



Hewlett Packard
Enterprise

How memory RAS technologies can enhance the uptime of HPE ProLiant servers

Reduce server crash rate by approximately 85%

Contents

Summary.....	2
Memory RAS technologies in HPE ProLiant scale-up servers.....	2
Why memory RAS is needed.....	5
Uncorrected memory device errors are the leading cause of server crashes.....	5
Memory RAS technologies reduce ASRs.....	6
A brief overview of SDDC, rank-sparing, and DDDC implementation.....	7
Cost trade-offs.....	8
HPE SmartMemory.....	9
Conclusion.....	9
Appendix.....	10

Summary

Memory RAS capabilities can reduce scale-up server crash rates by approximately 85%¹ and the annual service rate by 90%²

Memory device failures—if not corrected—can result in service events, or even server crashes. As modern servers implement ever larger memory arrays, the likelihood of a memory device failure increases, especially for scale-up³ servers due to their high memory capacity. To outpace this trend, HPE ProLiant servers provide an increasingly comprehensive suite of memory error detection and correction features—collectively called memory RAS (reliability, availability, serviceability).

It might surprise you to know that memory device failures are far and away the most frequent type of failure for scale-up servers. This paper reviews recent HPE engineering analyses which identify memory device failures as the #1 root cause of both crashes and service events for scale-up servers in which advanced memory protection technologies such as DDDC or SDDC with rank sparing, have not been implemented. However, when either of these technologies is implemented on a scale-up server, the annual crash rate (ACR) can be reduced by approximately 85%. Analysis data is also presented which shows that there is the opportunity to reduce the annual service rate (ASR) of your ProLiant servers by approximately 90% by implementing appropriate memory RAS technologies.

To help you select appropriate memory RAS capabilities of ProLiant servers based on Intel E5 or E7 processors, this paper provides descriptions of memory RAS technologies, their benefits, and the level of effort to implement them along with relative server pricing data. Also included is a discussion of HPE Memory Quarantine which is a technology implementation unique to HPE ProLiant servers to enhance the uptime of servers hosting virtual machines.

Overall, the data presented in this paper will help you make an informed decision about the most appropriate memory RAS technologies for your HPE ProLiant scale-up servers so that they can meet your demanding workload and data center service level requirements, especially for business-critical workloads.

Note

This paper focuses solely on memory RAS. It does not review the comprehensive suite of other RAS technologies found throughout the ProLiant portfolio.

Memory RAS technologies in HPE ProLiant scale-up servers

HPE ProLiant scale-up servers provide a comprehensive suite of advanced memory RAS technologies that help you to reduce system downtime due to memory failures. These memory RAS technologies have been introduced to outpace the increases in memory capacities of ProLiant servers.

The following descriptions provide an overview of the functionality of selected memory RAS technologies (please see the [Appendix](#) for a sample of ProLiant servers and the Intel processors that they use):

- **Error Correction Code (ECC):** handles minor amounts of data corruption via multi-bit error detection and single-bit error correction. Depending on the system, ECC can help identify failing DIMMs.
Availability: ECC is provided on all G7 and Gen8 ProLiant servers.
- **Advanced ECC:** detects and corrects some multi-bit errors. Advanced ECC can detect and correct up to 4-bit memory errors if all failed bits are on the same DRAM device on the DIMM, thus enabling continued memory operation.
Availability: Advanced ECC is provided on all G7 and Gen8 servers.
- **Single Device Data Correction (SDDC), also known as Single Device Disable Code:** detects and corrects multi-bit errors. It uses error checking and correction code to identify and disable a failed single DRAM device on a x4 DIMM. The disabled DRAM is removed from the memory map and its data is recovered onto a good spare DRAM device.
Availability: SDDC is provided on all G7 and Gen8 servers.

^{1,2} HP (now Hewlett Packard Enterprise) ISS R&D lab analyses and simulations performed during June 2012

³ Scale-up servers have four or more sockets available. Note that multiple processors are typically available per socket

- **Single Device Data Correction + 1 'bit' (SDDC+1):** a variation of SDDC, SDDC+1 can protect against the failure of an additional 'bit' in the same rank. This additional protection helps correct single-bit errors (e.g., due to cosmic rays) while a DIMM has entered single-device correction mode and is tagged for replacement.

Availability: DL580-G7, BL620-G7, and BL680-G7 provide SDDC+1.

- **Rank-sparing and its variant DIMM-sparing, or Online Memory Sparing or Online Spare with Advanced ECC Support:** provides protection against persistent DRAM failure. It tracks excessive number of correctable errors and copies the contents of an unhealthy rank to an available spare rank in advance of multi-bit or persistent single-bit failures that may result in future uncorrectable faults. It does not identify or disable individual failed DRAMs, but instead it disables the DIMM or rank. Since a DIMM or a rank is needed to perform sparing, this technique reduces the total amount of available memory by the amount of memory used for sparing. Sparing can only handle one failure per DIMM. DIMMs that are likely to receive a fatal/uncorrectable memory error are automatically removed from operation, resulting in less system downtime.

Rank-sparing is more efficient than DIMM-sparing since only a portion of the collection of DIMMs on a memory bus is set aside for memory protection. For example, in a memory bus configured with two dual-ranked DIMMs, rank-sparing sets aside one fourth of the memory capacity for sparing, whereas DIMM-sparing requires one half of memory capacity for sparing.

There is no performance penalty for rank-sparing during normal operation. Upon an error condition, the time it takes to copy the data from the failing rank to the spare rank is small.

Availability: Rank-sparing is provided on all G7 and Gen8 ProLiant servers.

- **Double Device Data Correction + 1 'bit' (DDDC+1), also known as Double-Chip Sparing:** a more robust and more efficient method of memory sparing. It can detect and correct single- and double-DRAM device errors for every x4 DIMM in the server. By reserving one DRAM device in each rank as a spare, and using the SDDC techniques on the remaining DRAMs, it ensures data availability after hardware failures with any x4 DRAM devices on any DIMM. In the unlikely occurrence of a second DRAM failure within a DIMM, the platform firmware alerts you to replace that DIMM gracefully without a system crash. The additional '1-bit' protection further helps correct single-bit errors (e.g., due to cosmic rays) while a DIMM has entered dual-device correction mode and is tagged for replacement. DDDC+1 provides the highest level of memory RAS protection with no performance impact and no reserved memory in Intel® Xeon® E7-based ProLiant-G7 servers; the full system memory is available for use.

Availability: DDDC+1 is available on ProLiant DL580-G7, DL980-G7, BL620-G7, and BL680-G7 servers.

The memory RAS technologies listed above represent just a portion of what is available on HPE ProLiant servers. Below is a sampling of other memory RAS technologies that might be of value to you for business-critical workloads and/or servers hosting significant numbers of virtual machines:

- **Memory Mirroring:** provides protection against uncorrectable memory errors that would otherwise result in system failure. In this mode, the system maintains two copies of all data. If an uncorrectable memory error occurs, the system automatically retrieves the good data from the mirrored (redundant) copy. The system continues to operate normally without any user intervention. By providing added redundancy in the memory sub-system, Memory Mirroring provides the greatest protection against memory failure not corrected by ECC, SDDC, DDDC, and online spare memory.

The performance impact for implementing Memory Mirroring is very small. While there is no READ performance impact and no WRITE performance impact in low memory traffic, WRITE performance impact is mostly hidden in heavy traffic.

Since Memory Mirroring consumes 50% of the system memory capacity, it is merited for server workloads that must receive the highest level of protection from memory device failures. You might want to consider Memory Mirroring for workloads which cannot have downtime and cannot risk waiting until scheduled downtime to replace degraded memory modules.

Availability: Memory Mirroring is available on selected G7 and all Gen8 servers.

- Lockstep, also known as x8 SDDC:** corrects a single x8 DRAM device failure on a DIMM. It extends SDDC capability from x4 DRAM devices to x8 devices by using two memory channels as a single-wide channel to transfer a longer data word. The long data word is transferred each time using 16 redundant bits to provide 8-bit error detection and 8-bit error correction to protect against a single DRAM failure. While Lockstep mode does not reduce the total amount of available memory, it has some performance impact to memory-intensive workloads. In this mode, similar to DDDC, the DIMMs in each paired memory channel must have identical HPE part numbers.

Availability: x8 SDDC is available on selected G7 and Gen8 servers.

- HPE Memory Quarantine:** Intel Xeon-E7’s Machine Check Architecture (MCA) Recovery, which is only implemented on HPE ProLiant servers, identifies memory regions containing uncorrectable hardware errors prior to data processing. It increases system availability by enabling the server and the operating system (or hypervisor) to work together so the server can recover from uncorrectable memory errors that would have otherwise caused a system crash. HPE Memory Quarantine isolates the bad memory location before it affects other data. It does this by using a patrol scrubber that constantly inspects the memory for errors (see figure 1), which provides early identification of failed components. If an error is found, the hardware attempts to correct it. If the hardware cannot correct the error, the platform firmware notifies the operating system. Then, the memory address is tagged as bad so that the operating system does not use this memory location.

HPE Memory Quarantine

Recovery from uncorrectable memory errors which may cause a system crash

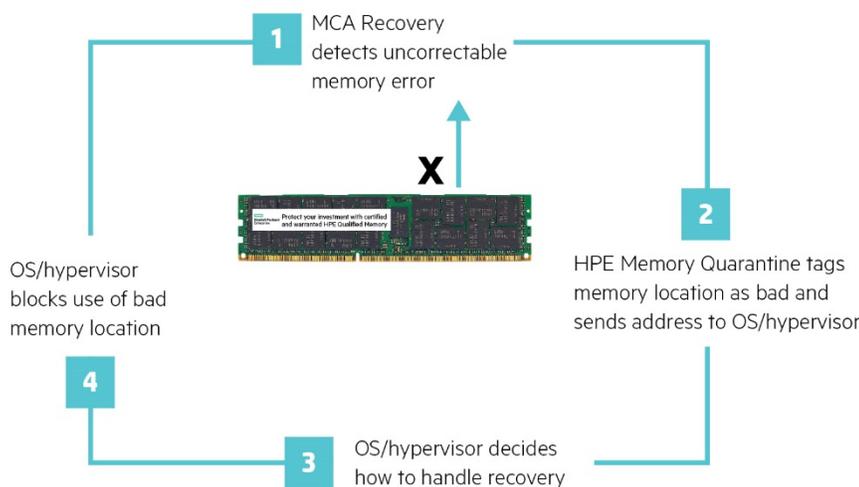


Figure 1: HPE Memory Quarantine limits the impact of uncorrectable memory errors

HPE Memory Quarantine can reduce the crash rate of hypervisors due to uncorrectable memory errors, thus making it ideal for use in servers that host a large number of virtual machines, and/or a small number of virtual machines hosting important workloads.

HPE Memory Quarantine also reduces service costs by identifying failing components that can be replaced during planned maintenance cycles, reducing the need for unplanned service, thus increasing server availability.

Availability: HPE Memory Quarantine is available on DL580-G7, BL680-G7, and BL620-G7 servers.

Note

The technologies cited above are not an exhaustive list of all memory RAS capabilities in ProLiant servers.

Why memory RAS is needed

When the ProLiant 1000 was announced in September 1993, it offered a maximum of 144 MB of system memory and only Advanced Error Checking and Correcting for memory failures. Then—like now—the root cause of memory device soft failures was from cosmic “rays” (which are actually the bare nuclei of a variety of atoms, and not an electromagnetic wave) and gamma rays (ultra-high energy electromagnetic waves). A soft error is a data error, but does not necessarily mean that the memory device has suffered a hard—physical—failure.

To pack more memory into a device, the size of each transistor has become smaller in successive generations of DRAMs. This silicon geometry compaction has resulted in the current high memory capacities within servers and other devices. When a cosmic ray or gamma ray now hits a device, it is much more likely to affect more than one memory cell, thus resulting in one or more soft errors.

The availability of higher memory capacity DIMMs for server system memory has supported the ongoing requirements of workloads that require large memory capacity and scale-up servers. HPE anticipates that for the foreseeable future, there will be demand for scale-up servers that offer ever larger system memory capacities. Among the workloads driving this are 1) hosting more and larger virtual machines per single processor socket, or single processor core; 2) in-memory analytics, such as SAP® HANA; and 3) large single-instance databases migrated off UNIX® platforms to x86. All of these workloads require a robust suite of memory RAS capabilities in the server to meet the operational requirements of the organization.

To stay ahead of these trends, Hewlett Packard Enterprise has continued to implement new memory RAS technologies to keep overall system crash rates and service events⁴ at low levels. However, ongoing conversations with customers indicate that many haven’t taken advantage of the full suite of memory RAS capabilities in their ProLiant Servers, and thereby are not experiencing the higher system availability that could be attained. To help achieve the most uptime from your HPE ProLiant server, this paper provides compelling data that demonstrates the value of using memory RAS capabilities.

It is important to note that by avoiding a critical failure, a system crash can be avoided. Failed memory devices are replaced as part of periodic service. Also, the memory RAS technologies can detect a device on a DIMM that has had numerous soft errors, and recommend replacing it before it has a hard failure.

Uncorrected memory device errors are the leading cause of server crashes

HPE engineers continually analyze root causes of server system crashes. Their recent findings may surprise you!

Figure 2 presents relative ACR data for scale-up servers in which neither memory sparing nor mirroring were implemented, but SDDC+1 was used. It may surprise you to see that DIMMs are the largest contributing factor, followed by “board related” group which represents multiple types of system elements that have failed. Clearly, the top two groupings represent the overwhelming majority of root causes of server crashes. (Note that the storage sub-system in these servers normally benefits from RAID, and HDD failures do not contribute to system crashes.)

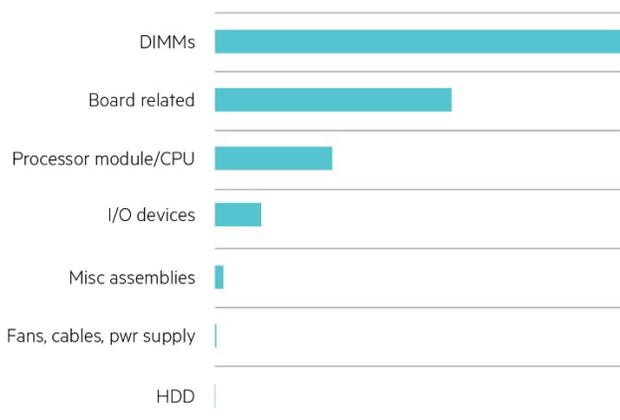


Figure 2: Pareto distribution of components of ACR with SDDC+1 implemented

No memory sparing or mirroring was utilized. Note that lower values in the chart are better.

⁴ A system crash refers to an event in which the server is not able to function and must be rebooted and/or repaired. A service event is one in which the server identifies the need for repair, but continues to function, albeit with diminished functionality and/or performance

Figure 3 shows that when DDDC+1 is used, the ACR due to memory drops approximately 85% compared to the system using SDDC+1.

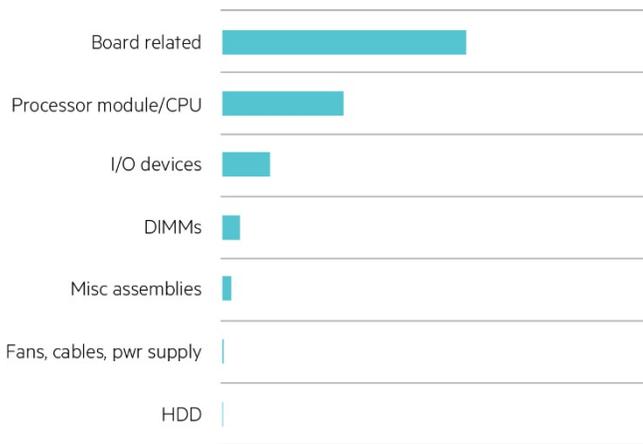


Figure 3: Pareto distribution of components of ACR with DDDC+1 implemented

No memory sparing or mirroring was utilized. Note that lower values in the chart are better.

The value of these relative comparisons is to help you know which memory RAS configurations will provide an ACR that is an appropriate match to the workloads on the server. In general, server crashes due to memory failures are relatively rarely observed because the default configurations of ProLiant servers provide a comprehensive suite of memory RAS technologies. However, data that is truly business-critical, or a server that must absolutely have crashes minimized, may merit the use of DDDC+1. Alternatively, you might want to consider use of Memory Mirroring, which was described in a prior section.

Memory RAS technologies reduce ASRs

The prior section identified the impact of different memory RAS technologies on server ACRs. The combination of the unpredictability of server crashes along with their severe business impact makes them the most important event to avoid.

HPE engineering analysis also demonstrates that ASRs decline when advanced memory RAS technologies are utilized. While service events are more manageable, they still consume significant amount of personnel time and may require server downtime to resolve. This section explores the impact of memory RAS technology on ASR.

In the prior section, anonymized crash data from servers in use with customers was analyzed. In this section, customer service data was combined with computer simulations to identify the impact on ASR of different memory RAS technologies. The reference server in this analysis was a four-processor HPE ProLiant scale-up server configured with 64 dual-rank x4 DIMMs (2 DIMMs per channel and 2 ranks per DIMM.)

As seen in figure 4, the ASR data was normalized to the DDDC scenario in which neither sparing nor Memory Mirroring is used. When only SDDC is used, the ASR is more than 10x the DDDC baseline. However, when SDDC is combined with a spare rank, an ASR comparable to that of DDDC is attained.

In terms of servers to implement, if an Intel Xeon E5-based server uses SDDC without spare ranks, it will have an ASR that is more than 10x that of a Xeon E7-based server which uses the more capable DDDC+1 technology. However, it is possible to greatly reduce the ASR for the E5-based server by adding rank-sparing to complement SDDC. But even when SDDC + spare rank is implemented, the E5-based server has an ASR that is 30% more than that of the E7-based server.

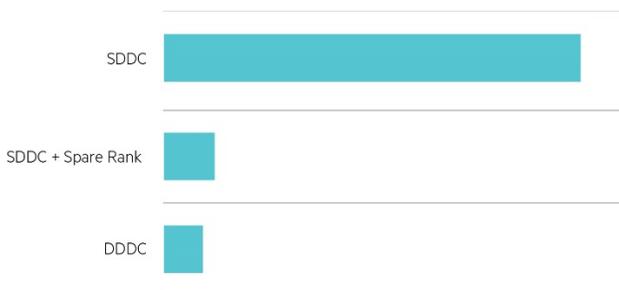


Figure 4: Relative ASR

Note that lower values in the chart are better.

In summary, you can reduce both ACR and ASR due to memory failures by either using an E7-based ProLiant server and benefit from its DDDC+1 capabilities, or use an E5-based ProLiant server and implement SDDC along with additional memory for sparing.

A brief overview of SDDC, rank-sparing, and DDDC implementation

For HPE ProLiant G7 servers, the HPE ROM-Based Setup Utility (RBSU) is an embedded configuration utility that performs a wide range of configuration activities, including configuring memory options. For more information on RBSU, see the HPE ROM-Based Setup Utility User Guide on the documentation CD.

For HPE ProLiant Gen8 servers and server blades, no media kit is included with the servers. It is instead replaced by Intelligent Provisioning. For more information, visit hp.com/go/intelligentprovisioning.

The balance of this section provides a brief overview of the level of administrative effort required to implement selected memory RAS capabilities.

- ECC, Advanced ECC, SDDC
 - Each are part of the default system configuration, and thus don't require any incremental administrative effort to invoke
- Rank-sparing
 - To implement rank-sparing, a portion of memory must be reserved for this purpose. Consequently, total available system memory is reduced by this reserved portion. For example, rank-sparing is typically implemented by setting aside one of every four or one of every eight DIMM ranks as a spare. In the latter scenario, the spare rank occupies 12.5% of capacity. Less frequently, sparing is implemented in one of every four DIMM ranks, thus consuming 25% of capacity. If a high level of sparing is thought to be needed, you might instead want to consider implementing Memory Mirroring. Note that rank-sparing cannot be enabled concurrently with Memory Mirroring.
 - There are some differences between the DRAM device sparing support across EX- and EP-based servers (i.e., E7 and E5):
 - EX supports SDDC+1 for x4 DRAM devices; EP supports SDDC
 - EX supports SDDC+1 for x8 DRAM devices; EP supports SDDC for x4 DRAM devices
 - Once the sparing is selected in the RBSU, system ROM configures and allocates spare rank(s) accordingly.
 - As part of the operating system boot up, the portion of physical memory allocated as spare memory is not visible by the operating system or hypervisor.
 - If at some point in the future it is determined that rank-sparing is no longer needed, the memory used for sparing can be reallocated using RBSU.
- DDDC+1
 - DDDC+1 is only supported on Intel Xeon E7 family processors using x4 DIMMs; x8 DIMMs are not supported in this mode.
 - More recently produced E7-based ProLiant-G7 servers ship with DDDC+1 enabled. Older ProLiant G7 servers require a firmware upgrade to support DDDC. No additional administrative effort is required.
 - DDDC is not available on E5-based servers.
- HPE Memory Quarantine
 - Included in the default system configuration.

Cost trade-offs

As seen in figures 2, 3, and 4, you can achieve fairly comparable ACR and ASR levels by purchasing an HPE ProLiant Server based on an Intel E7 processor and utilize DDDC+1, or alternatively purchasing an E5-based ProLiant Server and installing extra memory for spare ranks. As a guide, this section provides a comparison of the costs a ProLiant DL580 E7-based server configured with 256 GB of RAM versus a DL560 E5-based server configured with 256 GB of RAM plus 1:4 spare ranking, for a total configured RAM of 320 GB. Each of the servers was configured with four processors of comparable performance.

Table 1 presents a hardware capital cost comparison, with U.S. dollar-based prices normalized to the ProLiant DL580. The DL580 has a higher base price, but since it includes DDDC+1, there is no need to purchase extra RAM for rank sparing. With the spare ranks included in the pricing, the DL560 Gen8 capital cost is approximately 6% less than the DL580 G7. Also of importance is that the DL560 Gen8 is projected to use 23% less electrical power than the DL580 G7. The specific configurations used for this calculation are presented in table 2.

Table 1. Hardware capital cost comparison

	DL580 G7	DL560 Gen8
Configured RAM (total)	256 GB	256 GB + 64 GB
Server price including RAM	1.00*	-0.94**

* U.S. dollar-based price has been normalized to 1.00

**The DL560 Gen8 is approximately 94% of the price of the DL580 G7 when it is configured with 256 GB of RAM plus 64 GB for rank sparing

Table 2. Configurations used in hardware capital cost comparison

DL580 G7:

QUANTITY	MODEL #	DESCRIPTION
0		HPE DL580R07 (E7) CTO Server [#1]
1	643086-B22	HPE DL580R07 (E7) Br CTO Server
1	643086-B22 ABA	U.S. - English localization
1	643067-L21	HPE E7-4870 DL580 G7 2P FIO Kit
2	643067-B21	HPE E7-4870 DL580 G7 Kit
2	643067-B21 OD1	Factory integrated
4	644172-B21	HPE DL580G7/DL980G7 (E7) Memory Cartridge
4	644172-B21 OD1	Factory integrated
32	604506-B21	HPE 8GB 2Rx4 PC3L-10600R-9 Kit
32	604506-B21 OD1	Factory integrated
2	578322-B21	HPE 1200W CS Plat Ht Plg Pwr Supply Kit
2	578322-B21 OD1	Factory integrated
1	U4617E	HPE Install ProLiant DL58x Service
1	U2268E	HPE 3y 4h 24x7 DL58x ProCare Service
1	TC278AAE	HPE Insight Control ML/DL/BL Bundle E-LTU

DL560 Gen8:

QUANTITY	MODEL #	DESCRIPTION
0		HPE DL560 Gen8 SFF CTO Chassis [#1]
1	686792-B21	HPE DL560 Gen8 CTO Server
1	686792-B21 ABA	U.S. - English localization
1	686843-L21	HPE E5-4650 DL560 Gen8 FIO Kit
3	686843-B21	HPE E5-4650 DL560 Gen8 Kit
3	686843-B21 OD1	Factory integrated
32	647897-B21	HPE 8GB 2Rx4 PC3L-10600R-9 Kit
32	647897-B21 OD1	Factory integrated
16	647893-B21	HPE 4GB 1Rx4 PC3L-10600R-9 Kit
16	647893-B21 OD1	Factory integrated
1	684208-B21	HPE Ethernet 1GbE 4P 331FLR FIO Adptr
1	656364-B21	HPE 1200W CS Plat PL Ht Plg Pwr Supply Kit
1	656364-B21 OD1	Factory integrated
1	339773-B21	HPE Online Spare FIO Memory
1	U6H58E	HPE Install DL560 Service
1	U6H10E	HPE 3y 4h 24x7 DL560 ProCare Service
1	TC278AAE	HPE Insight Control ML/DL/BL Bundle E-LTU

HPE SmartMemory

As part of the holistic memory RAS technology, HPE SmartMemory plays an import role in reducing the incidence rate of memory failures as well as improving performance, quality, manageability, and efficiency while reducing downtime and energy costs.

HPE SmartMemory is included in each Gen8 ProLiant Server. Hewlett Packard Enterprise works closely with leading memory device manufacturers in the development, qualification and production testing of DIMMs to ensure that they meet the tighter tolerances specified by HPE. For example, voltage tolerance testing is used to validate that a narrower, lower voltage range can be used, resulting in reduced energy consumption. Noise margin testing enables HPE to support a higher memory bus bit rate.

Conclusion

HPE ProLiant servers include a comprehensive suite of memory RAS capabilities that markedly reduce the crash rate and service rate due to memory failures. As server system memory increases, it will be important to make use of the memory RAS capabilities in your ProLiant servers to avoid rapid increases in system crashes and service calls due to memory failures.

Appendix

Table 3 provides examples of selected ProLiant G7 and Gen8 servers and their use of Intel Xeon processors. Note that in this document the Intel Xeon E5 series is also referred to as EP, and the Intel E7 series is also referred to as EX.

Table 3. HPE ProLiant servers using Intel Xeon processors

	DL380P	BL460C	DL560	DL580	BL660C	BL680C	DL980
ProLiant Gen	Gen8	Gen8	Gen8	G7	Gen8	G7	G7
Processor	Intel Xeon E5-2600	Intel Xeon E5-2600	Intel Xeon E5-4600	Intel Xeon E7-4800 and E7-7500 series	Intel Xeon E5-4600	Intel Xeon E7-4800 and E7-8867L	Intel Xeon E7 7500 series



Sign up for updates

★ Rate this document



© Copyright 2013, 2016 Hewlett Packard Enterprise Development LP. The information contained herein is subject to change without notice. The only warranties for Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.

Intel Xeon is a trademark of Intel Corporation in the U.S. and other countries. SAP is the trademark or registered trademark of SAP SE in Germany and in several other countries. UNIX is a registered trademark of The Open Group.

4AA4-3490ENW, February 2016, Rev. 1