicon

- **SSD flavors and toppings**
  - Interfaces, channels, controller horsepower
  - Smart wear reduction
  - Specified lifespan

Considerations for Evaluating SSDs

- **Initial performance will not (and should not) last**
  - Evaluate your workload over time
  - Must utilize intended capacity, alignments, operation sizes, etc.
  - More than 2X or continued degradation is suspect

- **Being "too smart" is not always good**
  - Compression and de-duplication are great – **for IOMeter**
  - Compression can cause high and uneven churn for indirection tables
  - Any background process in device can take away from IO processing

- **Consistency and predictability are key**
  - Architecture can plan around known performance corners
  - Characterize device over specific data rates and thread depth
  - Allows smart monitoring to detect deviations early

- **Who makes the drive, controller and flash chips**
  - Manufacturing, testing capability and history
  - DPM levels and FA process