White Paper

An Integrated End-to-End Data Integrity Solution to Protect Against Silent Data Corruption

Abstract

This white paper describes how T10 PI prevents silent data corruption, ensuring that incomplete and incorrect data cannot overwrite good data. Without T10 PI, data corruption events may result in system downtime, lost revenue, or lack of compliance with regulatory standards.

August 2012

EMC²

Part Number H11058

Table of contents

Introduction

T10 Protection Information (T10 PI), previously known as Data Integrity Field (DIF), is an important standard that supports the industry's commitment to end-to-end data integrity validation. T10 PI prevents silent data corruption, ensuring that incomplete and incorrect data cannot overwrite good data. Without T10 PI, data corruption events may result in system downtime, lost revenue, or lack of compliance with regulatory standards.

As the industry leader in enterprise data protection and availability, EMC® Symmetrix® intends to be the first enterprise storage array to join with Emulex and Oracle in implementing end-to-end T10 PI.

The data protection information generated by Oracle Automatic Storage Management (ASM) is validated first by the host operating system, then by the Emulex LightPulse 8 Gb Fibre Channel Host Bus Adapter (HBA, model number LPe12000-E), and finally by EMC Symmetrix VMAX® 40k storage array with EMC Enginuity™ version 5876.82.57 or later, ensuring protection through the I/O stack.

Purpose

The purpose of this document is to provide information regarding the addition of T10 PI on EMC VMAX series products and the results of the joint testing effort of EMC, Oracle, and Emulex.

Scope

This document focuses on the initial release of the EMC, Oracle, and Emulex joint T10 PI solution.

Audience

This document is intended for those seeking a method to overcome silent data corruption and enhancing the integrity of their data stored on EMC storage.

Technology Overview

This section provides information on the three components used to achieve end-to-end data integrity, each discussed briefly in this section:

- Unbreakable Enterprise Kernel for Oracle Linux
- Emulex LightPulse 8 Gb Fibre Channel HBA
- EMC Symmetrix VMAX Family

Unbreakable Enterprise Kernel for Oracle Linux

For the implementation discussed in this White Paper, the Unbreakable Enterprise Kernel [kernel-uek-2.6.39-200.24.1.el6uek] for Oracle (also available as part of Oracle Linux 6.3 as a default kernel) is recommended.

Unbreakable Enterprise Kernel contains many new features that are relevant to Oracle Linux running in the data center, including data integrity features.

Unbreakable Enterprise Kernel, including the data integrity features, is provided under the GNU General Public License (GPL) and is available to anyone in both binary and source form. As of this writing, binary versions of the kernel are provided via Unbreakable Linux Network (ULN) and Oracle's public yum server.

Subsequent releases of Oracle Linux will include Unbreakable Enterprise Kernel as an option on the installation media, which can be downloaded for free from edelivery.oracle.com/linux. Existing Oracle Linux support customers receive full support for this kernel as part of their existing support subscriptions.

Bug fixes and security errata are delivered via ULN and announced through the el-errata mailing list.

Emulex LightPulse 8 Gb Fibre Channel HBA

The Emulex LightPulse 8 Gb FC HBAs (model numbers LPe12000-E and LPe12002-E), with the Emulex BlockGuard feature, is a key component in the Oracle's Data Integrity solution. BlockGuard ensures that data corruption events do not go undetected as data traverses the system, from the operating system and application to the disk array storing valuable data. The PCI Express 2.0 Emulex HBA includes BlockGuard, which provides T10 Protection Information (T10 PI) and Oracle Data Integrity Extensions (DIX).

As part of the overall ecosystem deployment, once the Oracle Database Application creates data in memory, the ASM generates protection information which the Oracle Linux kernel then forwards to the Emulex HBA using Data Integrity Extensions (DIX).

The Emulex HBA verifies that the data, protection information, and target location match and then interleaves the data and protection information and transmits 520-byte sectors to the storage.

At this point, the Emulex HBA has completed its job. Now, the storage array controller, followed by the disk drive firmware, verifies that the data, protection information, and target location match. If a successful I/O completion ensues, it is then reported back to the application.

Any mismatch detected by the HBA (or storage array and disk drive) causes the I/O to abort and the error is passed up the stack, preventing bad data from being written. Protection information is transmitted to read requests and the ASM verifies I/O before signaling completion to the application.

Lastly, when using legacy storage, protection information exchange is dynamically negotiated and automatically enabled between the application and HBA.

EMC Symmetrix VMAX Family

As the industry leader in enterprise data protection and availability, the EMC Symmetrix VMAX Family is the first enterprise storage array to join with Emulex and Oracle Linux in implementing end-to-end T10 PI. The data protection information generated by the Oracle ASM is validated by the Oracle Linux operating system, then passed on to the EMC Host Bus Adapter (HBA) and the EMC VMAX storage array, ensuring protection through the I/O stack.

The EMC Symmetrix VMAX Family consists of the VMAX 40K, VMAX 20K, and the VMAX 10K.

VMAX 40K

The VMAX 40K is built for Hybrid Cloud environments and provides the industry's highest levels of consolidation, performance and scalability.

VMAX 20K

The VMAX 20K is built for performance, consolidation, and automation for demanding virtual data center environments.

VMAX 10K

The VMAX 10K is the most affordable multi-controller array, and built for performance and efficiency to consolidate applications in virtual environments

Providing End-to-End Integrity

When fully implemented, end-to-end data integrity consists of components that support Data Integrity Extensions (DIX) and T10 Protection Information (T10 PI).

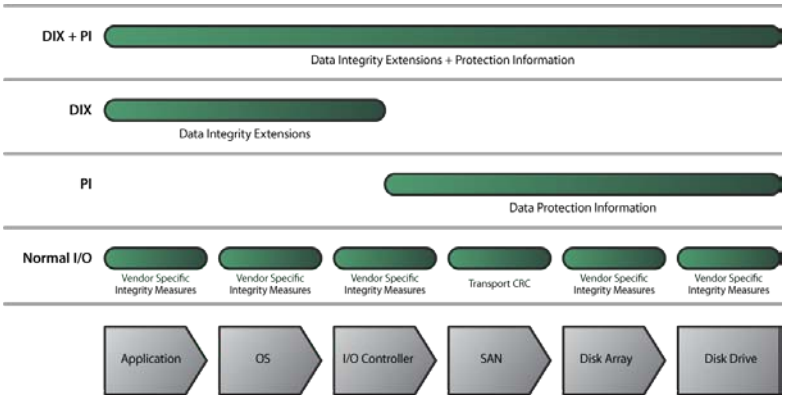

Figure 1. Achieving end-to-end data integrity

When writing date, end-to-end data integrity consists of the following steps:

1. The Oracle ASM library adds integrity metadata for each 512-byte sector as it is written to memory.
2. The integrity metadata is attached to the I/O request and passed through the layers in the operating system kernel to the Emulex driver.
3. The Emulex 8 Gb Fibre Channel adapter collects the information from memory buffers, verifies the data integrity, merges the data and integrity metadata, and sends out 520-byte sectors.
4. The EMC Symmetrix VMAX array firmware, Enginuity 5876.82.57, verifies the integrity metadata, and writes to disk.
5. The disk drive firmware verifies the integrity metadata before committing the data to physical media.

These steps are completed in reverse when reading data.

Solution Verification

Test Environment

The test environment consisted of the following, as shown in Figure 2 on page 7:

- Intel x86_64 based server with an Emulex LPe12002-E and firmware 2.01a10 installed
- Oracle Linux 6.3 with UEK kernel version 2.6.39-200.24.1.el6uek and the in kernel Emulex driver (lpfc) version 8.3.5.68.6p
- JDSU Fibre Channel analyzer

- SAN consisting of Brocade Fibre Channel switches
- VMAX 40k with Enginuity 5876.82.57

Configuration

The following figure depicts the configuration used to verify the solution.

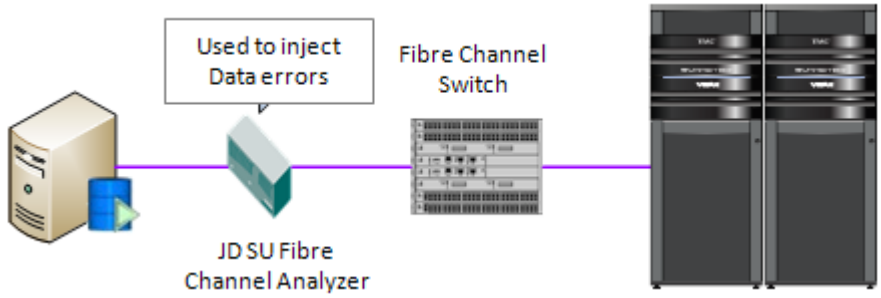

Figure 2. Testing configuration

Figure 3 explains how the stack works.

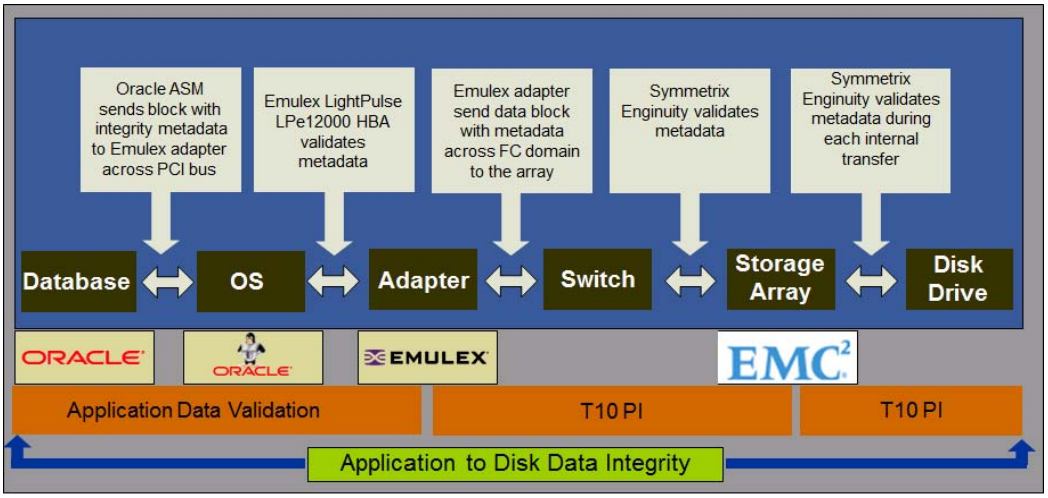

Figure 3. How the stack works

Method

EMC has tested the EMC/Oracle/Emulex T10 PI solution end-to-end in EMC's E-Lab™ and Symmetrix development labs. Testing included fault insertion through SAN events such as cable pulls, host failures, and storage failures, as well as the insertion of corrupted data between the host and target using the JDSU Fibre Channel analyzer. Additionally, corrupted data was inserted throughout the operating system stack to ensure compliancy with the specification. Target, initiator, and the operating system stack were monitored to ensure the T10 PI specification was met and that silent data corruption did not occur.

Conclusion

The stack as tested met the T10 PI specification, preventing the possible occurrence of silent data corruption. The insertion of faulty data on "the wire" was detected throughout the stack and appropriate responses were logged.

Actions were taken by the respective end-points (depending on the direction the errant data was injected) during both reads and writes to the array as well as when sent up the OS stack by injecting data errors within the operating system stack.

References

Refer to EMC Online Support website (registration required) for the following EMC documentation, at https://support.EMC.com:

- *EMC Host Connectivity Guide for Linux*
- *EMC Solutions Enabler Symmetrix Array Controls CLI V7.4 Product Guide*

Refer to the following website for Oracle Linux documentation:

- www.oracle.com

Refer to the following website for Oracle documentation:

- *Unbreakable Enterprise Kernel R.2 for Oracle Linux*
  www.oracle.com/linux

Refer to the following website for Emulex documentation:

- Emulex Product website for more information on Emulex-branded HBAs, at www.emulex.com/products/fibre-channel-hbas.html
- Emulex-EMC website for more information on EMC-branded HBAs , at www.emulex-emc.com