Case Study - CacheIO

Marty Czekalski

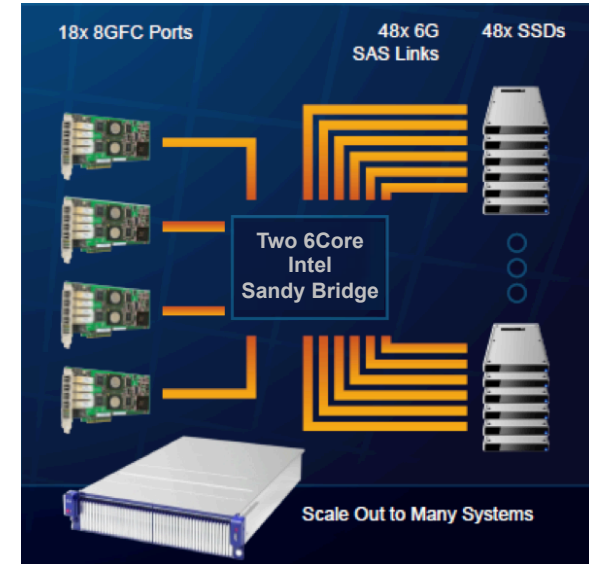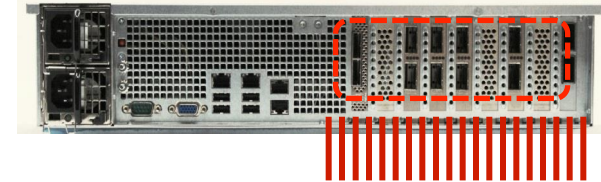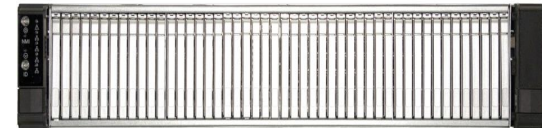## Seagate Technology

President, SCSI Trade Association

Challenge

- Design a high-performance SAN cache / high-performance solid state array
  - >1Million IOPs, cache operation
  - Multiple, uncompressed 4K UltraHD video streams (50MB/frame, 24 frames/sec)
  - Support scalable, standard interfaces
  - HA capable configurations
- Low cost
  - Off-the-shelf hardware and components
- Differentiation: Innovative software on commodity hardware

Components

- **SSDs**
  - 6Gb/s SATA, 7mm, 1TB each
- **Enclosure**
  - Echostreams 2U, 48 slots = 48TB
- **CPU**
  - 2P, Sandy Bridge, up to 512GB memory
- **Internal storage interfaces**
  - 3 x 16-port SAS HBAs
- **External interfaces**
  - 18 x 8Gb/s FC
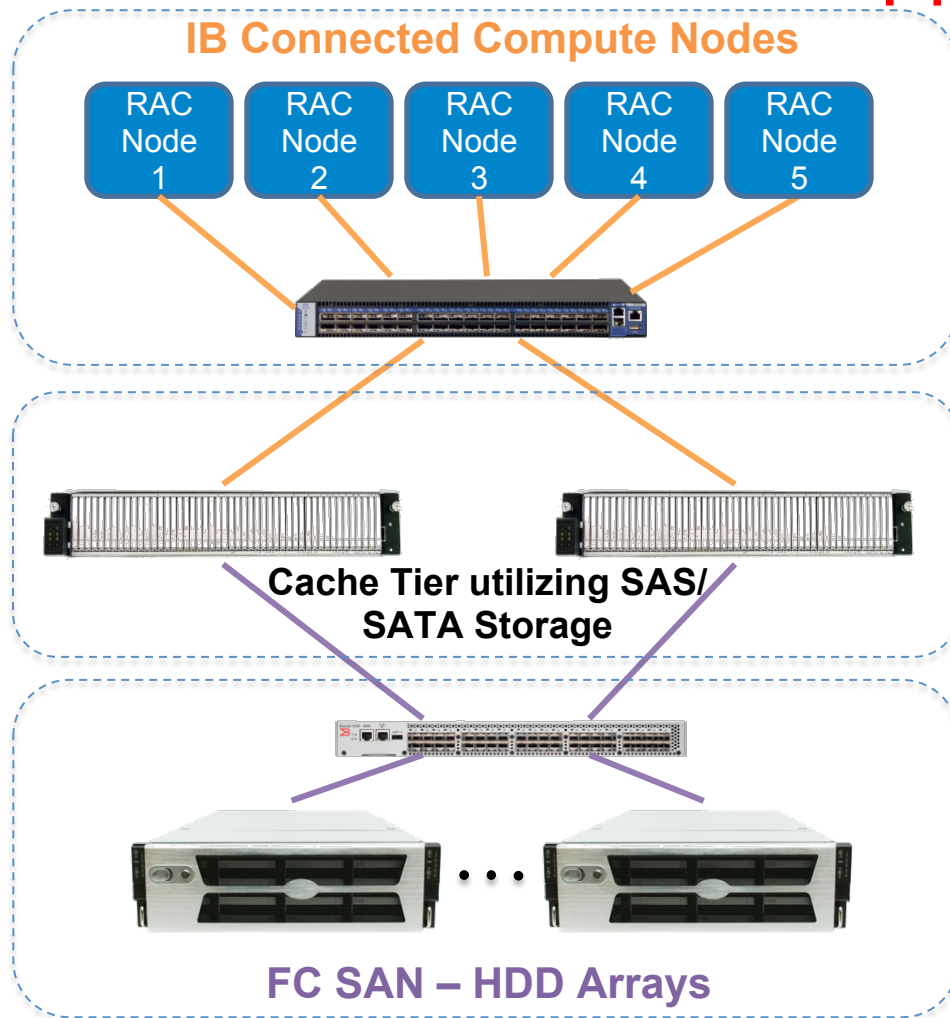 - 5 x 56Gb IB ports
  - iSCSI
- **Linux OS**

18x 8GFC Ports   48x 6G SAS Links   48x SSDs

Two 6Core Intel Sandy Bridge

Scale Out to Many Systems

Results

- **Performance**
  - **Cache**
    - 1.2 million random 4K IOPs on FC (sub ms access time)
    - 1.4 million random 4K IOPs on IB (sub ms access time)
    - Write latency 3 μs (write back, excluding network)
  - **Array**
    - Nine simultaneous uncompressed 4K video streams
- **HA with 2 units connected via InfiniBand**
- **SAS/SATA not a performance bottleneck**
  1. CPU cycles
  2. Memory Bandwidth
  3. Network

SCSI Trade Association

Flash Cache Performance with Oracle Real Application Clusters

IB Connected Compute Nodes

| RAC Node 1 | RAC Node 2 | RAC Node 3 | RAC Node 4 | RAC Node 5 |

Cache Tier utilizing SAS/ SATA Storage

FC SAN – HDD Arrays

- Accelerates Oracle RAC performance
- Mixed tier one downstream storage arrays
- Speeds up reads, redo log, writes
- Multi-path redundancy

2M DB Transactions per minute

How CacheIO increases Transactions Per Minute

- Reduces db read latency from 26.89 ms to 0.98 ms
- Reduces log sync latency from 2.33 ms to 0.48 ms

Before CacheIO

Wait		Event		Wait Time		
Class	Event	Waits	%Timeouts	Total(s)	Avg(ms)	%DB time
User I/O	db file sequential read	634,850	0.00	17,070.32	26.89	55.13
	DB CPU			5,773.42		18.64
Commit	log file sync	1,340,054	0.00	3,118.59	2.33	10.07
Configuration	log file switch (checkpoint incomplete)	636	0.00	2,651.99	4169.80	8.56

After CacheIO

Wait		Event		Wait Time		
Class	Event	Waits	%Timeouts	Total(s)	Avg(ms)	%DB time
	DB CPU			87,427.57		73.79
Commit	log file sync	24,788,838	0.00	11,960.47	0.48	10.09
User I/O	db file sequential read	4,488,001	0.00	4,376.98	0.98	3.69
Other	enq: FB - contention	111,451	0.00	2,154.44	19.33	1.82
Configuration	log file switch (checkpoint incomplete)	2,142	0.00	1,833.44	855.95	2.65

SCSI Trade Association

Summary

- **SAS and SATA provide a high performance storage subsystem at low cost with proven off-the-shelf components**
  - Performance bottleneck lies with the processor system
  - No benefit to using higher performance SSDs or interfaces which would only add cost