# Reducing Human Intervention in the Maintenance of Mass Storage Systems

May 5, 2008

Dave Fellinger, CTO
dfellinger@datadirectnet.com

# Reducing Human Intervention

- **Program defined in 2006**
  - Coding started in Q1 2006
  - Beta code tested in Q3 2006
  - Process continues with 3.11 delivered in Q1 2008
- **Program Goals**
  - Reduce instances of NTF drives and hardware
  - Increase mean time to human intervention
  - Change failure criteria to handle instances of drive and hardware anomalies
- **NTF drives reduced from >65% to <20% on product that can support RAID 6 error recovery**

# Agenda

- **Balanced I/O Systems**
- **MTBF and Availability**
- **Recovery Operations**
- **Parallel Storage Systems**
- **Conclusion**

# Balanced I/O in Clusters

- **Balance is essential for minimizing I/O in cluster computing**
- **DOE has generally used 1GB/s/TF**
- **Current systems range to 400GB/s**
- **Petaflop scale clusters assume 1TB/s**
  - At 100mB/s/drive one system would contain 10,000 drives at 90kW for SATA

# Challenges to I/O Balance

- **Current SAS drives still have full stroke seeks at 6.5mS**
- **Current SATA drives still have full stroke seeks at 22mS**
- **Recovery mechanisms are slowed by increased data density**
  - SAS drives take 75 steps
  - SATA drives take 300 steps
- **Block reassign times can range to tens of seconds including recovery**
  - SAS drives at 1.5s/LBA
  - SATA drives at 6s/LBA

# Agenda

- Balanced I/O Systems
- **MTBF and Availability**
- Recovery Operations
- Parallel Storage Systems
- Conclusion

# MTBF and Availability

- **MTBF is established by running a large drive sample**
  - 1000 drives running for 1000 hours without failure asserts a specification of 1x10^6 hours
- **Array MTBF=Drive MTBF/array size**
- **Availability=MTBF/(MTBF+MTTR)**
- **The goal in a large system is to reduce MTTR to a minimum to reduce data vulnerability**
- **Individual failure events must never affect data availability or performance**

# MTTDL and Redundancy

- **MTTDL relates to the number of redundant elements in a group**
- **In a RAID 6 with dual parity**
  - A single drive failure results in a redundant system
  - A dual drive failure results in continued data availability
  - A triple drive failure results in data loss

# MTTDL and Redundancy

- **Assuming that each drive has an MTBF of 1 million hours..**
  - The probability of one drive failing in an hour is 1/1,000,000
  - The probability of two drives failing in an hour is 1/1,000,000^2
  - The probability of three drives failing in an hour is 1/1,000,000^3

# Failure Criteria

## What constitutes a failure?

- **Hard failures include**
  - Head crashes
  - Bearing wear
  - Motor failure
  - Electronic hardware failure (ASC\ASCQ 04)
- **Soft failures must include**
  - Rereads
  - Dynamic block reallocation
  - Complete sector loss
  - Data corruption
  - Data recovery timeouts in excess of 20s

# Agenda

- Balanced I/O Systems
- MTBF and Availability
- **Recovery Operations**
- Parallel Storage Systems
- Conclusion

# Drive Error Recovery

- **Enterprise drives have less than 100 recovery steps**
- **SATA drives have over 200 recovery steps**
- **Each execute “free retries” regardless of the retry settings**
- **SATA recovery can range to 30 seconds**
  - Read and write of non-user data
  - Vary read amplifier characteristics
  - Re-read at +/- 6% of track width
  - Re-read at +/- 12% of track width
  - Adjustment of ECC parameters

# Background Error Recovery

- **Enterprise drives execute background data verification only when there is no host activity**
- **SATA drives can execute background data verification after any error recovery operation**
- **Recovery operations always include**
  - Optimization of servo position
  - Optimization of read amplifier operation
  - Read and write of sectors adjacent to a recovered block

# Agenda

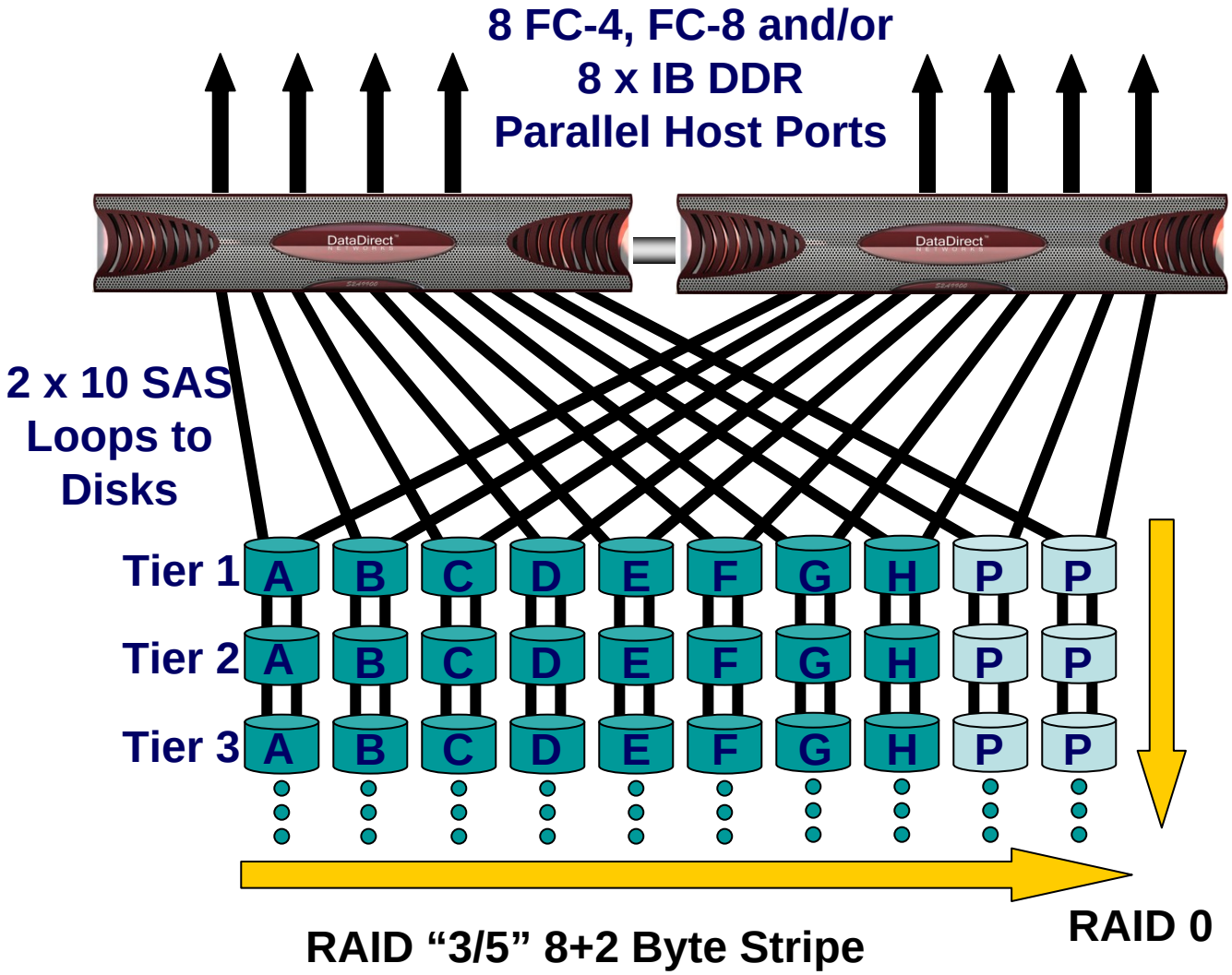
- Balanced I/O Systems
- MTBF and Availability
- Recovery Operations
- **Parallel Storage Systems**
- Conclusion

# Parallel Storage Goals

## Low Latency - High Performance, Silicon Based Storage Appliance

- Parallel Access For Hosts
- Parallel Access To A Large Number Of Disk Drives
- True Performance Aggregation
- Reliability From A Parallel Pool
- Quality Of Service
- Scalability
- Drive Error Recovery In Real Time
- True State Machine Control
  - 10 Virtex 4 FPGAs, 16 Intel embedded processors, 8 Data FPGAs

# An Implementation of Parallelism w/ Double Parity RAID Protection

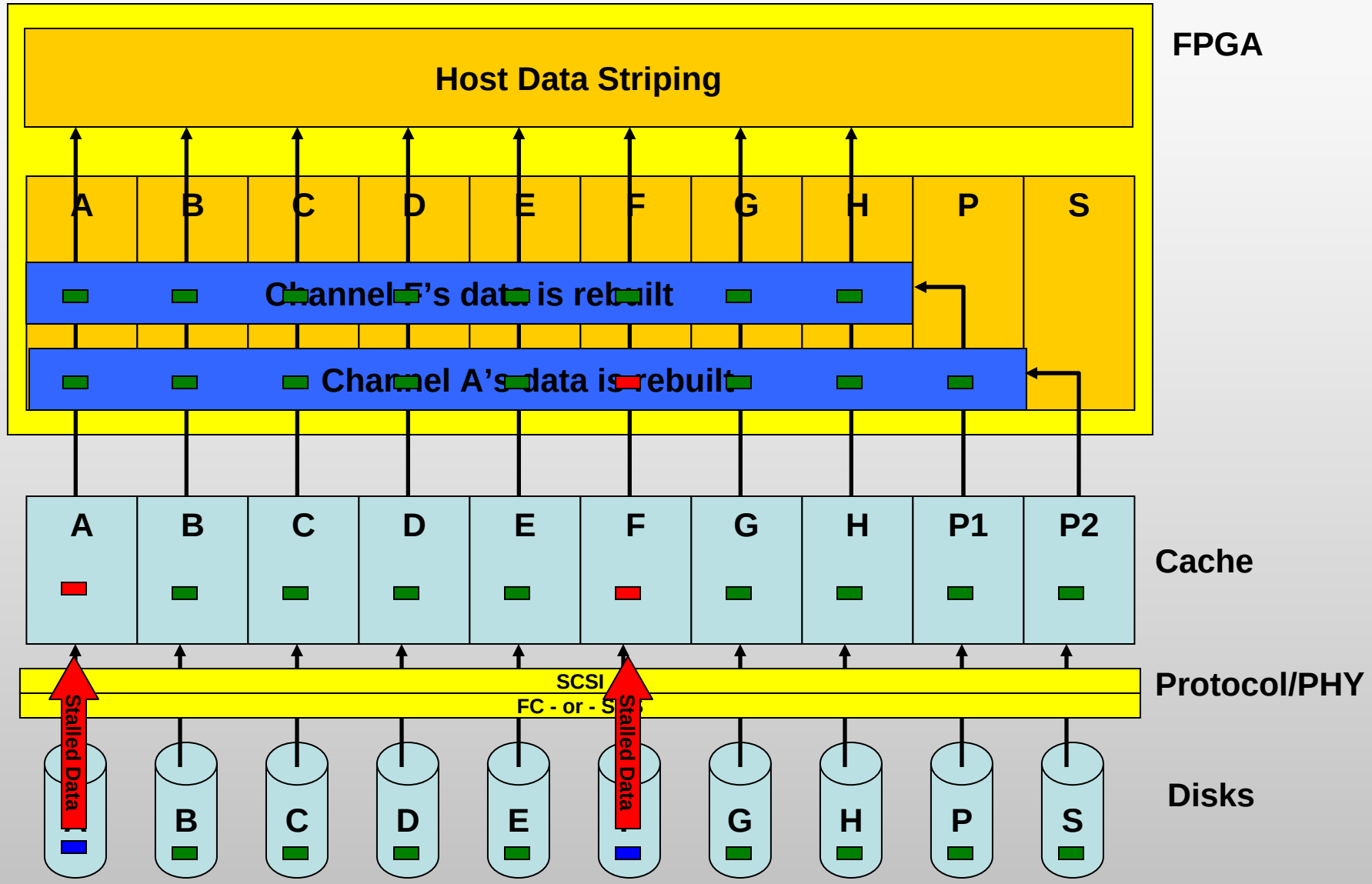



- Double Disk Failure Protection
- LUNs can span tiers
- All ports access all storage
- Implemented in Hardware State Machine
  - No penalty for RAID 6!
- Parity Computed On Writes AND Reads
- No loss of performance on any failure
- Multi-Tier Storage Support, SAS or SATA Disks
- Up to 1200 disks total
  - 960 formattable disks

# Quality of Service

- S2A always reads (and writes) to all members of a RAID group
- FPGA designed to generate host data with missing elements
- If a single member of RAID group is slowed by internal error recovery S2A can still provide host data at a high level of QOS
- All data passes through a Parallel Data Recovery Engine which recovers stalled or missing data

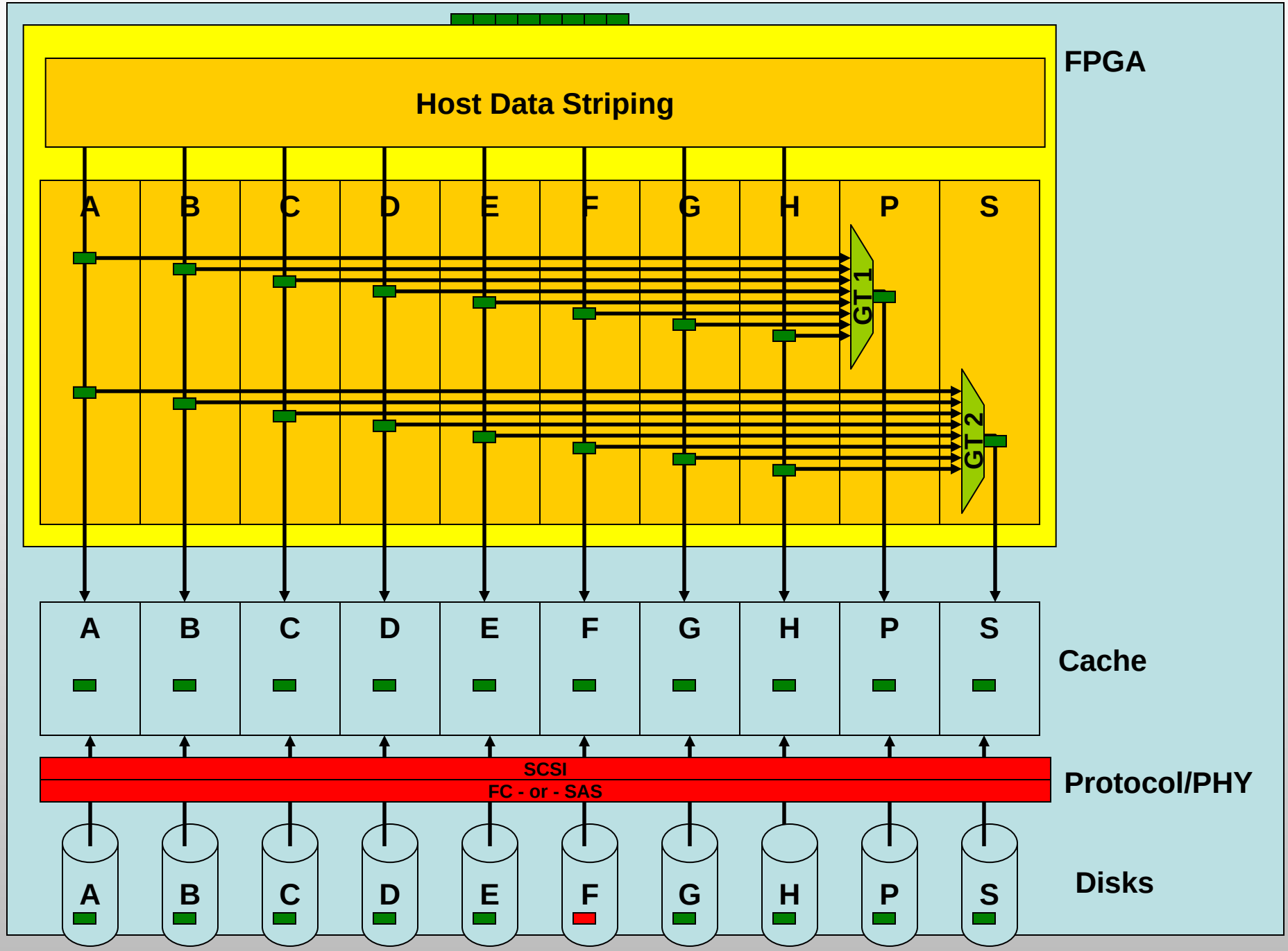
# Quality of service



# Data Corruption Error Handling

**Note that the Cache and Disks have not corrected the data corruption.**

**We will need to rebuild the data into the cache and flush the data back to the disk in order to repair the problem fully.**


FPGA
Host Data Striping
A
B
C
D
E
F
G
H
P
S
GT 1
GT 2
A
B
C
D
E
F
G
H
P
S
Cache
SCSI
FC - or - SAS
Protocol/PHY
A
B
C
D
E
F
G
H
P
S
Disks

# SATAssure Data Integrity

- **The Parallel Data Recovery Engine allows data reconstruction and integrity checking**
  - S2A hardware enables SATAssure software to verify all data read from the disks
  - S2A hardware allows SATAssure to send hosts “fixed” data (data integrity is assured)
  - S2A hardware enables SATAssure to correct data on the disk for future accesses (self-healing array)
  - Multiple levels of disk recovery attempted before failing drives (replace fewer drives)
  - S2A controller journaling allows partial rebuilds (less time in degraded mode)

# Worst Case Recovery

- **Disks can become completely unresponsive to all commands**
  - The internal OS can enter a loop that does not enable external commands
  - A power cycle always recovers the drive
  - S2A 9900 automatically power cycles a drive in place
- **SAS drives can be issued LLF in place**
  - Platters are rechecked for integrity and the sectors are rewritten
- **Drives that issue SMART warnings or grow defects at an increasing rate are copied to channel spares**

# Agenda

- Balanced I/O Systems
- MTBF and Availability
- Recovery Operations
- Parallel Storage Systems
- **Conclusion**

# Conclusion

- Bit error rates and drive error recovery mechanisms are a statistical reality
- Large checkpoint systems must maintain a very high data rate to minimize the I/O cycle
- MTBF and MTTHI must be disconnected
- Storage systems must execute self test and repair to minimize human intervention in the machine room
- If human intervention is required every possible automation assist must be employed

DataDirect NETWORKS

S2A TECHNOLOGY ADVANTAGE

# Thank You.