Dell EMC Fault Resilient Memory

This technical white paper briefs about the Dell EMC's Fault Resilient Memory (FRM) operating mode. This document covers the behavioral aspects of FRM on yx4x PowerEdge servers and later generation servers with VMware ESXi.

Abstract

This technical white paper briefs about Fault Resilient Memory (FRM) mode on VMware ESXi, and the Reliable Memory (ReM) technology enabled from VMware ESXi, which uses the resilient region exposed by the platform.

November 2019

Revisions

Date	Description
December 2013	Initial release
November 2019	Updated

Acknowledgements

This paper was produced by the following:

Author: Krishnaprasad K, Principal Engineering Technologist

Support: Ramya D R, Technical Writer, IDD team

Others: Mukund Khatri, Sandeep J

Table of contents

Executive summary

This technical white paper elaborates on the behavioral aspects of Fault Resilient Memory operating mode on Dell EMC's yx4x PowerEdge servers and later. This document briefs about the command line options and VMware documentation which helps the administrators to know the memory reserved for reliable region, map the memory address range used by userworld and virtual machines to the reliable memory region.

1 Introduction to Dell Fault Resilient Memory (FRM)

Fault Resilient Memory (FRM) is an operating mode introduced from yx2x PowerEdge servers. The mode establishes an area of memory that is fault resilient and can be used by hypervisors such as VMware ESXi to load vmkernel, critical applications or services to maximize system availability etc. The operating systems uses the resilient region exposed from the platform and map the process's address ranges to enforce resiliency. This paper explicitly details on the behavioral aspects of yx4x PowerEdge servers and later generation servers.

1.1 Audience and Scope

The intended audience for this white paper includes IT Administrators and Channel Partners planning to use this new memory operating mode introduced from yx4x PowerEdge servers. Fault Resilient Memory is a feature aimed at customers concerned with highly reliable virtualization platforms. It is not targeted at customers leveraging highly overcommitted memory resources or needing the highest possible memory performance.

1.2 Benefits of enabling FRM

FRM creates a highly resilient memory zone for the hypervisor, protecting it from severe memory errors. With the VMware Reliable Memory feature, vSphere 5.5 and later revisions can take advantage of this zone, providing the hypervisor with a strong protection from memory faults that would bring down the entire system. When used in conjunction with other Dell PowerEdge reliability and redundancy features for virtualization -- such as the Internal Dual SD Modules usable with ESXi installations and Memory Page Retire, which helps the hypervisor to locate and fence off areas prone to many correctable errors that could lead to an uncorrectable error in the standard memory space -- it creates a highly reliable VMware virtualization server environment for customers concerned with high availability and uptime.

1.3 Prerequisites

The following are the prerequisites for utilizing FRM:

- The FRM operating mode is introduced from yx2x PowerEdge servers, and it is continued to support the yx3x and yx4x PowerEdge servers.
- FRM memory mode is supported only on systems with Intel Xeon Gold and Platinum series processors.

Note: Dell AMD servers don't support FRM.

- There are only specific server models to support this feature. Ensure that you refer to the respective server hardware Owner's Manual. In the Owner's Manual, see the **Memory Settings→ Memory Operating Mode** section to know if this feature is supported. If this feature is supported, it is mentioned as **Dell Fault Resilient Mode**.
- DIMMs population to adhere a specific memory matrix which is supported. See the respective server hardware Owner's Manual which briefs on the specific slots that needs to be populated to enable this feature.
- VMware vSphere 5.5 version or later offers the Reliable Memory (ReM) feature (Enterprise or Enterprise Plus editions). There is no separate token required to enable ReM from VMware ESXi as it detects automatically when FRM is enabled from the server hardware BIOS.
- The VMware Reliable Memory (ReM) is supported only when the user use "Enterprise" or "Enterprise Plus" edition of license.

1.4 Behavioral description of FRM on yx4x PowerEdge servers and later

On the yx4x PowerEdge server, only FRM operating mode is supported. Overall 25 percentage of system memory is reserved for reliable memory region. The 25 percentage of the total system memory populated on the server is reserved for reliable region. The 75 percentage of the total system memory populated is available for the end user applications for general purpose use.

- When FRM is enabled with the Node Interleaving option set to enable, then the 25 percentage reliable memory region is occupied across the sockets.
- When FRM is enabled with the Node Interleaving option set to disable (NUMA Enabled), then the 25 percentage reliable memory region is occupied from 1st socket and then followed by the 2nd socket until the 25 percentage is reserved. However, if there is a case arises where complete NUMA Node memory needs to be used to complete 25 percentage reservation, then BIOS further divides the allocation between NUMA nodes to ensure there is no major NUMA imbalance.

1.5 Enabling FRM in PowerEdge servers

To enable FRM on the supported Dell EMC PowerEdge server, complete the following steps:

1. In the BIOS setup, select **Memory Settings**.
2. Under the **Memory Operating Mode** section, select **Dell Fault Resilient Mode**.
3. Click **Save** and exit from the BIOS setup.

Installing VMware ESXi 5.5 GA or later on the server (ensure to look at VMware HCL to select the supported ESXi version against the specific server model) makes vmkernel and other critical applications or services loads into the reliable region by default. If you are using an edition of VMware vSphere with the Reliable Memory feature enabled, it will automatically locate the FRM space and load the hypervisor into the protected zone. The user does not have to do any configuration changes within ESXi or vCenter to enable ReM.

Note: It is recommended that you must NOT enable FRM on earlier vSphere versions than 5.5, or any other OS. If you do so, while the system will function, but the memory used to create the fault tolerant zone will be still consumed and the protected zone will be used randomly, thus causing the wastage of system memory.

The below screenshot is taken from a Dell EMC PowerEdge R840 indicating the BIOS token to enable FRM in server BIOS.

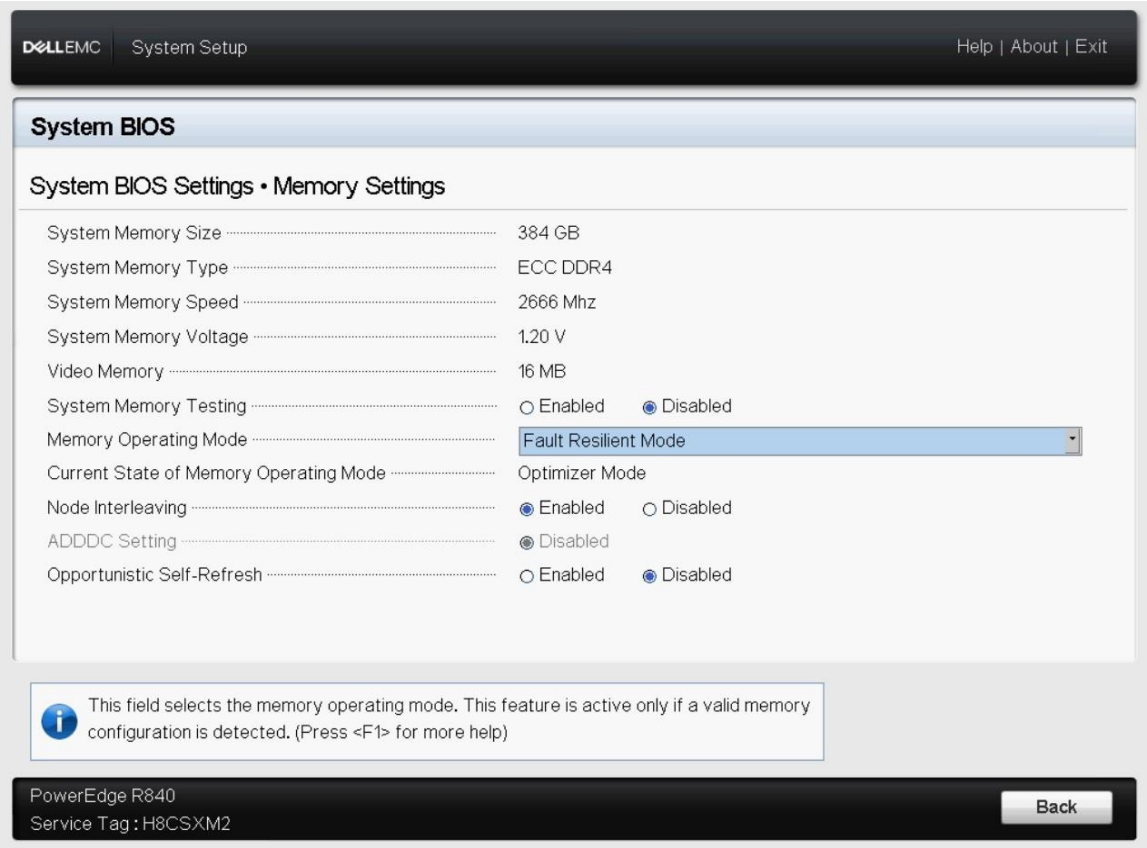

Figure 1 : Enabling FRM Memory Operating Mode in the BIOS

1.6 Verifying ReM from VMware ESXi

VMware ESXi 5.5 and later versions enables Reliable Memory (ReM) by default (for supported versions - Enterprise, Enterprise Plus). The user does not have to make any configuration changes in ESXi to enable ReM. There are few utilities in VMware, that provides the status of the reliable memory region. For example, the command: esxcli sched reliablemem get, mentioned in the below screenshot indicates that the system is FRM/ReM enabled or not. A value of true (set by default) indicates that the feature is enabled.

```
~ #
~ # esxcli sched reliablemem get
true
~ #
```

Figure 2 : FRM Token in ESXi

1.7 Validate Fault Resilient Memory reservations in VMware ESXi

This section briefs about ESXi commands which helps the end users to understand and configure FRM accordingly.

The command: `esxcli system settings kernel list | grep useReliableMem` indicates the status of kernel parameter `useReliableMem`. By default, this parameter is set to TRUE.

Note: It is recommended that end user do not change this setting unless instructed by Dell EMC or VMware for any debugging purpose.

```
~ # esxcli system settings kernel list | grep useReliableMem
useReliableMem                          Bool    System is aware of reliable memory.
                                                                        TRUE        TRUE        TRUE
~ #
```

Figure 3 :Listing ESXi ReM kernel parameter

Commands such as esxtop and esxcli hardware memory get help the user to understand the memory reserved for fault resiliency, when this option is enabled from sever BIOS.

The below screenshot is an output when you run the command Esxtop, and then press m. This output shows the memory specific attributes. NUMA/MB field indicates the overall NUMA nodes created on the system (when Node Interleaving is disabled in server BIOS). For example, the below screenshot is captured from a Dell EMC PowerEdge R840 with FRM enabled and Node Interleaving disabled (NUMA Enabled).

```
 1:14:42am up 24 min, 1174 worlds, 0 VMs, 0 vCPUs; MEM overcommit avg: 0.00, 0.00, 0.00
PMEM  /MB: 294104   total: 3074     vmk,177 other, 290853 free
VMKMEM/MB: 293719 managed:  3551 minfree, 10642 rsvd, 283077 ursvd,  high state
NUMA  /MB: 48343 (46375), 49152 (47659), 98304 (98216), 98304 (98216)
PSHARE/MB:      23  shared,      23  common:       0 saving
SWAP  /MB:       0    curr,       0 rclmtgt:                0.00 r/s,   0.00 w/s
ZIP   /MB:       0  zipped,       0   saved
MEMCTL/MB:       0    curr,       0  target,       0 max

   GID NAME              MEMSZ    GRANT     CNSM     SZTGT     TCHD   TCHD_W    SWCUR    SWTGT   SWR/s   SWW/s LLSWR/s  LLS
  6023 hostd.2100997     85.49    52.98    57.46    62.75    14.48    10.01     0.00     0.00    0.00    0.00    0.00
 10185 vpxa.2101569      28.65    16.65    19.89    21.55     6.05     2.82     0.00     0.00    0.00    0.00    0.00
  1079 vmsyslogd.20983   17.49    13.91    14.54    15.93     7.31     6.68     0.00     0.00    0.00    0.00    0.00
 11362 dcui.2101768      16.13     3.85     4.56     4.95     1.48     0.77     0.00     0.00    0.00    0.00    0.00
  5700 hostdCgiServer.   16.10     8.82     9.52    10.40     1.01     0.31     0.00     0.00    0.00    0.00    0.00
  1223 vobd.2098362      14.85     3.16     3.96     4.28     0.91     0.10     0.00     0.00    0.00    0.00    0.00
  6184 rhttpproxy.2101   13.00     2.56     5.43     5.69     3.39     0.52     0.00     0.00    0.00    0.00    0.00
  1071 vmsyslogd.20983   11.77     9.59    10.07    11.03     9.67     9.18     0.00     0.00    0.00    0.00    0.00
 12978 esxtop.2101999     9.74     5.82     6.48     7.06     3.64     2.98     0.00     0.00    0.00    0.00    0.00
```

Figure 4 :NUMA Representation from ESXi

In the above example, when Node Interleaving is disabled, there would be 4 NUMA nodes created in the system each of 96 GB. With FRM enabled and Node Interleaving disabled together, first two NUMA Nodes are reported with 48 GB of memory and last two socket NUMA nodes capacity is 96 GB each. This is because 25 percentage of memory reservation to be allocated in the yx4x system and hence 96 GB (25 percentage of 384 GB) is reserved for memory fault resiliency. The 96 GB is further divided between first two NUMA nodes equally to ensure that there is no NUMA node imbalance.

The command in the below screenshot depicts the reliable memory reported with in ESXi, as the same example mentioned above. As noted above, system BIOS reserves approximately 96 GB as reliable memory and hence 288 GB is reported as system memory. You may observe a slightly lower memory reported in ESXi (287.2 GB) comparing the system memory reported in BIOS (288 GB). This is a known issue in ESXi and is documented in VMware KB 2149889.

```
[root@sc-pesx03:~] esxcli hardware memory get
   Physical Memory: 308391202816 Bytes
   Reliable Memory: 102231154688 Bytes
   NUMA Node Count: 4
[root@sc-pesx03:~] _
```

Figure 5 :Reliable Memory reported with in ESXi

1.8 Configure Reliable memory for Virtual Machines

VMware define a priority mechanism for various processes running on ESXi to ensure that the reliable memory region is mapped to processes based on their priority. VMKernel and the VMM gets the highest priority (Priority 0) and they make use of the reliable memory region when it is enabled from the platform. Some of the critical userworld processes such as hostd, vpxa services running on ESXi also make use of reliable memory region using the tag 'memreliable'. These userworld critical processes are marked as Priority 1. Similarly, an administrator can tag a virtual machine memory address ranges to the protected region by explicitly adding a parameter to the virtual machine's .vmx file. For more information, see the VMware KB 2146595.

1.9 Monitoring FRM redundancy failure

This section describes about monitoring the memory redundancy related events from iDRAC System Event Log (SEL). When an uncorrectable error (UCE) occurs on a reliable region for the first time, the SEL logs an entry, but the hypervisor does not crash. This clearly shows that the FRM provided memory redundancy is lost and occurrence of one more uncorrectable error to any of the memory addresses' in Socket 0 will result in a Purple Screen Of Death (PSOD).

Note that the system continues working with more than one UCE, as long as the error is not persistent. The system always write-back data when UCE is detected. If the UCE is still persistent after the write-back, then the memory redundancy is lost.

Figure 6 : FRM Memory redundancy lost when uce occurs in the memory channels in Socket 0

2 FAQ

1. What are the Dell server models that support FRM?

- For more information, see your respective server's Installation and Service Manual available at www.dell.com/support. See the **Memory Settings** section to know if your server supports FRM.

2. Does FRM require any specific DIMM population configurations and what are the configurations?

- Yes, FRM requires specific DIMM population configuration. For more information. See the **General memory module installation guidelines** section in your respective server's Installation and Service Manual. This is also applicable to FRM.

3. Can FRM be utilized in conjunction with other memory RAS technologies such as SDDC (Single Device Data Correction), ADDDC (Adaptive Double Device Data Correction), MPR (memory page retire), Scrubbing, ChipKill, etc.?

- FRM can be used in conjunction with other memory RAS technologies as listed above except ADDDC. However, FRM cannot be used in conjunction with full memory mirroring, single rank sparing, and multi-rank sparing.

4. In BIOS, **Optimizer Mode** and **Advanced ECC mode** are listed under **Memory operating mode**. Is the **Advanced ECC** option still available to customers when selecting FRM?

- **Advanced ECC** does not exist in yx4x PowerEdge servers. The Advance ECC option is used to imply SDDC, which required running the memory in lockstep mode. Therefore, there is a separate memory mode option. In yx4x PowerEdge servers you get SDDC (Advanced ECC) in an independent mode. You will still get SDDC in FRM for the memory regions that are not being mirrored.

5. Can FRM be leveraged on servers with NVDIMMs?

- No, FRM is unsupported when NVDIMMs are used.

6. What hypervisor supports FRM?

- VMware ESXi is the only hypervisor that supports FRM.

7. Which operating system and hypervisors (for example, Windows Server, RHEL) supports FRM?

- VMware vSphere 5.5 and later versions support FRM.

8. After configuring FRM, if a server is populated with 1 TB physical RAM, then how much reliable and standard memory space is available to the Hypervisor/OS?

- In 1 TB physical RAM, 768 GB of memory capacity is available for hypervisor and applications. Remaining 256 GB memory is reserved as reliable region and can used by vmkernel and critical applications to get loaded.

9. Are there any methods to test the functionality of FRM?

- Yes, there are internal test methodologies that Dell EMC uses to validate FRM, but that is not available to the end users.

3 Summary

This white paper details both Dell EMC's Fault Resilient Memory mode and VMware's Reliable Memory technology. It explains about the feature, the prerequisites required to enable the reliable memory and the steps needed to enable FRM from the PowerEdge server BIOS. The paper also provides some of the command line utilities to monitor the size reserved for reliable region, and documentation on marking virtual machine's memory to the reliable region etc.

4 References

- What's New in the VMware vSphere 6.0 Platform
- Reliable Memory
- Dell Fault resilient memory – Initial release