Why Data Integrity is important to you

Richard Vanderbilt, LSI
Why Data Integrity is important to you!

© 2008 Storage Networking Industry Association. All Rights Reserved.
Abstract

Why Data Integrity is important to you

- This session will appeal to Data Center Managers, Development Managers, and those that are seeking a fundamental understanding how and why silent data corruptions occur, the affects it has on a data center and how the industry is moving to minimize its occurrence.
Data Integrity VS Corruption

❖ At the Storage Level, There Are Two Kinds of Data Corruption
   ♦ Latent sector errors
   ♦ Silent data corruption

❖ For a Storage Device, it is Usually the Case That…

it is Better to Not Return Any Data, Than Return the Wrong Data!
Typical Causes of Data Integrity Problems

- Operating System Bugs
  - Core O/S
  - Device drivers
- Storage Hardware and Firmware Bugs
  - HBAs
  - Arrays
  - Disks
- Administrative Errors- “The human factor”
  - System administrators
  - Database administrators
  - Storage administrators
Technology transitions can add risk!

Server and Storage ecosystem is changing

- Multi Core processors 2, 4, 8, 16….
- Ever expanding memory 2GB, 4GB, 8GB, 32GB…
- Doubling wire speeds 8GFC, 10GbE, FCoE, Infiniband…
- Higher density drives, doubling every six months
- Geographically dispersed processing, virtual data centers, Grids, Clouds …
- Commodity SATA drives entering the enterprise

Technology sectors progress independently of each other

- Plugfests focus on their particular sector, not the entire solution
Growth can add risk!

Petabyte Shipment Trends — External Controller-Based Disk Storage

Raw Petabyte Shipments 2002-2006
- Modular Petabytes
- Monolithic Petabytes

Raw Petabyte Shipment Forecast 2007-2011
- Modular Petabytes
- Monolithic Petabytes

Ref 1 Gartner '07
Why Data Integrity is important to you!

The Cause!

- Hardware or system problem
- Human error
- Software corruption or problem
- Hardware or system problem

Perception vs. Reality

<table>
<thead>
<tr>
<th>Category</th>
<th>Perception</th>
<th>Reality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disaster</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td>Computer viruses</td>
<td>2%</td>
<td>4%</td>
</tr>
<tr>
<td>Human error</td>
<td>9%</td>
<td>11%</td>
</tr>
<tr>
<td>Hardware or system problem</td>
<td>78%</td>
<td>56%</td>
</tr>
</tbody>
</table>

Ref 2 Source: Kroll Ontrack 2007
The Effect!

- A well known e-commerce company was forced to shut down for days when a bug in the file manager caused bad data to be written onto their database.

- A leading financial services company experienced repeated corruptions when a problem in the virtual memory system caused the wrong data to be written onto the database.

### The Cost of Downtime

Average financial impact per hour of interrupted computer operations by industry

<table>
<thead>
<tr>
<th>Industry</th>
<th>Average Financial Impact Per Hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retail Brokerage</td>
<td>$6,450,000</td>
</tr>
<tr>
<td>Credit Card Sales</td>
<td>$2,600,000</td>
</tr>
<tr>
<td>Pay-Per-View</td>
<td>$150,000</td>
</tr>
<tr>
<td>Home Shopping (TV)</td>
<td>$113,000</td>
</tr>
<tr>
<td>Catalog Sales</td>
<td>$90,000</td>
</tr>
<tr>
<td>Airline Reservations</td>
<td>$89,500</td>
</tr>
</tbody>
</table>

Source: Gartner Group & Contingency Planning Research, Inc.

Why Data Integrity is important to you!
© 2008 Storage Networking Industry Association. All Rights Reserved.
File Systems Today

- **Observations:**
  - Bugs are common
  - Numerous bugs across the file systems tested, some of which are serious, and many of which are not found by other sophisticated techniques

- **Detection:**
  - Sanity checking is of limited utility
  - Many of the file systems use sanity checking
  - Modern disk failure modes such as misdirected and phantom writes lead to cases where:
    - Bad block thus passes sanity checks, is used and can corrupt the file system
    - Indeed, all file systems tested exhibit this behavior

- **Recovery:**
  - Automatic repair is used rarely by the file systems
  - Most of the file systems require manual intervention . . . (i.e., running fsck)

Ref 4: Phd thesis of Vijayan Prabahakaran
Data Integrity Study 1

Disk Errors

- A special 2 GB file was written to more than 3,000 nodes every 2 hours and read it back
- Errors were checked for 5 weeks
- 500 errors were found on 100 nodes
  - Single bit errors: 10% of disk errors
  - Sector (512 bytes) sized errors: 10% of disk errors
  - 64 KB regions: 80% of disk errors

RAID Errors

- Verify command was run on 492 RAID systems each week for 4 weeks
- The disks are spec’ed at a Bit Error Rate of 10^14 read/written
- The good news
  - That the observed BER was only about a 3rd of the spec’d rate
- The bad news
  - That in reading/writing 2.4 Petabytes of data there were some 300 errors

Ref 5: CERN Study, April ’07
Memory Errors

- **Good news:**
  - 600,000 single bit errors were detected (1/3 of the expected rate at $10^{12}$)
  - only 3 double-bit errors in 3 months on 1300 nodes

- **Bad news:**
  - according to the spec there shouldn’t have been any double bit errors AND these can’t be corrected

All of These Errors Will Corrupt User Data

- 8.7 TB of user data was checked for corruption, some 33,700 files...
- They found 22 corrupted files
- Which is 1 in every 1500 files

Ref 5: CERN Study, April ‘07
Data Integrity Study 2

- 1.53M Drives Monitored Over 41 Months
  - 1.17M FC drives, 358K SATA drives
- Three Classes of Corruption Were Identified
  - Random
  - Lost writes (365 disks)
  - Parity inconsistencies (data scrubbing)
- Random Corruptions Detected With Data Scrubbing
  - 49% for SATA
  - 73% for FC
- Silent Data Corruption
  - Any media error not detected by the drives internal protection
  - 400,000 instances were recorded
  - 8% (on average) were found during RAID reconstructions
  - 3088 SATA drives developed silent data corruption (0.86%)
  - 767 FC drives developed silent data corruption (0.065%)

Ref 6: FAST’08 Paper
Individual companies strive to build better protection mechanisms within their own products
- This addresses component failures only
- Many disparate and proprietary technologies addressing device failures

True end to end data integrity solutions require cooperation of everyone in the stack
- Exchanging protection metadata between application and spindle
- Application and or file system creates protection metadata
- Metadata is verified in-flight and at rest
- Corruption detection can occur before data is written to disk
Application-to-Disk Data Integrity

- Data Integrity Solution Overview
  - Create integrity metadata coincident with data creation
  - Validate the metadata throughout the data path
  - Errors are directed to the application for remediation

Application and/or File System creates metadata tag

HBA validates metadata while in flight

Storage Array validates metadata in flight

Disk Drives executes final validation before writing to platter

Why Data Integrity is important to you!
© 2008 Storage Networking Industry Association. All Rights Reserved.
Data Integrity Model

SNIA Storage Model

Case 1: Filesystem/Database
PI Creator, Owns the Metadata

Errors are detected and sent to creator

PI Consumer, Uses the Metadata
Why Data Integrity is important to you!

SNIA Storage Model

Case 2: Application
PI Creator, Owns the Metadata

Errors are detected and sent to creator
Why Data Integrity is important to you!

Data Integrity Model

SNIA Storage Model

Case 3: Block Aggregation
PI Creator, Owns the Metadata
PI Consumer, Uses the Metadata
Data Integrity Schemes

T10-DIF + DIX (Data integrity plus extensions)

T10-DIF (Data Integrity Field)

Other industry solutions

Normal I/O

Vendor Specific Integrity measures
Application
Vendor Specific Integrity measures
OS/File System
Vendor Specific Integrity measures
I/O Controller
Vendor Specific Integrity measures
Disk Array
Vendor Specific Integrity measures
Disk Drive

Validates Data From End to End (True E2E)
Validates Data from Driver to Spindle
Validates Data in flight to the Array

From SNIA DITWG 080131
T10 Data Integrity Field (DIF)

- Interleaved With Data Sectors on the Wire
- Three Protection Schemes: Type 1, 2 & 3
  - All have guard tag defined
  - Type 1 reference tag is lower 32 bits of target Logical Block Address
  - Type 2 reference tag is seeded in 32-byte Command (CDB)
  - Type 3 reference tag and application tag are combined
T10 Data Integrity Field (DIF) + Data Integrity Extensions (DIX)

Why Data Integrity is important to you!
© 2008 Storage Networking Industry Association. All Rights Reserved.
Why Data Integrity is Important to you

- Silent data corruptions can have irreparable damage
- Backups may have bad data, and can’t be trusted
- A data corruption could cause your DBA to take the wrong corrective action
- Downtime or corruptions will be costly, and may be fatal to a company

What Should You Do?
- Promote Standards - SNIA has recently approved Data Integrity (DI) as a formal Technical Working Group (TWG)
- Promote the Technology – Research with your server/storage supplier their plans to enhance data integrity
Q&A / Feedback

Please send any questions or comments on this presentation to SNIA: trackstorage@snia.org

Many thanks to the following individuals for their contributions to this tutorial.
- SNIA Education Committee

Dave Crespi
Tom Hammond-Doel
Brian McKea
Supporting Studies

- Ref 1: Gartner Study 2007
- Ref 2: Kroll On track 2007
- Ref 3: Gartner Group & Contingency Planning Research Inc.
- Ref 4: IRON File Systems
  - V. Prabhakaran, L. Bairavasundaram, N. Agrawal, H. Gunawi – University of Wisconsin, Madison
- Ref 5: Data Integrity
  - Bernd Panzer-Steindel CERN/IT, April ’07
- Ref 6: Disk Failures in the Real World
  - L. Bairavasundaram, G. Goodson, B. Schroeder, A. Arpaci-Dusseau, R. Arpaci-Dusseau, FAST’08