Using Memory Mirroring and Address Range Mirroring in VMware ESXi on Lenovo ThinkSystem Servers

Planning / Implementation

Introduction to Memory Mirroring

Servers based on the Intel Xeon Scalable processor family support a Reliability Availability Serviceability (RAS) feature called Memory Mirroring. Memory Mirroring allows users to configure the memory in a highly reliable mode when memory component is affected by uncorrectable fault, so that in the event of a DIMM failure the server will keep on running. It provides full memory redundancy while reducing the total system memory capacity in half. Memory channels are grouped in pairs with each channel receiving the same data. If an uncorrectable fault occurs, the memory access controller switches from the DIMMs on the primary channel to the DIMMs on the mirrored channel.

The workflow of Memory Mirroring data writes and reads are shown in the following figure.

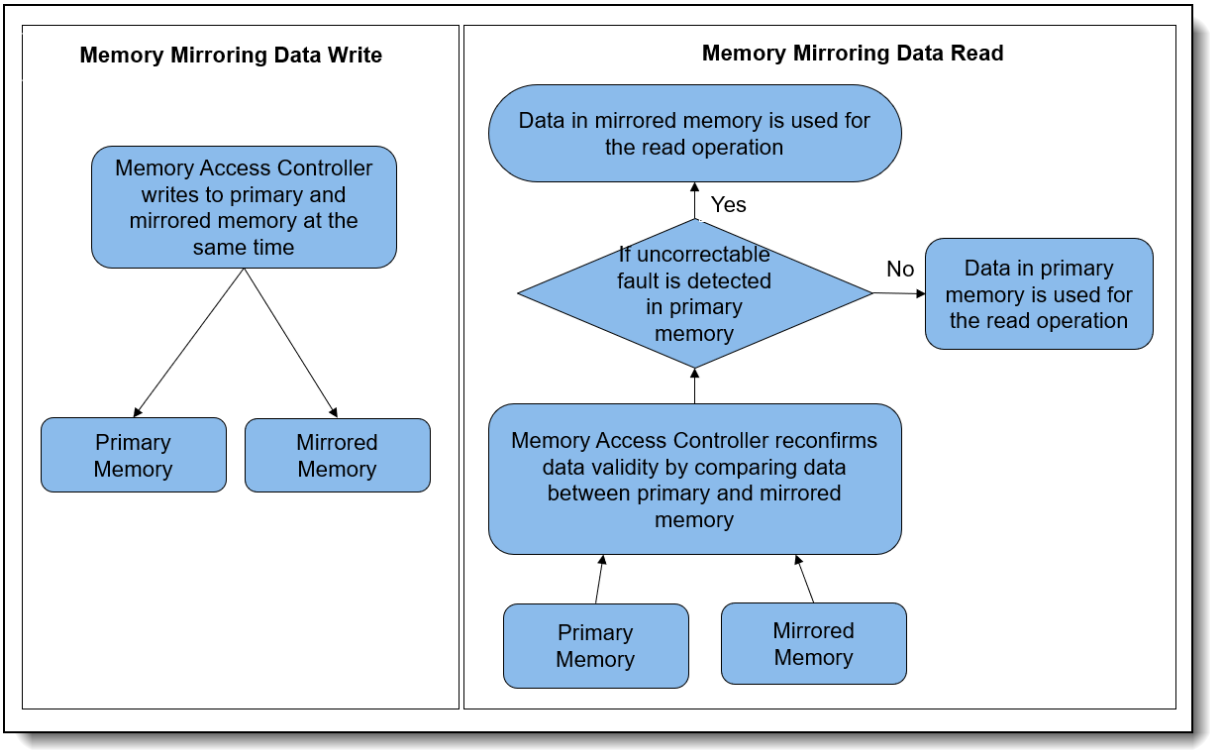

Figure 1. Workflow of Memory Mirroring data writes and reads

ThinkSystem servers with any Intel Xeon Scalable processor can support the Memory Mirroring feature. VMware ESXi supports the Memory Mirroring feature and the memory scheduler can put the memory pages consumed by critical services on reliable memory regions.

Memory mirroring reduces the maximum available memory by half of the installed memory. For example, if the server has 128 GB of installed memory, only 64 GB of addressable memory is available for ESXi when memory mirroring is enabled.

An illustration of Memory Mirroring is shown in the following figure.

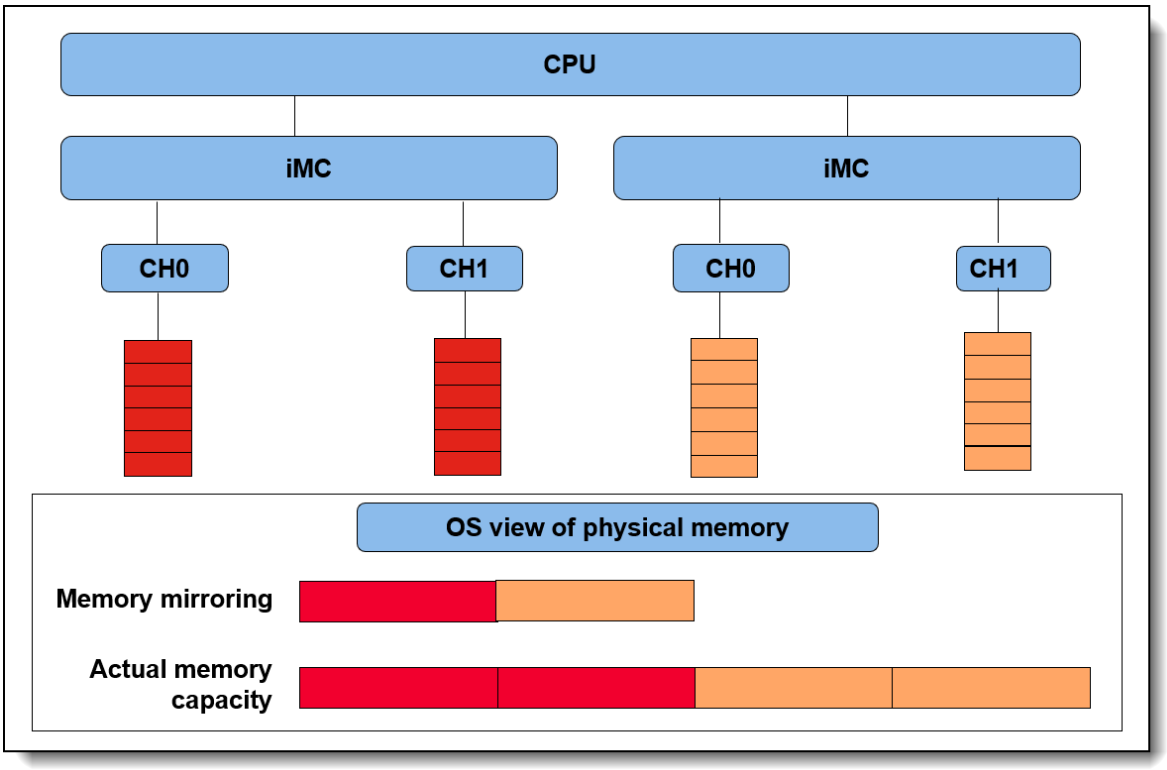

Figure 2. Memory Mirroring

Introduction to Address Range Mirroring

The Intel Xeon Gold and Platinum processors also offer support for partial memory mirroring which also called Address Range Mirroring. This feature allows greater granularity in selecting how much memory is dedicated for redundancy and it can reduce the amount of memory reserved for redundancy.

When Address Range Mirroring is used, the platform allows customer to specify a subset of total available memory for mirroring. This capability allows customers to make an appropriate trade-off between non-mirrored memory range and mirrored memory range, thus optimizing total available memory while keeping highly reliable memory range (the mirrored portion of the address space) available for mission-critical workloads and kernel space.

Lenovo ThinkSystem servers with Platinum or Gold processors can support Address Range Mirroring feature. VMware ESXi also support Address Range Mirroring feature and memory scheduler will do its best at putting all critical code and data in reliable memory. VMware refers to memory that is enabled for mirroring as *reliable memory*.

Address Range Mirroring can reduce the amount of memory reserved for redundancy by specify the desired subset of memory to mirror. For example, if the server has 128 GB of installed memory and mirror 25% of memory, 96 GB of addressable memory is available for ESXi when Address Range Mirroring is used. An illustration of Address Range Mirroring is shown below.

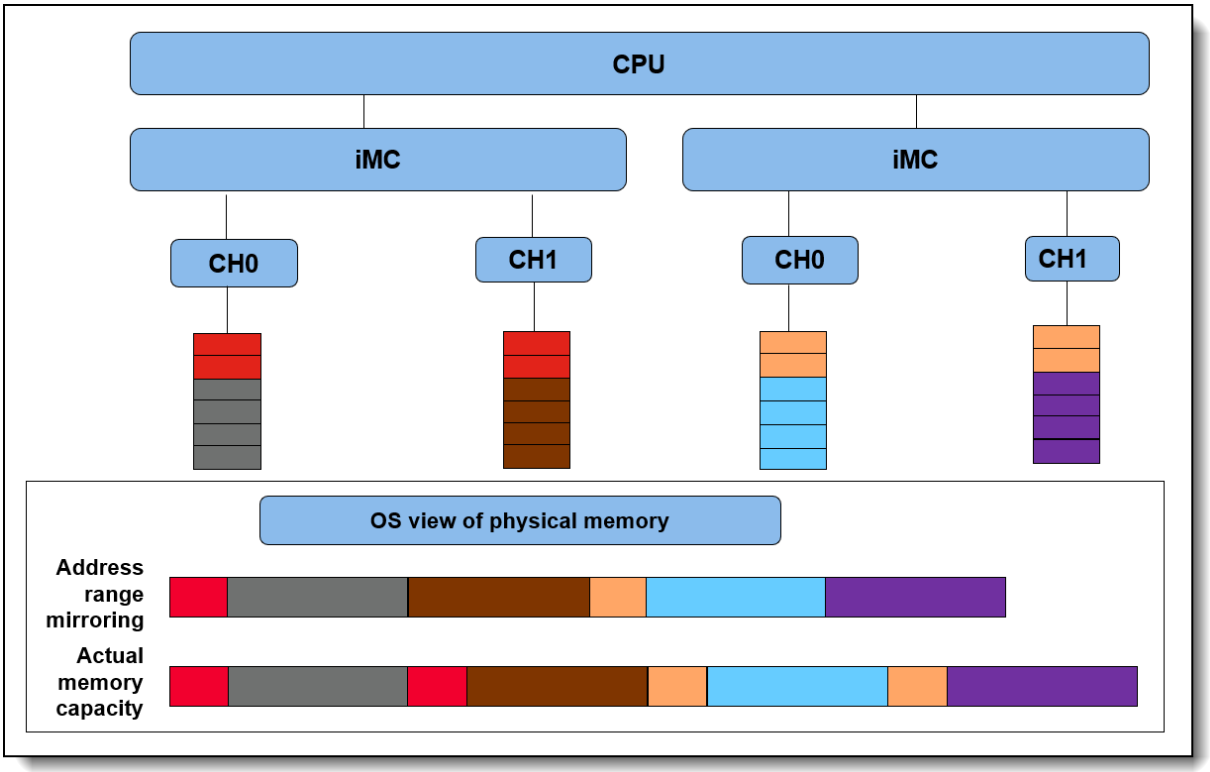

Figure 3. Address Range Mirroring

Address Range Mirroring offers the following benefits for customers:

- Provides greater granularity to memory mirroring by allowing customer to determine a range of memory addresses to be mirrored and leaving the rest of the memory in non-mirror mode.
- Reduces the amount of memory reserved for redundancy.
- Optimizes total available memory while keeping highly reliable memory range available for mission-critical workloads and kernel space.

vSphere support

Memory Mirroring is supported by all Intel Xeon Scalable processors, starting from 1st Gen processors. Address Range Mirroring is only supported by Intel Xeon Platinum processors and Intel Xeon Gold processors. The following table lists the Intel Xeon CPUs that support Memory Mirroring or Address Range Mirroring.

Table 1. Processor support of Memory Mirroring and Address Range Mirroring (all generations)

Mirroring mode	Intel Xeon Bronze Processors	Intel Xeon Silver Processors	Intel Xeon Gold Processors	Intel Xeon Platinum Processors
Memory Mirroring	Support	Support	Support	Support
Address Range Mirroring	No support	No support	Support	Support

Memory Mirroring and Address Range Mirroring are supported in vSphere ESXi 5.5 and later versions. If the server platform supports Memory Mirroring or Address Range Mirroring feature, ESXi can put the memory pages consumed by critical services in the mirrored regions.

Reliable memory: VMware ESXi refers to memory that is mirrored as *reliable memory*.

If at any point of time the system has insufficient reliable memory, ESXi falls back to allocating regular memory. At that point, using the reliable memory for critical services is a best effort.

As a minimum, we recommend booting ESXi with 3GB of reliable memory. If the amount of reliable memory on a system is too small to contain all the critical services at boot time, the host might hang or PSOD. To guarantee that all the critical services remain in reliable memory, it is recommended not to exhaust the reliable memory. In other words, configuring virtual machines with more reliable memory than the host capacity is not recommended.

vmkernel and monitor are categorized as priority 0 so the memory pages consumed by them are on high priority to be put in the reliable memory area. Some system processes running on ESXi userworld are marked as memory reliable and they are categorized as priority 1. Therefore, they are secondary priority to be put on reliable memory regions.

It's worth noting that more than the kernel can use this feature. We can also place the memory pages consumed by virtual machine (VM) on reliable memory area to protect the VMs from memory failure. They are categorized as priority 2, so they are on thirdly prioritized to be put in reliable memory area.

Note that even though vSphere ESX supports reliable memory, vSphere Distributed Resource Scheduler (DRS) does not support for reliable memory and DRS is a feature included in the vSphere Enterprise Plus.

DIMM installation

In Memory Mirroring mode, each memory module in a pair must be identical in size and architecture. The channels are grouped in pairs with each channel receiving the same data. One channel is used as a backup of the other, which provides redundancy. If a failure occurs, the memory access controller switches from the DIMMs on the primary channel to the DIMMs on the backup channel. The DIMM installation order for memory mirroring varies based on the number of processors and DIMMs installed in the server.

Address Range Mirroring is a sub-function of Memory Mirroring, so it requires the same memory installation rules and order as Memory Mirroring.

Follow the rules below when installing memory modules in Mirroring Mode:

- DIMMS are installed in pairs for each processor.
- All memory modules to be installed must be of the same type with the same capacity, frequency, voltage, and ranks.

- Mirroring can be configured across channels in the same iMC, and the total TruDDR5 memory size of the primary and secondary channels must be the same.
- 9x4 RDIMMs do not support mirroring mode.
- Partial Memory Mirroring is a sub-function of memory mirroring. It requires following the memory installation order of memory mirroring mode.

DIMM installation order for Mirroring Mode with one processor

Table 2 shows the sequence of populating memory modules for mirroring mode when only one processor is installed on ThinkSystem SR650 V3.

Table 2. Installation order - Mirroring mode with one processor on SR650 V3

	Processor 1															
Total DIMMs	**16**	**15**	**14**	**13**	**12**	**11**	**10**	**9**		**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
8 DIMMs	16		14		12		10				7		5		3		1
16 DIMMs	16	15	14	13	12	11	10	9		8	7	6	5	4	3	2	1

DIMM installation order for Mirroring Mode with two processors

The following table shows the sequence of populating memory modules for mirroring mode when two processors are installed on Lenovo ThinkSystem SR650 V3.

Table 3. Installation order - Mirroring mode with two processors on SR650 V3

	Processor 1																
Total DIMMs	**16**	**15**	**14**	**13**	**12**	**11**	**10**	**9**		**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
16 DIMMs	16		14		12		10				7		5		3		1
32 DIMMs	16	15	14	13	12	11	10	9		8	7	6	5	4	3	2	1
	Processor 2																
Total DIMMs	**32**	**31**	**30**	**29**	**28**	**27**	**26**	**25**		**24**	**23**	**22**	**21**	**20**	**19**	**18**	**17**
16 DIMMs	32		30		28		26				23		21		19		17
32 DIMMs	32	31	30	29	28	27	26	25		24	23	22	21	20	19	18	17

Refer to the following Lenovo Pubs site for more information on installing DIMMs correctly in Mirroring mode on Lenovo ThinkSystem servers:
https://pubs.lenovo.com/

For example, you can find the Memory Mirroring mode installation order for the Lenovo ThinkSystem SR650 V3 at the following web page:
https://pubs.lenovo.com/sr650-v3/memory_module_installation_order_mirroring

Server configuration

To use Memory Mirroring and Address Range Mirroring, the server configuration must meet the following requirement:

1. Lenovo ThinkSystem servers with Intel Xeon processors can support Memory Mirroring.
2. Lenovo ThinkSystem servers with Intel Xeon Platinum Processors or Intel Xeon Gold Processors can support Address Range Mirroring feature.
3. 9x4 Dual In-line Memory Modules do not support Memory Mirroring and Address Range Mirroring.
4. “ADDDC Sparing” should be disabled in UEFI settings if you want to use the Memory Mirroring or Partial Memory Mirroring features. When ADDDC Sparing is enabled, both Full Mirroring and Partial Mirroring setup options get grayed out and mirroring feature cannot be enabled.

Server UEFI settings for Memory Mirroring

In UEFI, Memory Mirroring is referred to as Full Mirror.

When memory is configured in full mirror mode, it provides full memory redundancy while reducing the total system memory capacity in half. Primary memory and mirrored memory are receiving the same data. If an uncorrectable error is detected in Primary memory, data in mirrored memory will be used for the read operation and the server will keep on running. Memory Mirroring is transparent to the OS.

Adaptive Double Device Data Correction (ADDDC) Sparing is another memory Reliability Availability Serviceability feature that is deployed at runtime to dynamically map out the failing DRAM device and continue to provide SDDC ECC coverage on the DIMM, translating to longer DIMM longevity. The operation occurs at the fine granularity of DRAM Bank and Rank to have minimal impact on the overall system performance. It's disabled by default in Lenovo ThinkSystem UEFI settings. Please note that ADDDC Sparing will not take effect if memory is set to full mirror mode or partial mirror mode. So, if we want to use memory Mirroring or Partial Memory Mirroring features, we need to disable “ADDDC Sparing” in UEFI settings.

The following are steps for configuring Memory Mirroring in UEFI Settings on Lenovo ThinkSystem server SR650 V3.

1. Power on Lenovo ThinkSystem server SR650 V3 and then press F1 to enter System Setup, go to **System Settings** > **Memory** page and make sure that **ADDDC Sparing** is set to **Disabled** as shown in the following figure.

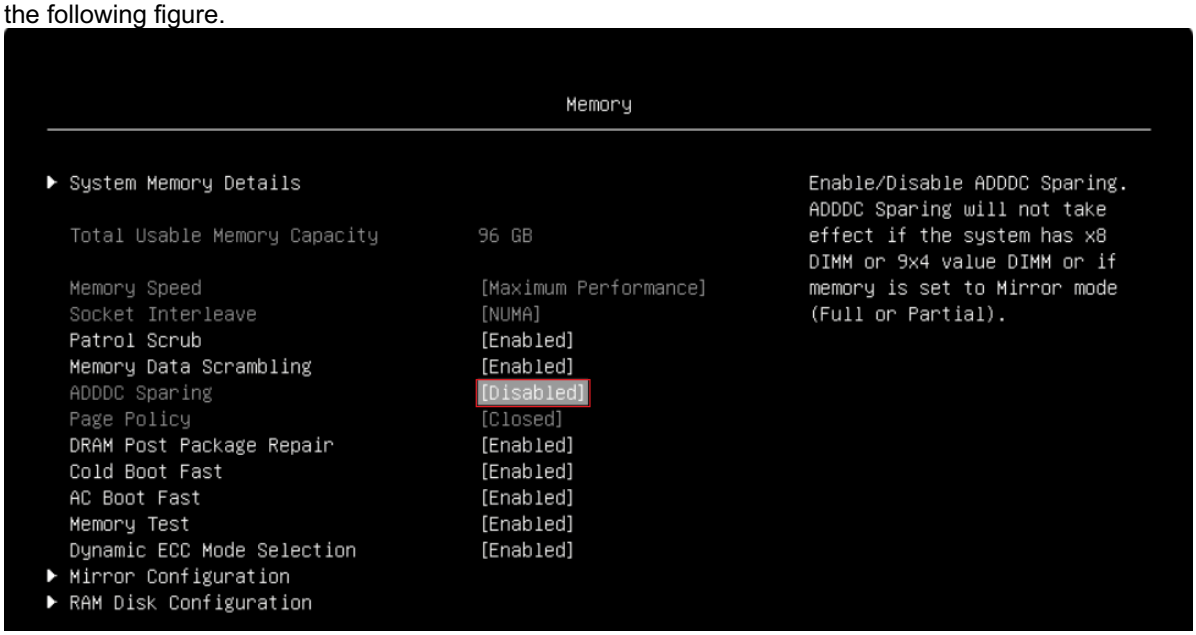

Figure 4. Disable ADDDC Sparing in UEFI Settings

When **ADDDC Sparing** is set to **Enabled**, both Full Mirroring and Partial Mirroring setup options get grayed out and mirroring feature cannot be enabled as shown in the following figure.

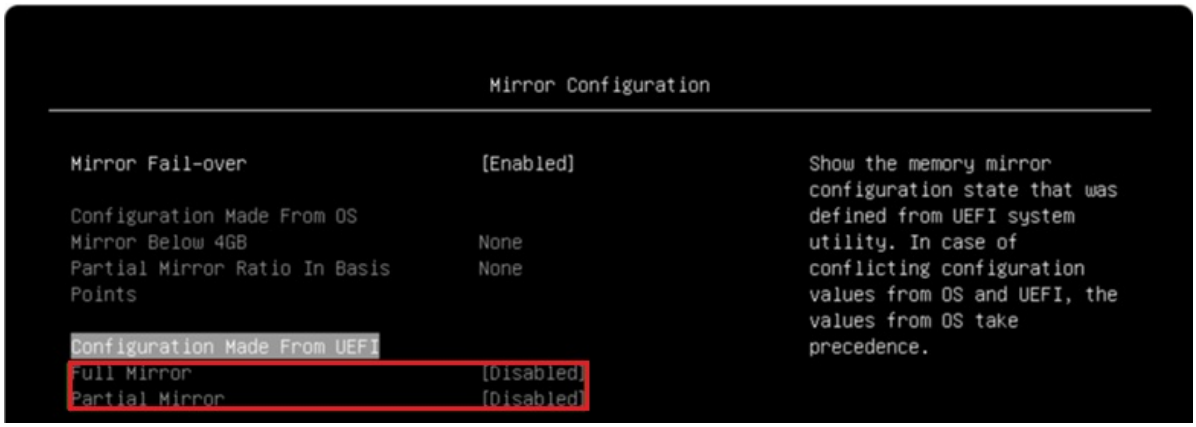

Figure 5. Full Mirror and Partial Mirror cannot be enabled when ADDDC is enabled

2. Go to **System Settings > Memory > Mirror Configuration** page, you can set **Mirror Fail-over** for Mirror Configuration. When **Mirror Fail-over** is Enabled, a persistent memory uncorrectable error will trigger mirror failover. When **Mirror Fail-over** is disabled, Lenovo UEFI will skip the mirror failover even when a persistent uncorrectable error occurs. The default setting of **Mirror Fail-over** is **Enabled** as shown in the following figure.

Figure 6. Mirror Fail-over configuration in UEFI Settings

3. On **System Settings > Memory > Mirror Configuration** page, enable **Full Mirror** as shown in the figure below.

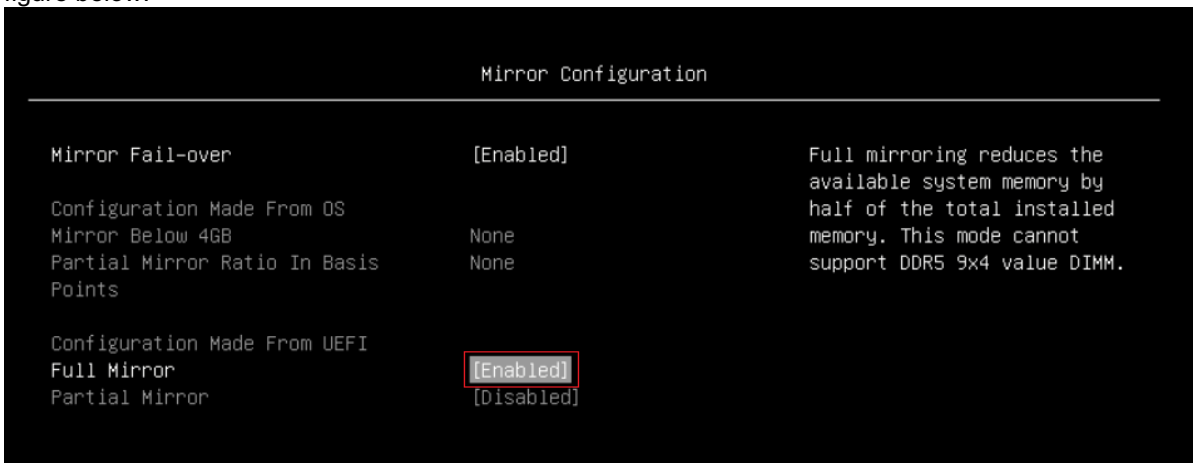

Figure 7. Enable Full Mirror in UEFI Settings

4. Save Settings and reboot host to make full mirror configuration take effect.

5. After configured Full Mirror in UEFI settings, the splash screen displays a total of 128 GB memory detected, Mirrored mode enabled, usable capacity 64 GB as shown in the figure below.

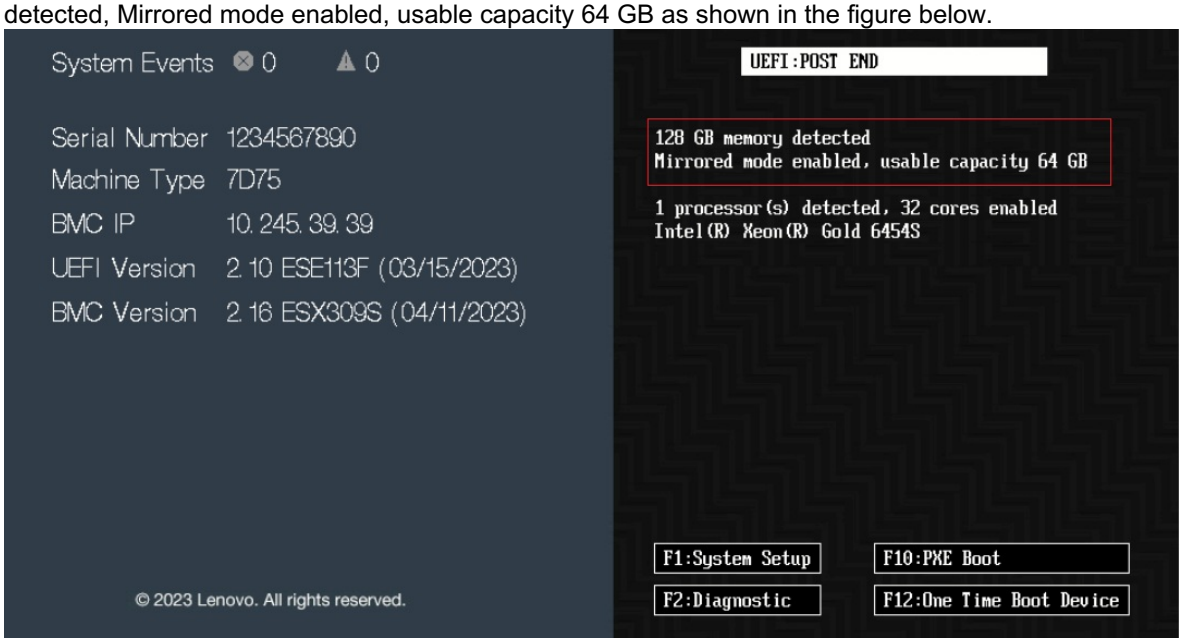

Figure 8. Splash screen with full mirror

6. Check Memory Map under EFI shell. The current memory map can be shown via the **memmap** If the full mirror mode is enabled, the memory map will be changed with mirrored size reduction. The top memory address is `0X207fffffff` (130 GB) in independent mode as shown in Figure 9, and the top memory address is `0x107fffffff` (66 GB) in full mirror mode as shown in Figure 10. As there's a 2GB MMIO size under 4GB address space, the memory size in independent mode is 128 GB (130 GB – 2 GB) and memory size in full mirror mode is 64 GB (66 GB – 2 GB).

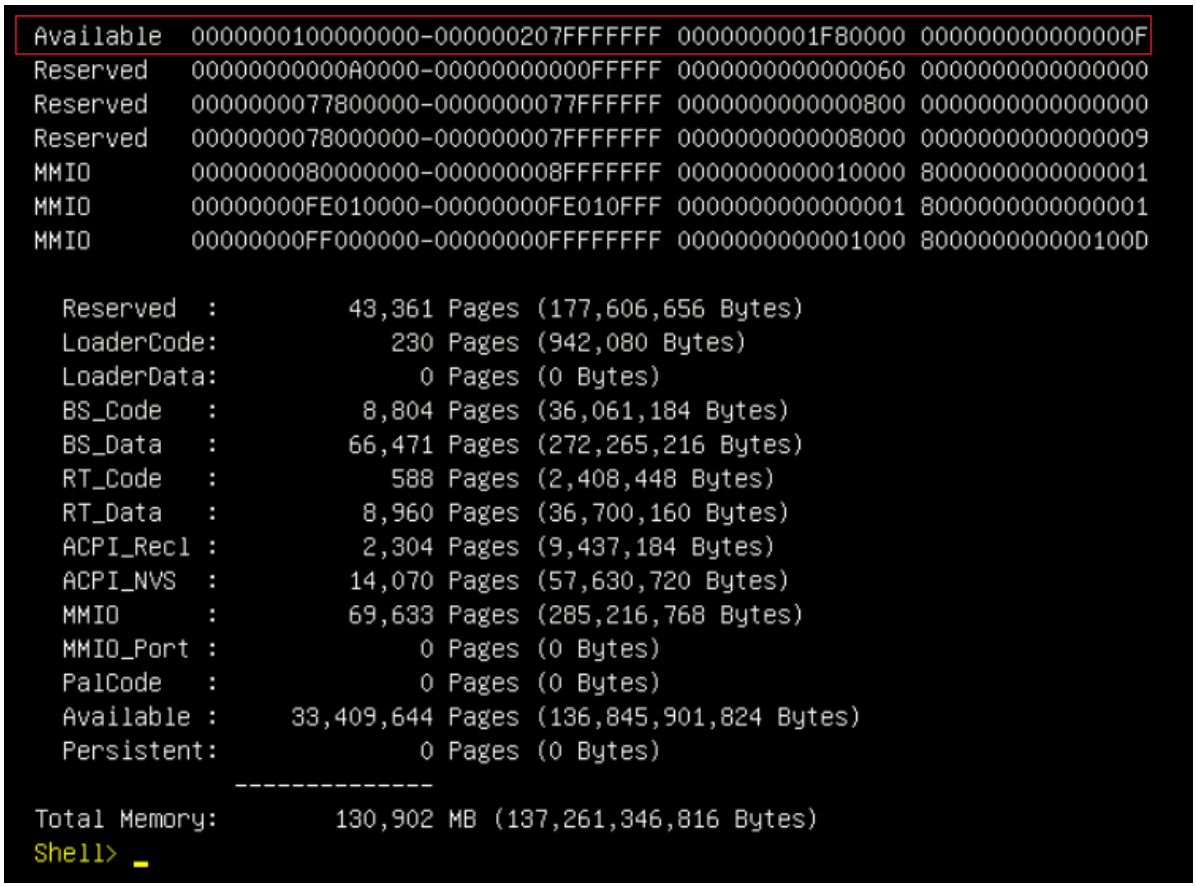

Figure 9. Independent Mode Memory Map

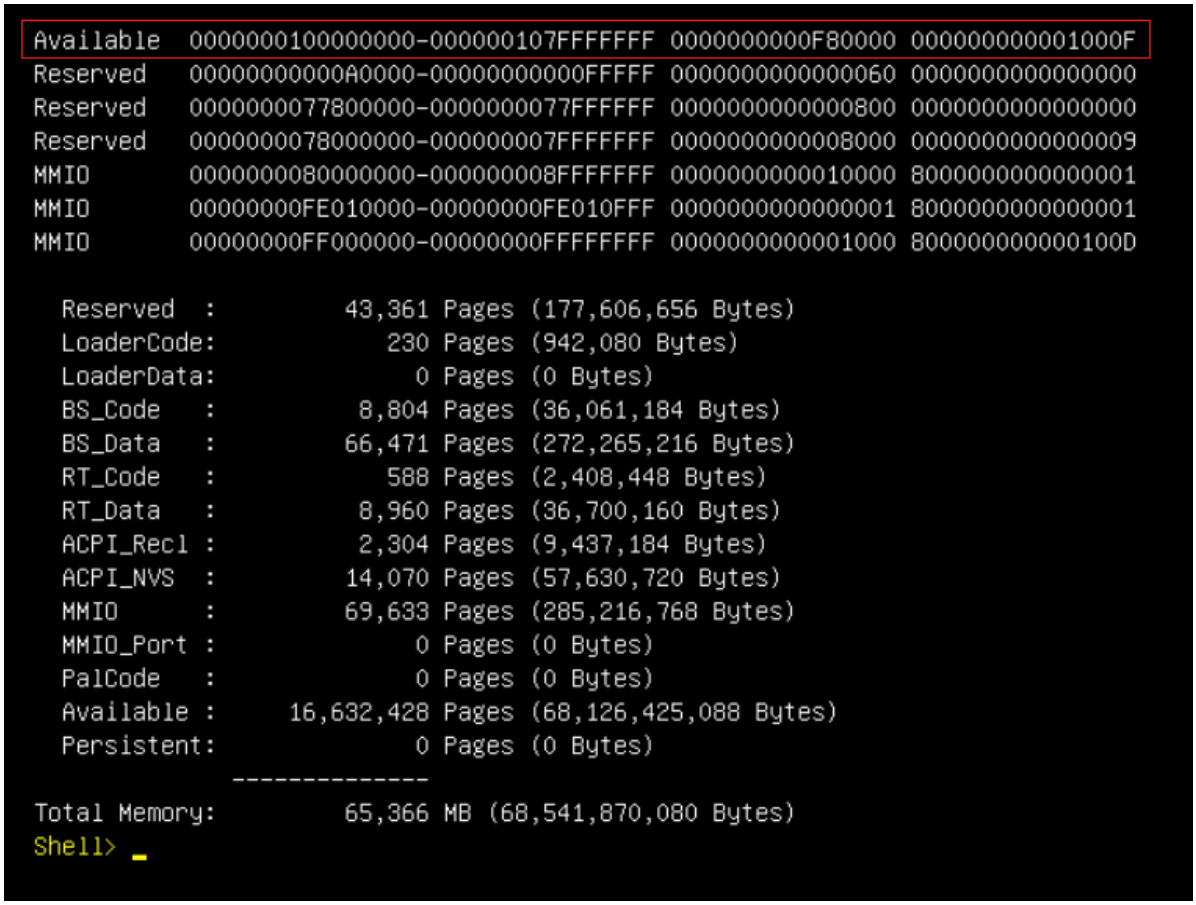

Figure 10. Full Mirrored Mode Memory Map

Server UEFI settings for Address Range Mirroring

In UEFI, Address Range Mirroring is referred to as Partial Mirror. It reduces the available system memory by percentage of up to 50% per processor. The percentage is set by the Partial Mirror Ratio In Basis Points setting.

When memory is configured in Partial Mirror mode which is also called Address Range Mirroring, a subset of memory is mirrored and the rest of the memory in non-mirrored mode. Address Range Mirroring allows greater granularity in selecting how much memory is dedicated for redundancy.

Address Range Mirroring requires a firmware-OS interface for a user to specify the desired subset of memory to mirror. Currently Lenovo UEFI settings provides options for user to configure partial mirroring configuration and we don't have an ESXi based tool to configure partial mirroring configuration from OS side.

The following are steps for configuring Address Range Mirroring in UEFI Settings on the ThinkSystem SR650 V3:

1. Configure Adaptive Double Device Data Correction (ADDDC) Sparing as described in the Server UEFI settings for Memory Mirroring section.
2. Configure Mirror Fail-over as described in the Server UEFI settings for Memory Mirroring section.
3. Go to the **System Settings > Memory > Mirror Configuration** page, do the following, as shown in the figure below.
    1. Enable **Partial Mirror**
    2. Enable or disable **Mirror Below 4G**. You can choose enable or disable "Mirror Below 4G" for Partial Mirror. When "Mirror Below 4G" is enabled, all available system memory below the 4GB address limit will be mirrored. When "Mirror Below 4G" is disabled, all available system memory

below the 4GB address limit won't be mirrored.

3. Input **Partial Mirror Ratio in Basis Points.** This option is used for percentage setting of memory mirror for each processor. For example, to mirror 12.75% of memory, input the value 1275; to mirror 25% of memory, input the value 2500.

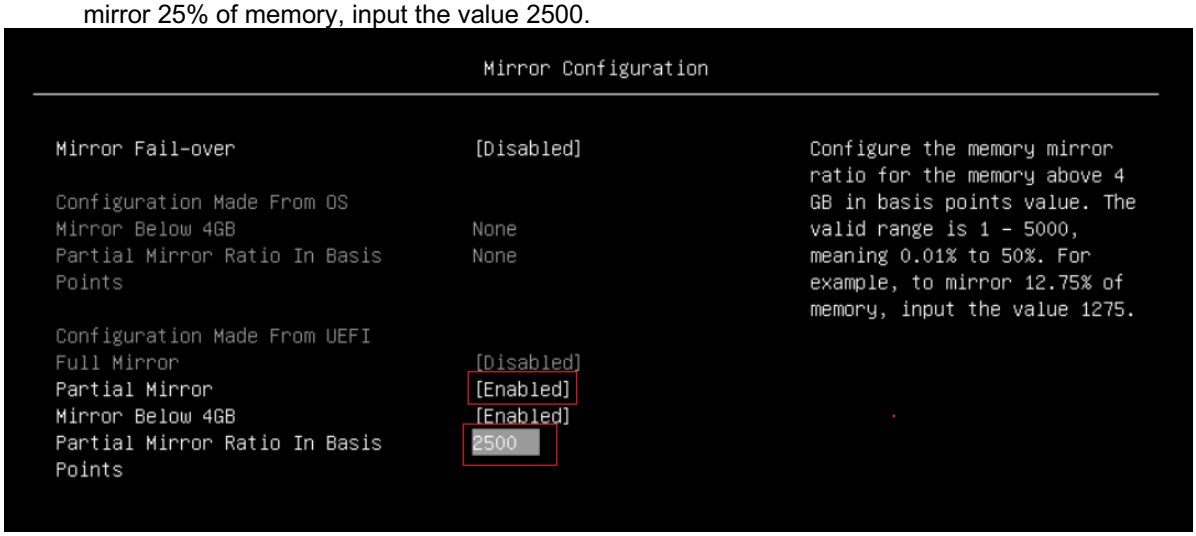

Figure 11. Configure Partial Mirror in UEFI Settings

4. Save Settings
5. Reboot the host so that the partial mirror configuration takes effect.
6. Check if mirrored mode is enabled on POST (Power On Self Test) stage, and the message “Mirrored mode enabled, usable capacity xx GB” is expected on POST page. The system has a total of 128 GB memory detected, mirror 25% of memory, and the usable capacity is 96 GB as shown below.

Figure 12. Splash screen with partial mirror

7. Check Memory Map under EFI shell, and the current memory map can be shown via the **memmap** If memory in independent mode, the top memory address is `0X207fffffff` (130 GB) as shown in Figure 9. If memory is in Address Range Mirroring mode and mirrored 25% of memory, the top memory address is `0x187fffffff` (98 GB) as shown in the figure below. As there’s a 2GB MMIO size under 4GB address space, the memory size in independent mode is 128 GB (130 GB - 2 GB) and memory

size in Address Range Mirroring mode is 96 GB (98 GB - 2 GB).

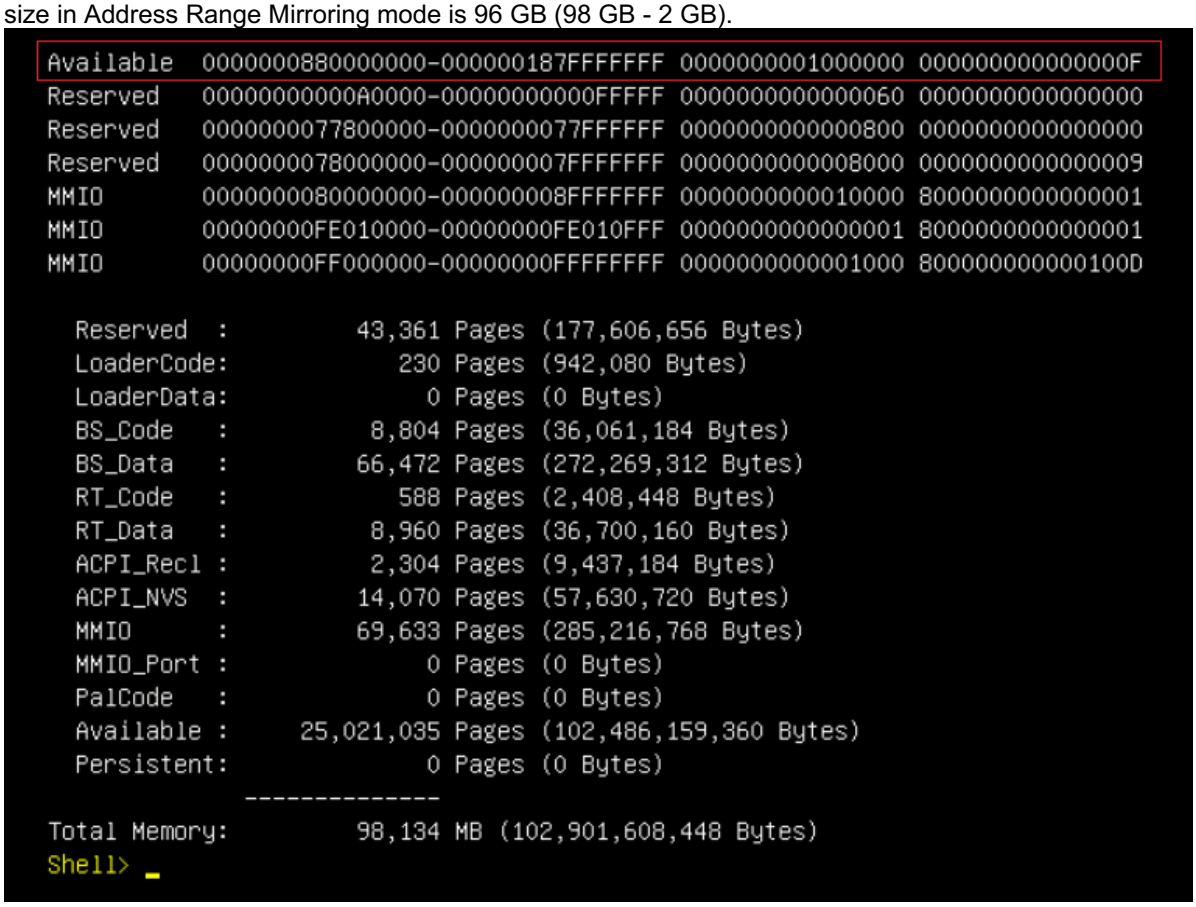

Figure 13. Address Range Mirror Enabled Memory Map

Configuring VMware ESXi for mirroring

vSphere ESXi supports both Memory Mirroring and Address Range Mirroring features. Memory scheduler puts the memory pages consumed by critical services on reliable memory regions to provide highly reliable when memory occurred uncorrectable fault. It's worth noting that more than the kernel can use this feature. We can also configure VM to place the memory pages consumed by VM on reliable memory area to protect the VMs from memory failure.

The following are steps for using reliable memory in ESXi on the ThinkSystem SR650 V3:

1. SSH to ESXi and run the following ESXCLI command to check reliable memory after configuring Memory Mirroring or Address Range Mirroring in UEFI settings.

```
~# esxcli hardware memory get
~# vsish -e get /memory/comprehensive
```

The following figure shows reliable memory in ESXi when memory is configured in Memory Mirroring mode.

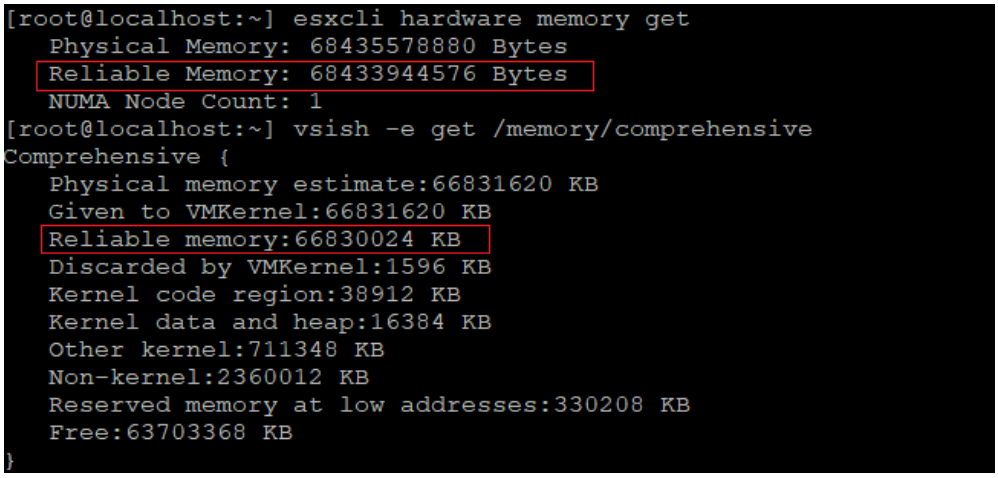

Figure 14. Check full mirror memory in ESXi

The following figure shows reliable memory in ESXi when memory is configured in Address Range Mirroring mode and mirrored 25% of memory.

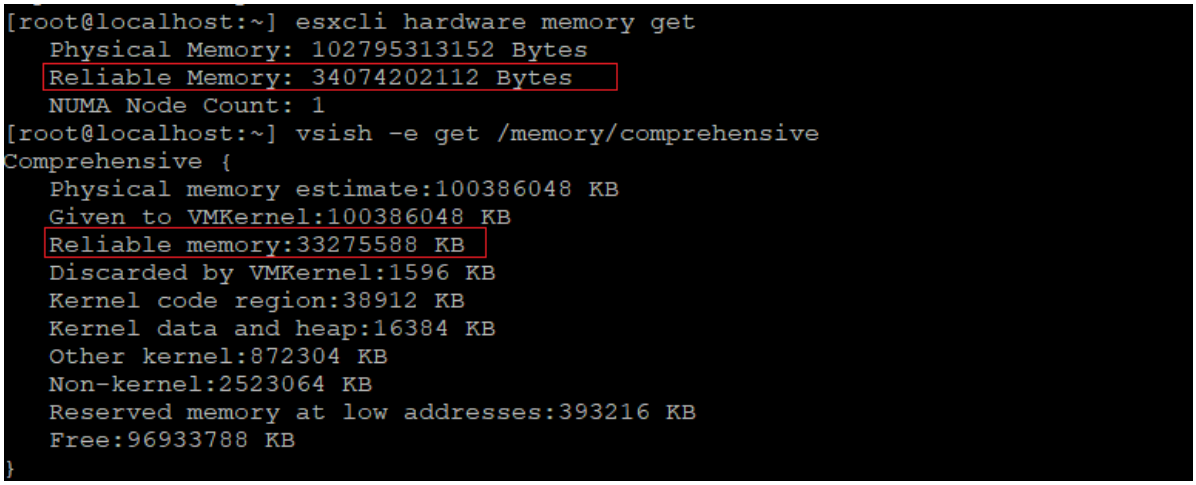

Figure 15. Check partial mirror memory in ESXi

2. We can inject an uncorrectable error (UCE) within mirroring range to verify the memory mirroring feature. If an UCE within mirroring range can be treated as a corrected error (CE) and ESXi keep on running, the test passes. If not, the test fails.

   ESXi Beta build required: Error injection testing requires ESX beta build type as the error injection capabilities are enabled in only ESXi beta build type.

   ESXi provides a vmkernel module mceInjACPI which can be used to inject UCE via VSI interface and use ACPI standard EINJ defined interface. We need to install the einj test vib for the corresponding ESXi build and then you can use mceInjACPI kernel module for error injection.

3. Run the following commands to install the einj test vib on ESXi as shown in the figure below.

   ```
   ~# esxcli software vib install -v
   ```

```
[root@localhost:~] esxcli software vib install -v /opt/einj-test-8.0.2-0.0.22380527.i386.vib
Installation Result
   Message: Operation finished successfully.
   VIBs Installed: VMware_bootbank_einj-test_8.0.2-0.0.22380527
   VIBs Removed:
   VIBs Skipped:
   Reboot Required: false
   DPU Results:
```

Figure 16. Install einj test vib on ESXi

4. In order to test error injection, we need to enable Direct Connect Interface (DCI) for Lenovo ThinkSystem servers due to security policy. Please note that only internal error injection testing requires to enable DCI. Run the following IPMI commands to enable DCI and get DCI status as shown in the figure below.

```
~# ipmitool -I lanplus -H BMC_IP -U USERID -P PASSW0RD raw 0x3a 0x39 0x0d
1
~# ipmitool -I lanplus -H BMC_IP -U USERID -P PASSW0RD raw 0x3a 0x39 0x0c
```

```
D:\ipmitool>ipmitool -I lanplus -H 10.245.39.39 -U USERID -P PASSWORD!! raw 0x3a 0x39 0x0d 1

D:\ipmitool>ipmitool -I lanplus -H 10.245.39.39 -U USERID -P PASSWORD!! raw 0x3a 0x39 0x0c
 01
```

Figure 17. Enable and get DCI status

5. Run the following commands to inject an UCE to mirroring range on beta type ESXi as shown in the figure below. We can refer to memmap in Figure 10 or Figure 13 to select a mirroring memory address or non-mirroring address for error injection.

```
~# vmkload_mod mceInjACPI
~# vsish -e set /system/loglevels/MceInj 3
~# vsish -e get /vmkModules/mceInjACPI/help
~# vsish -e set /vmkModules/mceInjACPI/errorType 4
~# vsish -e set /vmkModules/mceInjACPI/errorAddress 0x21d261040 0xffffffff
ffffffc0
~# vsish -e set /vmkModules/mceInjACPI/trigger 1
~# vsish -e set /vmkModules/mceInjACPI/inject 1
```

```
[root@localhost:~] vmkload_mod mceInjACPI
Module mceInjACPI loaded successfully

[root@localhost:~] vsish -e set /system/loglevels/MceInj 3
[root@localhost:~] vsish -e get /vmkModules/mceInjACPI/help
errorType: type of error to inject
errorAddress: physical address of the error
trigger: whether to execute the trigger action
inject: inject error
stopInject: needed for error types that autorepeat
Error Types supported by the platform:
Memory Correctable  3
Memory Uncorrectable non-fatal  4
Memory Uncorrectable fatal  5
PCI Express Correctable  6
PCI Express Uncorrectable non-fatal  7
PCI Express Uncorrectable fatal  8
Unknown/Reserved  12
Vendor Defined  31

[root@localhost:~] vsish -e set /vmkModules/mceInjACPI/errorType 4
[root@localhost:~] vsish -e set /vmkModules/mceInjACPI/errorAddress 0x21d261040 0xffffffffffffffc0
[root@localhost:~] vsish -e set /vmkModules/mceInjACPI/trigger 1
[root@localhost:~] vsish -e set /vmkModules/mceInjACPI/inject 1
```

Figure 18. Inject UCE to mirrored range

6. Check vmkernel.log to see if UCE can be downgraded to CE and check if ESXi keep on running, and see the vmkernel log as shown in the following figure:

```
2023-12-07T07:16:11.079Z In(182) vmkernel: cpu15:1000868216)MceInj: MCEInjACPI_SetErrorType:139: Set type to Memory
Uncorrectable non-fatal (4)
2023-12-07T07:16:24.978Z In(182) vmkernel: cpu26:1000868219)MceInj: MCEInjACPI_InjectError:223: Injecting type=Memory
Uncorrectable non-fatal (4), flags=0x2, address=0x21d261040, rangeMask=0xfffffffffffffc0  pcpu=0, sbdf=0x0
2023-12-07T07:16:25Z In(182) vmkernel:
2023-12-07T07:16:25.032Z In(182) vmkernel: cpu26:1000868219)MceInj: MCEInjACPI_InjectError:232: Inject return status:
Success
2023-12-07T07:16:25.032Z In(182) vmkernel: cpu27:1000867121)MCA: MCELogBank:201: CE Intr G0 Bc S8c00004001010090
A21d261040 M4800043020f86086 P21d261040/40 Memory Controller Read Error on Channel 0.
2023-12-07T07:16:25.032Z In(182) vmkernel: cpu26:1000868219)MceInj: MCEInjACPI_InjectError:239: Trigger return status:
Success
2023-12-07T07:16:25.032Z In(182) vmkernel: cpu26:1000868219)MceInj: MCEInjACPI_InjectError:246: EndOperation return
status: Success
2023-12-07T07:16:25.056Z In(182) vmkernel: cpu13:1000866220)MCA: MCELogBank:201: CE Intr G0 B13 S8c400043001000c0
A21d261040 M900018172806886 P21d261040/40 Memory Controller Scrubbing Error on Channel 0.
```

Figure 19. UCE downgraded to CE in vmkernel log

7. To compare the UCE injection to mirroring range, we can inject an UCE to non-mirroring address range as shown in following figure.

```
[root@localhost:~] vsish -e set /vmkModules/mceInjACPI/errorType 4
[root@localhost:~] vsish -e set /vmkModules/mceInjACPI/errorAddress 0x187fffffff 0xfffffffffffffc0
[root@localhost:~] vsish -e set /vmkModules/mceInjACPI/trigger 1
[root@localhost:~] vsish -e set /vmkModules/mceInjACPI/inject 1
```

Figure 20. Inject UCE to non-mirrored range

The memory controller detects the UCE and then triggers machine check exception (MCE) to ESXi, and system will run into purple screen of death (PSOD) as shown in following figure.

```
VMware ESXi 8.0.2 [BETAbuild-22380527 x86_64]
Machine Check Exception on PCPU10 in world 1001390798:tq:tcpip4
System has encountered a Hardware Error - Please contact the hardware vendor

UC Excp G5 Ba Sfe20000000031146 A1880000000 M80613c610300486 P1880000000/40 Cache Hierarchy: Level 2 Data Cache DataWrite Error.

Uncorrectable/unrecoverable machine check error

cr0=0x8001003d cr2=0x3d1dc4d3bf0 cr3=0x20e000 cr4=0x14216c
FMS=06/8f/6 uCode=0x2b000461
frame=0x4520400a5eb0 ip=0x42000934d0b8 err=0x12 rflags=0x86
rax=0x2636eb rbx=0x1925c0552f4 rcx=0x2b
rdx=0x1 rbp=0x4300a5209fa0 rsi=0x10
rdi=0x4300a520ac70 r8=0x0 r9=0x41ffc92b61a0
r10=0x1 r11=0x4538d709f000 r12=0x10
r13=0x4300a8211000 r14=0x0 r15=0x4300a8211000
*PCPU10:1001390798/tq:tcpip4
PCPU  0: UUUUIUUUUISUSSUISUUSSUSUSISUUSSIUUUSUSIIUSUISUSUUIUISIUSSUUIUSUS
Code start: 0x42000920000 VMK uptime: 0:00:13:29.024
0x4538d709a938:[0x42000934d0b8]Histogram_InsertN@vmkernel#nover+0x4 stack: 0xef
0x4538d709a940:[0x420009326e03]BH_EnablePreemptionStatsKeeping@vmkernel#nover+0x98 stack: 0x80000000
0x4538d709a960:[0x420009809dfa]CpuSched_PreemptEnabledStatsUpdate@vmkernel#nover+0x2f stack: 0x1
0x4538d709a980:[0x42000980a222]CpuSchedEnablePreemption@vmkernel#nover+0x5f stack: 0xef
0x4538d709a9a0:[0x42000935312c]IntrCookie_DoInterrupt@vmkernel#nover+0x891 stack: 0x4538d709aac0
0x4538d709aa60:[0x4200093535591]IntrCookie_VmkernelInterrupt@vmkernel#nover+0x5e stack: 0xef
0x4538d709aa90:[0x4200093fde4d]IDT_IntrHandler@vmkernel#nover+0x132 stack: 0x0
0x4538d709aac0:[0x4200093f40c5]gate_entry@vmkernel#nover+0xb6 stack: 0x8001003d
0x4538d709ab80:[0x4200092cf909]Power_ArchPerformWait@vmkernel#nover+0xed stack: 0x2
0x4538d709aba0:[0x4200092cfaac]Power_ArchSetCState@vmkernel#nover+0xd1 stack: 0x100000020
0x4538d709abf0:[0x42000981053c]CpuSchedIdleLoopInt@vmkernel#nover+0x515 stack: 0x420042801040
0x4538d709ac70:[0x420009816e0]CpuSchedChooseAndSwitch@vmkernel#nover+0x2179 stack: 0x4538d709f000
0x4538d709ad60:[0x420009817a2]CpuSchedDispatch@vmkernel#nover+0x233 stack: 0x1
0x4538d709add0:[0x420009817c51]CpuSchedWait@vmkernel#nover+0x486 stack: 0xffffffffffffffff
0x4538d709af40:[0x420009247a18]VmkTimerQueueWorldFunc@vmkernel#nover+0x461 stack: 0x390079f000
0x4538d709afd0:[0x420009818b6c]CpuSched_StartWorld@vmkernel#nover+0x135 stack: 0x0
0x4538d709b000:[0x42000932cdc3]Debug_IsInitialized@vmkernel#nover+0x18 stack: 0x0
lastClrIntrRA:0x420009394b76 base fs=0x0 gs=0x420042800000 Kgs=0x0
No place on disk to dump data.
Finalized dump header (16/16) FileDump: Successful.
```

Figure 21. ESXi PSOD after inject UCE to non-mirrored range

8. We can configure VM to place the memory pages consumed by VM on reliable memory area to protect the VMs from memory failure.
   Create a VM on ESXi and edit the .vmx file to configure reliable memory for the VM. The .vmx file is typically located in the directory where you created the virtual machine. You can also run command `find / -name "*.vmx"` on ESXi to get the location of the .vmx file.

9. Power off the VM, edit the .vmx file and add the following parameter and then save the settings, as shown in figure below.

```
sched.mem.reliable = "TRUE"
```

```
usb:1.speed = "2"
usb:1.present = "TRUE"
usb:1.deviceType = "hub"
usb:1.port = "1"
usb:1.parent = "-1"
svga.guestBackedPrimaryAware = "TRUE"
usb:0.present = "TRUE"
usb:0.deviceType = "hid"
usb:0.port = "0"
usb:0.parent = "-1"
tools.remindInstall = "TRUE"
sched.mem.reliable="TRUE"
:wq
```

Figure 22. Configure reliable memory for VM

10. Power on VM, then memory pages consumed by virtual machines will be on reliable memory area.

References

For additional information, see the following:

- Address Range Partial Memory Mirroring
  https://www.intel.com/content/www/us/en/developer/articles/technical/address-range-partial-memory-mirroring.html
- vSphere Reliable Memory
  https://docs.vmware.com/en/VMware-vSphere/8.0/vsphere-resource-management/GUID-5639BA75-E1C0-4137-BB77-20829E673740.html

Author

Chengcheng Peng is a VMware Engineer in the Lenovo Infrastructure Solutions Group in Beijing, China. As a VMware engineer with 6 years' experience, she mainly focuses on vSphere security and storage.

Thanks to the following people for their contributions to this project:

- Boyong Li, Lenovo OS Technical Leader
- Skyler Xing12 Zhang, Lenovo VMware Engineer
- Alpus Chen, Lenovo VMware Engineer
- David Hsia, Lenovo VMware Engineer
- Chia-Yu Chu, Lenovo VMware Engineer
- Gary Cudak, OS Architect and WW Technical Lead
- David Watts, Lenovo Press

Related product families

Product families related to this document are the following:

- Memory
- Processors

Notices

Lenovo may not offer the products, services, or features discussed in this document in all countries. Consult your local Lenovo representative for information on the products and services currently available in your area. Any reference to a Lenovo product, program, or service is not intended to state or imply that only that Lenovo product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any Lenovo intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any other product, program, or service. Lenovo may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

Lenovo (United States), Inc.
8001 Development Drive
Morrisville, NC 27560
U.S.A.
Attention: Lenovo Director of Licensing

LENOVO PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary. Lenovo may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any references in this publication to non-Lenovo Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this Lenovo product, and use of those Web sites is at your own risk. Any performance data contained herein was determined in a controlled environment. Therefore, the result obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

This document, LP1877, was created or updated on December 27, 2023.

Send us your comments in one of the following ways:

- Use the online Contact us review form found at:
  https://lenovopress.lenovo.com/LP1877
- Send your comments in an e-mail to:
  comments@lenovopress.com

This document is available online at https://lenovopress.lenovo.com/LP1877.

Trademarks

Lenovo and the Lenovo logo are trademarks or registered trademarks of Lenovo in the United States, other countries, or both. A current list of Lenovo trademarks is available on the Web at https://www.lenovo.com/us/en/legal/copytrade/.

The following terms are trademarks of Lenovo in the United States, other countries, or both:
Lenovo®
ThinkSystem®

The following terms are trademarks of other companies:

Intel® and Xeon® are trademarks of Intel Corporation or its subsidiaries.

Other company, product, or service names may be trademarks or service marks of others.