White Paper

Managing Correctable Memory Errors on Cisco UCS Servers

This document provides empirical evidence that shows no correlation between correctable and uncorrectable errors. Furhermore, using industry-standard benchmarks, this document demonstrates that systems with correctable errors do not exhibit system performance degradation. Given these findings, starting with UCS Manager (UCSM) 2.2(7), 3.1(1), and Cisco Integrated Management Controller (CIMC) 2.0(9) for rack standalone, Cisco UCS server memory-error threshold policies will not declare modules with correctable errors to be degraded. For customers who are not on UCSM 2.2(7), 3.1(1), or CIMC 2.0(9) or newer that experience a degraded memory alert for correctable errors, the Cisco UCS team recommends that memory modules with correctable errors not be replaced immediately upon alert. Instead please reset the memory-error counters and resume operation.

Contents

Field Recommendations: Correctable Errors and Threshold Policies **3**

Overview of Memory Errors **3**

Classification of Memory Errors 3
Correctable Versus Uncorrectable Errors 3

Trends in Server Memory Systems **4**

Increased Capacity 4
Increased Bandwidth 4
Reduced Operating Voltages 4

Handling of Memory Errors **5**

Cisco UCS Server ECC Capabilities 5
Scrub Protocol 5
Field Memory-Error Threshold Policies 5

Correctable Errors and Uncorrectable Errors **6**

Correctable Errors and System Performance 7

Future Enhancements to Cisco UCS Error Management **7**

Conclusion **7**

Appendix A: How to Reset Memory Errors **8**

Cisco UCS B-Series and C-Series Operating in UCSM 2.2 and 3.1 8
Cisco UCS B-Series and C-Series Operating in UCSM 2.1 8
Cisco UCS C-Series Rack Servers Operating in Standalone Mode 8

Appendix B: Additional Resources **9**

Field Recommendations: Correctable Errors and Threshold Policies

All Cisco UCS servers have ECC and scrub patrol capabilities that can effectively address modules with correctable errors. Findings from a multiyear Cisco study showed no correlation between incidences of correctable errors and uncorrectable errors on a module. Also, the incidence of correctable errors does not degrade system performance. For details on this study, please read the Correctable Errors and Uncorrectable Errors section of this paper. For these reasons, starting with UCS Manager (UCSM) 2.2(7), 3.1(1), and Cisco Integrated Management Controller (CIMC) 2.0(9) for rack standalone servers, Cisco UCS server memory-error threshold policies will not declare modules with correctable errors to be degraded. Note: For this change to be effective, both the UCS infrastructure firmware including UCS Manager and the host firmware must be at UCSM 2.2(7), 3.1(1), or newer.

For customers who are not on UCSM 2.2(7), 3.1(1), or CIMC 2.0(9) or newer, that experience a degraded memory alert for correctable errors please reset the memory-error counters and resume operation. Appendix A provides steps for resetting memory-error counters. If the correctable error condition persists for a particular memory module after resetting of error counters, reset the errors again and resume normal operation, but also plan for a replacement of the DIMM when convenient.

Overview of Memory Errors

Memory errors are encountered when an attempt is made to read a memory location. The value read from the memory does not match the value that is supposed to be there.

Classification of Memory Errors

Detected Versus Undetected Errors

A system without error-correcting code (ECC) memory will not detect hardware errors. Hence, memory errors will silently lead to data corruption, incorrect processing of the operating system or application, and eventually system failures. Cisco Unified Computing System™ (Cisco UCS®) servers use ECC memory. Therefore, powerful error correcting codes such as those provided by the Intel® Xeon® processors in Cisco UCS servers can detect memory errors so that silent data corruption does not occur.

Hard Versus Soft Errors

Errors that are caused by a persistent physical defect are traditionally referred to as “hard” errors. A hard error may be caused by an assembly defect such as a solder bridge or cracked solder joint, or may be the result of a defect in the memory chip itself. Rewriting the memory location and retrying the read access will not eliminate a hard error. This error will continue to repeat.

Errors caused by a brief electrical disturbance, either inside the DRAM chip or on an external interface, are referred to as “soft” errors. Soft errors are transient and do not continue to repeat. If the soft error was the result of a disturbance during the read operation, then simply retrying the read may yield correct data. If the soft error was caused by a disturbance that upset the contents of the memory array, then rewriting the memory location will correct the error.

Hard errors are typically detected by memory tests run by the Cisco UCS BIOS at boot time, and any modules containing hard errors are mapped out so that they cannot cause errors during runtime. Cisco UCS servers employ memory patrol scrubbing to automatically detect and correct soft errors during runtime.

Correctable Versus Uncorrectable Errors

Whether a particular error is correctable or uncorrectable depends on the strength of the ECC code employed in the memory system. Dedicated hardware is able to fix correctable errors when they occur with no impact on

program processing. Uncorrectable errors generally cannot be fixed and may make it impossible for the application or operating system to continue processing.

Trends in Server Memory Systems

DRAM has become ubiquitous as a main-memory solution in almost all computer systems. In such systems, memory errors have been a concern for a long time. The trends in complex computer systems, particular in today's servers, exacerbate the incidence of these memory errors. Such trends include increased memory capacity, increased bandwidth, and reduced operating voltages.

Increased Capacity

A primary reason for increased error rates is the rapid increase in the size of memory systems. As more bits of memory are added to the system, the likelihood that any one of them will encounter an error increases. Such increases in system memory capacity are the result of the shrinking size of DRAM modules (that is, more bits can be packed on a single die). Since 2008, DRAM capacities have increased 16x, from 512 megabits to 8 gigabits. As chip capacity has increased, individual bit cells have been getting smaller. As the bit cell gets smaller, the number of stored charges per bit decreases, making it more difficult to distinguish between a stored 1 and a stored 0. The basic storage element, or bit cell, in a DRAM chip is a tiny capacitor. DRAM bit cells are inherently leaky, and smaller bit cells storing fewer chargers are less tolerant of this leakage. Additionally, smaller bit cells are more easily upset by external sources such as alpha particles and cosmic rays. Today's advanced DRAM technologies deliver up to 8 gigabits of memory on a single die and up to 64 gigabytes of memory on a single memory module. In addition, today's processors incorporate multiple memory channels on each processor socket and multiple modules on each channel.

Increased Bandwidth

Memory-system bandwidth has also been increasing steadily. Not only does each processor socket have multiple memory channels, but the speed of those channels has increased. Just a few years ago, the top speed for DDR2 memory interfaces was 800 mega-transfers per second (Mtps). Using advanced DDR4 memory, the Cisco UCS B200 M4 Blade Server supports memory channels operating at 2133 Mtps. Ever-increasing operating frequencies, while providing higher bandwidth, also result in smaller bit times. As individual bit times decrease, timing margin also decreases, making it more difficult for receiving circuitry to separate each bit from those that precede and follow it.

Reduced Operating Voltages

Another underlying technology trend is an ongoing reduction in operating voltages, to lower power and cooling requirements and to accommodate the smaller transistors associated with advances in processor technology. DRAM voltages have decreased over the years, declining from 2.5V to 1.8V, 1.5V, 1.35V, and then to 1.2V as the industry has shifted from DDR to DDR2, DDR3, and then to DDR4. As the operating voltages decrease, the available noise margin also decreases, making it more difficult for receivers and sense amplifiers to distinguish between a 1 and a 0.

Handling of Memory Errors

As explained in the previous section, increased error rates can be attributed to multiple trends in the server industry. Cisco UCS servers handle memory errors in a way that does not compromise the reliability, availability, and serviceability (RAS) of the server. Memory errors are handled through three main technologies: server ECC capabilities, scrub protocols, and field memory-error threshold policies.

Cisco UCS Server ECC Capabilities

All Cisco UCS servers use memory modules with ECC codes applied across 64-bit (8-byte) data words protected by 8 check bits to form a 72-bit code word. Such single error correcting and double error detecting (SECDED) ECC codes could correct any single-bit error and detect any double-bit error. In addition, Cisco UCS servers built from Intel Xeon EP-class processors employ ECC codes that not only correct any single-bit error, but also correct errors confined to a single x4 DRAM chip and detect errors in up to two devices. This capability is known as single-device data correction (SDDC).

Additionally, when a system is operating in lockstep mode, which spreads the ECC code words across a pair of memory channels, SDDC is extended to correct errors in any x8 bit DRAM chip (or adjacent pair of x4 DRAM chips). To provide even greater reliability and availability, Cisco UCS servers built from the Intel Xeon EX-class processors can correct errors in any (not necessarily adjacent) pair of x4 devices and can detect errors in up to three devices. This capability is known as double-device data correction (DDDC).

Scrub Protocol

In all normal memory read accesses, the memory controller checks for and corrects single-bit errors. However, sometimes the data in the entire memory array may not be accessible for reasons related to data locality. Thus, scrub patrol protocols provide additional correction capabilities that are needed beyond the usual SECDED ECC codes. The scrub patrol routine reads the entire memory array and corrects any single-bit errors. This patrol routine occurs periodically at a predetermined interval, usually once every 24 hours.

Field Memory-Error Threshold Policies

In addition to ECC capabilities and scrub protocol, Cisco UCS servers employ field memory-error threshold policies that flag certain memory modules as candidates for replacement after the module reaches a certain memory-error threshold. If any memory module generates an uncorrectable error, that module is flagged as degraded. A similar logic is used to flag modules that generate correctable errors also as degraded.

A problem with current threshold policies is that both correctable and uncorrectable errors are flagged as degraded even though advanced ECC capabilities and scrub protocols effectively address all correctable errors. Flagging a module with only correctable errors as degraded is premature, because data shows that in many cases correctable errors are transient and will resolve themselves with little user intervention. In addition, when a module is flagged as degraded from correctable errors, it is replaced, causing unnecessary disruption to server uptime.

Historically, memory modules with correctable errors were treated the same as modules with uncorrectable errors based on the following two premises:

- Premise 1: Occurrence of correctable errors on a module indicates a higher probability of an occurrence of an uncorrectable error on the same module.
- Premise 2: Correctable errors degrade system performance.

The next section debunks these two premises.

Correctable Errors and Uncorrectable Errors

Traditionally, modules with correctable errors were quickly replaced in the field. Modules with only correctable errors were replaced based on the premise that correctable errors indicate a higher probability of an imminent uncorrectable error, resulting in system failure or downtime. Therefore, server vendors advised customers to replace any modules with correctable errors immediately. However, a review of field failures shows no correlation between correctable errors and uncorrectable errors.

Cisco conducted a multiple-year study of field memory failures on Cisco UCS servers. The data for this study came from 8,322 servers built with 163,850 memory modules on both Cisco UCS M3 and M4 servers running DDR3 and DDR4 memory. In this study, all data related to any module that generated a correctable error or an uncorrectable error was collected weekly. The count of correctable and uncorrectable errors for every module was plotted to see whether correctable errors occurred before the same memory module experienced an uncorrectable error. As shown in Figure 1 and Table 1, 296 memory errors occurred, with an average count of about 5.9 million correctable errors per week that never preceded an uncorrectable error. From this data, the following can be inferred:

- Even though several of these modules ran more than 10 weeks with a large number of correctable errors every week, they never had a single uncorrectable error during or after. This finding suggests that incidences of correctable errors are not a reliable indicator of subsequent uncorrectable errors.
- Six modules experienced an average of 66 uncorrectable errors per week, but had no prior correctable errors. This data suggests that there is no correlation between incidences of correctable errors and uncorrectable errors.

Figure 1. Correctable Errors Versus Uncorrectable Errors

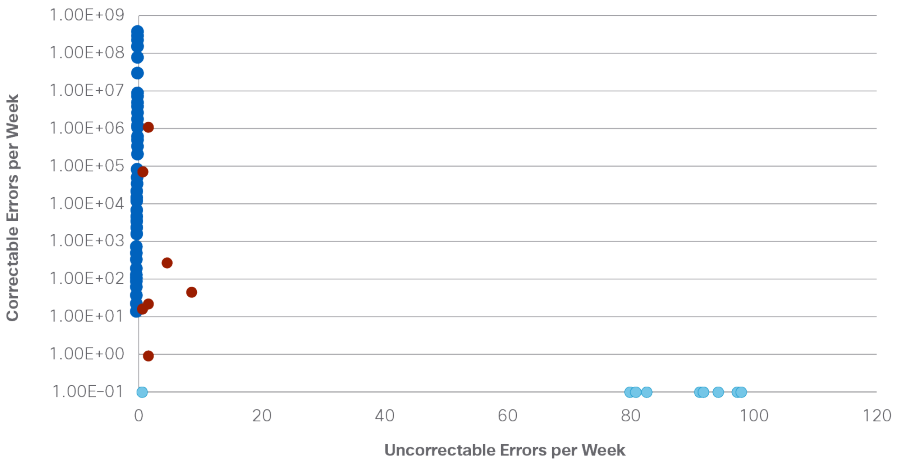

Table 1. Count of Correctable Errors Versus Uncorrectable Errors

Error Types	Count of Errors	Count of Modules	Average Error Count per Week
Correctable errors only	296	144	5,950,523
Uncorrectable errors only	16	6	66
Both correctable and uncorrectable errors	7	5	Correctable errors: 157,615 Uncorrectable errors: 3

In seven cases (see the red dots in Figure 1), modules experienced both correctable and uncorrectable errors during the same week. However, the average correctable error count on these modules was several orders of magnitude lower than on modules with correctable errors only (157,000 versus 5.9 million). Hence, this data debunks the premise that there is a strong correlation between correctable errors and uncorrectable errors.

Given that there is no correlation between correctable errors and uncorrectable errors based on this study, flagging a module with only correctable errors for immediate replacement to avoid uncorrectable errors may create unnecessary server disruptions without actually reducing the risk that an uncorrectable error will occur.

Correctable Errors and System Performance

To address the second premise, that modules with correctable errors degrade system performance and therefore should be replaced immediately, Cisco built servers with instrumented modules that generated a large number of correctable errors: for instance, in the range of 250,000 correctable errors in 24 hours. The industry-standard STREAM benchmark was used to measure performance on these instrumented servers and compare the results with those from servers with no correctable errors. The servers with correctable errors showed no performance degradation when compared to the servers with no errors. No degradation in performance was observed even in LINPACK and blended LINPACK and STREAM benchmark tests.

Future Enhancements to Cisco UCS Error Management

Based on the analysis of DIMMs returned from the field and internal validation testing, it is clear that error management policies should handle correctable errors in a different way than uncorrectable errors. Whereas modules with uncorrectable errors must be replaced immediately as they might cause systems to crash, immediate replacement of DIMM modules with only correctable errors is not recommended as systems can continue to operate normally after resetting of the errors. Future enhancements to error management policies will distinguish memory errors based on their locations in the DIMM module, and provide recommendations about the appropriate corrective actions to take, if necessary, to avoid unplanned outages.

Conclusion

Industry demands for greater capacity, greater bandwidth, and lower operating voltages lead to increased memory-error rates. Traditionally, the industry has treated correctable errors in the same way as uncorrectable errors, requiring the module to be replaced immediately upon alert. Given extensive research that correctable errors are not correlated with uncorrectable errors, and that correctable errors do not degrade system performance, the Cisco UCS team recommends against immediate replacement of modules with correctable errors. Customers who experience a degraded memory alert for correctable errors should upgrade to that latest UCSM 2.2(7), 3.1(1), or CIMC 2.0(9) for racks standalone mode. If the customer does not wish to upgrade, then reset the memory error and resume operation. Following this recommendation will avoid unnecessary server disruption.

Appendix A: How to Reset Memory Errors

Cisco UCS B-Series and C-Series Operating in UCSM 2.2 and 3.1

To reset memory-error counters on a Cisco UCS B-Series or C-Series server in UCSM 2.2 and 3.1, run the following script on the CLI:

```
ca-1-A# scope server 1/8
ca-1-A /chassis/server # reset-all-memory-errors
ca-1-A /chassis/server* # commit
```

Cisco UCS B-Series and C-Series Operating in UCSM 2.1

To reset memory-error counters on a Cisco UCS B-Series or C-Series server in UCSM 2.1, run the following script on the CLI:

```
Switch-A # scope server 1/1
Switch-A /chassis/server # scope memory-array 1
Switch-A /chassis/server/memory-array # scope dimm 2
Switch-A /chassis/server/memory-array/dimm #  reset-errors
```

Cisco UCS C-Series Rack Servers Operating in Standalone Mode

To reset memory-error counters on a Cisco UCS C-Series Rack Server operating in standalone mode, run the following script on the CLI:

```
C240-FCH092779J# scope reset-ecc
C240-FCH092779J /reset-ecc # set enabled yes
C240-FCH092779J /reset-ecc *# commit
```

Appendix B: Additional Resources

For additional information about memory, please refer to these resources:

- DIMMs: Reasons to Use Only Cisco Qualified Memory on Cisco UCS Servers

For additional information about Cisco UCS servers, please visit the Cisco UCS server webpage.

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**

Printed in USA C11-736116-03 04/16