



曦云®系列通用计算 GPU

Driver EID Message

CSRD-24004-020-F3_V01

2024-09-06

沐曦专有和三级保密信息

本文档受 NDA 管控

更新记录

版本	日期	更新说明
V01	2024-09-06	首次发布

目录

1. 背景	1
2. 概述	2
3. 日志收集方法	4
4. METAX 常见错误	5
4.1 METAX-EID-1: ATU Fault	5
4.2 METAX-EID-2: XCORE Shader Error	6
4.3 METAX-EID-3: CE Dequeue Timeout	7
4.4 METAX-EID-4: HBM 显存坏页错误	7
4.5 METAX-EID-5: 引用 GPU 驱动失败	8
4.6 METAX-EID-6: 等待 SMP 响应超时	8
4.7 METAX-EID-7: PCIe BAR 空间不够	8
4.8 METAX-EID-8: 虚拟地址冲突错误	9
4.9 METAX-EID-9: HBM 显存耗尽	9
4.10 METAX-EID-10: 系统内存耗尽	10
4.11 METAX-EID-11: INT 中断溢出错误	10
4.12 METAX-EID-12: PCIe AER Error	11
4.13 METAX-EID-13: PCIe 通讯断开	11
5. MXCD 常见错误	12
5.1 MXCD-EID-1: CE Dequeue Error	12
5.2 MXCD-EID-2: IOCTL 操作设备失败	12
6. MXGVM 常见错误	13
6.1 MXGVM-EID-1: 映射 GPU 卡 PCIe BAR 失败	13
6.2 MXGVM-EID-2: GPU 卡初始化由于 SMP 通信超时而失败	13
6.3 MXGVM-EID-3: GPU 卡 VBIOS 与 GPU 驱动版本不兼容	14
6.4 MXGVM-EID-4: GPU 卡不支持 SRIOV 虚拟化	14
6.5 MXGVM-EID-5: GPU 卡 SRIOV BAR 空间不够	15
6.6 MXGVM-EID-6: 开启 GPU 卡 SRIOV 虚拟化失败	15
6.7 MXGVM-EID-7: GPU 卡 FLR 复位失败	15
6.8 MXGVM-EID-8: GPU 卡 PCIe 通讯断开	15
6.9 MXGVM-EID-9: GPU 卡 ATU Fault	16

表目录

表 2-1 LOGLEVEL 字段定义 2

表 3-1 日志收集方法 4

表 4-1 METAX 常见错误列表 5

表 5-1 MXCD 常见错误列表 12

表 6-1 MXGVM 常见错误列表 13

1. 背景

本文档描述了曦云®系列 GPU KMD 内核驱动程序报告的 EID（Error Message ID）信息格式，旨在帮助开发人员、FAE 和系统管理员理解这些信息具体的含义，以达到对 GPU 相关问题分析和解决的目的。

2. 概述

当 GPU KMD 内核驱动检测到错误后，会按照特定格式将错误信息打印到操作系统内核日志中，这些日志记录了包括但不限于 GPU 硬件错误、系统故障或应用程序问题等信息。

在 Linux 系统中，GPU KMD 内核驱动打印日志通常会放在 `/var/log/messages` 路径下，通过以下命令可以查看 GPU KMD 打印的错误信息：

```
$ sudo grep -P "METAX.*ERROR|MXCD.*ERROR|MXGVM.*ERROR" /var/log/messages
```

说明

Linux 内核日志存放路径会因 Linux 发行版本不同而有所区别，主流 Linux 发行版内核日志存放路径如下：

- Ubuntu: `/var/log/kernel.log`
- CentOS: `/var/log/messages`

如果存在 GPU KMD 内核驱动错误记录，比如 ATU Fault，会得到类似如下日志：

```
[...] METAX.B10200.ATU.ERROR atu 0x0 pde_base_addr 0x50016c0000
[...] METAX.B10200.ATU.ERROR xpid 1 vfid 0 vf 0 page va 224200000
[...] METAX.B10200.ATU.ERROR hwip DMAW0 fault code(7) perm code: 2
```

每一条日志记录的格式为：

```
MODULE.DBDF.SUBMOD.LOGLEVEL CONTENT
```

其中，每个字段的具体定义如下：

- **MODULE**：表示打印日志的主模块，目前包括 METAX、MXCD 和 MXGVM 几个主模块。
- **DBDF**：表示日志所属的 GPU 卡，以 GPU 卡的 PCIe DBDF（Domain Bus Device Function）作为 GPU 标识号。

该 GPU 标识号具体格式为 `Bxxxx`，字母 `B` 为 GPU 标识号的前导编字母，`xxxx` 为 GPU 标识号的 PCIe DBDF 16 进制编码值（32 比特位宽）。其中，bits[31:16]为 Domain，bits[15:8]为 Bus，bits[7:3]为 Device，bits[2:0]为 Function。

比如，GPU 标识号 B10200，对应标准 PCIe DBDF 格式件为 0001:02:00.0。

- **SUBMOD**：表示打印日志的子模块，根据主模块的不同，子模块的具体定义也不一样。
- **LOGLEVEL**：表示日志等级，定义参见下表：

表 2-1 LOGLEVEL 字段定义

编号	LOGLEVEL	说明
1	EMERG	紧急错误，系统不可用

编号	LOGLEVEL	说明
2	ALERT	告警错误，需要立即处理
3	CRIT	严重错误
4	ERROR	一般错误
5	WARN	告警提示
6	INFO	普通信息
7	DEBUG	调试信息

- CONTENT：表示日志信息的具体内容。

3. 日志收集方法

GPU KMD 内核驱动打印日志，可以通过如下方法收集：

表 3-1 日志收集方法

编号	方法	说明
1	dmesg 命令	Linux 系统通用命令，只能查看当前运行时内核日志信息。同时由于 Linux 内核日志缓存区大小限制，可能无法显示当前运行时内核的完整日志。 命令如下： <pre>sudo dmesg grep -P "METAX.*ERROR MXCD.*ERROR MXGVM.*ERROR"</pre>
2	journalctl 命令	基于 systemd 启动的 Linux 系统上的强大日志工具，没有 dmesg 命令的内核日志缓存区限制问题，还可以查看系统的历史运行日志。 命令如下： <pre>sudo journalctl -k -b 0 grep -P "METAX.*ERROR MXCD.*ERROR MXGVM.*ERROR"</pre>
3	/var/log/message 或 /var/log/kernel.log 文件	Linux 系统内核日志在磁盘存放位置，用户需要注意收集的内核日志是否属于当前系统运行时日志，还是系统以前运行的历史日志。 CentOS 系统命令如下： <pre>sudo grep -P "METAX.*ERROR MXCD.*ERROR MXGVM.*ERROR" /var/log/messages</pre> Ubuntu 系统命令如下： <pre>sudo grep -P "METAX.*ERROR MXCD.*ERROR MXGVM.*ERROR" /var/log/kernel.log</pre>
4	mx-report 工具	mx-report 工具的使用方法参见《曦云®系列通用计算 GPU 用户指南》

4. METAX 常见错误

表 4-1 METAX 常见错误列表

编号	错误信息	错误说明
METAX-EID-1	atu x pde_base_addr xxx	ATU Fault
METAX-EID-2	shader exception, pasid x error_type sxxx	XCORE Shader Error
METAX-EID-3	kill queue timeout, pipe x, queue x	CE Dequeue Timeout
METAX-EID-4	xxx-yyy page_offset x size	HBM 显存坏页错误
METAX-EID-5	failed to get module!	引用 GPU 驱动失败
METAX-EID-6	wait SMPx timeout	等待 SMP 响应超时
METAX-EID-7	PCIe vram bar init failed, rrr	PCIe BAR 空间不够
METAX-EID-8	bo xxx va xxx-xxx conflict with yyy-yyy	虚拟地址冲突错误
METAX-EID-9	vram try to use (xxx kB) beyond total memory (xxx kB), failed	HBM 显存耗尽
METAX-EID-10	(xtt + userptr) memory (xxx bytes) beyond limit (xxx bytes)	系统内存耗尽
METAX-EID-11	xxx: add cookie client x src x to fifo failed	INT 中断溢出错误
METAX-EID-12	detect uncorrectable AER, xxxx	PCIe AER Error
METAX-EID-13	get vm fault status with full f, device may be disconnected	PCIe 通讯断开

4.1 METAX-EID-1: ATU Fault

问题现象：

在 GPU 运行过程中，GPU 卡 ATU 检测到存在非法 VA 地址访问（ATU Fault），会打印如下错误信息：

```
METAX.Bxxxx.ATU.ERROR atu x pde_base_addr xxx
METAX.Bxxxx.ATU.ERROR xpid x vfid x vf x page va x
METAX.Bxxxx.ATU.ERROR hwip x fault code x(x) perm code x
```

打印信息解读：

- page va x：表示触发 ATU 错误的虚拟地址
- hwip x：表示触发 ATU 错误的硬件 IP
- fault code x：表示错误类型
- perm code x：表示虚拟地址对应的页表项权限配置

可能原因以及建议操作：

- 页表所处的 HBM 不稳定，报 bug 给硬件厂商。
- KMD 页表管理逻辑出错，报 bug 给硬件厂商。
- 非法 VA 或者越界访问，用户用 GDB 跟踪或者检查自己的代码是否可能导致非法 VA 或者越界访问。

查看进程页表的方法：不要退出或者 kill 用户进程，新建一个窗口用 root 用户登录，执行 `cat /sys/kernel/debug/mxcd/vm` 命令，这个 debugfs 接口会列出当前所有进程使用的内存块的 GPU 虚拟地址和物理地址，然后查看出现 fault 的 VA 是否包含在任何一个内存块中：

```
user va 0x803400000-> pa 0xff264000 size 0x100000 domain XTT cset 0x3
user va 0x804200000-> pa 0x21b000 size 0x1000 domain VRAM cset 0x3
```

4.2 METAX-EID-2: XCORE Shader Error

问题现象：

在执行 GPU 应用程序过程中，如果分发到 GPU 的 kernel 程序执行出现问题，GPU 驱动可能会记录一些 GPU 硬件内部的 XCORE shader 错误信息，并打印到内核日志，具体格式如下：

```
METAX.Bxxxx.XCORE.ERROR shader exception, pasid x error_type xxx(x)
```

打印信息解读：

- pasid x：表示哪个进程下发的任务触发了 XCORE shader error
- error_type xxx：表示触发了哪种类型的 XCORE shader error
 - 0x1: fwe_err
 - 0x2: illegal_inst
 - 0x4: mem_viol (non align access)
 - 0x8: xnack_error
 - 0x10: out_of_range
 - 0x20: fuc_error
 - 0x40: timeout_error

可能原因以及建议操作：

- 若 error type = 1，即 fwe error，一般是卡内部 SRAM 没有初始化导致，报 bug 给硬件厂商。
- 若 error type = 2，即非法指令，确保用户编写的 kernel 函数正确并且编译器正确使用，如果无法解决，报 bug 给硬件厂商。
- 若 error type = 4，即 memory violation，一般是非对齐访问导致的，检查用户编写的 kernel 函数。

- 若 error type = 8, 即 xnack error, 会有对应的 ATU fault 触发, 根据提示信息进一步调试, 参见 4.1 METAX-EID-1: ATU Fault。

4.3 METAX-EID-3: CE Dequeue Timeout

问题现象:

在 GPU 应用进程结束退出时, 驱动会释放 GPU 硬件相关资源。当遇到 GPU 应用程序结束过慢或卡住的问题时, 其中可能原因就是 CE Dequeue Timeout, 并在内核日志看到如下错误信息:

```
METAX.Bxxxx.XCORE.ERROR kill queue timeout, pipe x, queue x, rrr
```

打印信息解读:

- pipe x: 释放失败的 pipe 编号
- queue x: 释放失败的 queue 编号

可能原因以及建议操作:

GPU 内部资源释放失败, 报 bug 给硬件厂商, 然后尝试 warm reset 或重启服务器进行恢复。

4.4 METAX-EID-4: HBM 显存坏页错误

问题现象:

当 GPU 内核驱动检查到 HBM 显存坏页错误时, 会尝试对这些 HBM 显存坏页进行隔离处理, 避免 GPU 应用程序继续使用这些 HBM 显存坏页, 并打印如下信息:

```
METAX.Bxxxx.RAS.ERROR xxx-yyy page_offset x size x(Bytes) info x
```

打印信息解读:

- xxx: 触发 HBM 显存坏页的 IP 名称, 包括 MCCTL0/MCCTL1/MCCTL2/MCCTL3
- yyy: HBM 显存坏页错误类型, 包括 UE/RE/CE/FATAL/FREQ_CALL
- page_offset x: 出现 HBM 显存坏页的显存地址

可能原因以及建议操作:

- 遇到 HBM 显存坏页错误, 可以尝试将业务进程全部退出, 重新执行业务。若重新执行业务还会遇到同样的坏页错误, 可以尝试 warm reset 或重启服务器进行恢复。
- GPU HBM 内存颗粒有问题, 可以尝试换卡。

4.5 METAX-EID-5: 引用 GPU 驱动失败

问题现象:

在执行 `mx-smi` 命令返回失败时, 内核日志打印如下错误信息:

```
METAX.Bxxxx.SMI.ERROR failed to get module
```

可能原因以及建议操作:

在 GPU 内核驱动加载或卸载过程中, 用户正在尝试通过 SMI 接口去访问 GPU 相关资源。请在 GPU 内核驱动加载或卸载完成后, 再执行 GPU 业务操作。

4.6 METAX-EID-6: 等待 SMP 响应超时

问题现象:

```
METAX.Bxxxx.SMP.ERROR wait SMPx timeout
```

打印信息解读:

SMP 可能值有两个: 0 或者 1。在 GPU 初始化过程中, KMD 会和 SMP 0 与 SMP 1 进行同步, 该信息表明 KMD 和 SMP 通信失败了, 驱动无法正常加载。

可能原因以及建议操作:

导致这个问题的原因比较复杂, 需要结合软硬件环境综合判断, 联系硬件厂商解决。

4.7 METAX-EID-7: PCIe BAR 空间不够

问题现象:

在 GPU 卡初始化过程中, 检测到 GPU 卡 PCIe BAR 空间不够, 会打印如下错误信息:

```
METAX.Bxxxx.BASE.ERROR PCIe vram bar init failed, rrr
```

同时, 在内核启动枚举 PCIe 过程中, 还会打印 GPU 卡 BAR 空间资源分配失败的日志:

```
pci xxx BAR x: no space for [mem size xxxx]
...
pci xxx BAR x:failed to assign [mem size xxxx]
```

可能原因以及建议操作:

- 修改服务器 BIOS, 启用 large BAR 功能, 有些服务器选项为启用超过 4GB PCI 地址空间, 具体可联系硬件厂商确认。

- 若启用 large BAR 无法解决问题，则需确认 VBIOS 是否启用了 VF BAR 导致服务器 PCI 空间不够，联系硬件厂商解决。

4.8 METAX-EID-8：虚拟地址冲突错误

问题现象：

当出现 GPU 卡的虚拟内存管理错误时，内核日志会打印如下错误信息：

```
METAX.Bxxxx.MC.ERROR bo xxx va xxx-xxx conflict with yyy-yyy
METAX.Bxxxx.MC.ERROR failed to map va xxx in vm, rrr
METAX.Bxxxx.MC.ERROR failed to map bo to xvm, rrr
```

打印信息解读：

Log 里面打印出两段有重叠区域的 VA 区间范围 (xxx-xxx 和 yyy-yyy)，rrr 为返回值。

可能原因以及建议操作：

UMD 申请内存使用的 VA 有重叠，KMD 会检测每个进程各 GPU VA 到 PA 的 entry 映射是否与其他 entry 有重复（包括部分重复），一旦检测到 GPU VA 到 PA 存在多重映射，KMD 日志就会记录冲突地址，标记当前映射失败，并返回错误码。这种问题需要联系硬件厂商解决。

4.9 METAX-EID-9：HBM 显存耗尽

问题现象：

当 GPU 应用程序使用的 HBM 显存空间超过 GPU 卡的 HBM 显存总大小时，会打印如下错误信息：

```
METAX.Bxxxx.MC.ERROR vram try to use (xxx kB) beyond total memory (xxx kB),
failed
```

打印信息解读：

- try to use (xxx kB)：已使用的 HBM 显存空间
- total memory (xxx kB)：GPU 卡 HBM 显存总大小

可能原因以及建议操作：

- GPU 应用程序使用了过多的 HBM 显存，GPU 应用程序编程模型是否合理，减少对 HBM 显存过多申请，或者及时释放没用的 HBM 显存。
- GPU SDK 存在内存管理相关问题，可以联系硬件厂商解决。

4.10 METAX-EID-10：系统内存耗尽

问题现象：

当 GPU 应用程序使用的系统内存空间超过 GPU 驱动对系统内存（XTT）限制总大小时，会打印如下错误信息：

```
METAX.Bxxxx.MC.ERROR (xtt + userptr) memory (xxx bytes) beyond limit (xxx bytes)
```

打印信息解读：

- memory (xxx bytes)：已使用的系统内存总空间
- beyond limit (xxx bytes)：当前 GPU 驱动对系统内存限制总大小

可能原因以及建议操作：

- GPU 应用程序使用了过多的系统内存，GPU 应用程序编程模型是否合理，减少对系统内存过多申请，或者及时释放没用的系统内存。
- GPU SDK 存在内存管理相关问题，可以联系硬件厂商解决。

4.11 METAX-EID-11：INT 中断溢出错误

问题现象：

GPU INT 中断消息发送到 host 后，GPU 内核驱动将这些 INT 中断消息缓存到 FIFO 队列失败后，会打印如下错误信息：

```
METAX.Bxxxx.INT.ERROR xxx add cookie client x src x to fifo failed
```

打印信息解读：

- client：触发 INT 中断的 IP 编号
- src：触发 INT 中断的源编号

可能原因以及建议操作：

因为 GPU 触发 INT 中断和 CPU 处理中断的过程是异步的，当 GPU 触发 INT 中断消息的速度超过 CPU 处理中断的速度，会出现 GPU 驱动的中断 FIFO 队列溢出问题，可以尝试以下操作：

确认 CPU 是不是运行在最高频率，通过 `cat /sys/devices/system/cpu/cpu0/cpufreq/scaling_governor`，确认是否为 performance 模式；如果不是，则执行 `echo performance > /sys/devices/system/cpu/cpu<x>/cpufreq/scaling_governor` 命令切换为 performance 模式，其中 <x> 为所有 CPU 的 index。

如果还有问题，请联系硬件厂商。

4.12 METAX-EID-12: PCIe AER Error

问题现象：

当 host 与 GPU 卡 PCIe 通讯链路存在某些异常的时候（比如常见的 Completion Timeout、Unsupported Request 与 ECRC Error 等），GPU 驱动通常会在内核记录如下 PCIe AER 错误信息：

```
METAX.Bxxxx.PCI.ERROR detect uncorrectable AER, xxxx
```

打印信息解读：

xxxx：表示 PCIe uncorrectable AER 错误信息，与 PCIe spec 规范 AER 定义一致。

可能原因以及建议操作：

- PCIe 链路不稳定，可以尝试重新插拔 GPU 卡或更换 PCIe 卡。
- 服务器 PCIe 相关配置不对

这两种情况都需要联系硬件厂商和服务器厂家分析 PCIe AER 产生的原因。如果是偶尔发生这种情况，可以尝试 warm reset 或重启服务器进行恢复。

4.13 METAX-EID-13: PCIe 通讯断开

问题现象：

在 GPU 运行过程中，如果 PCIe 通讯断开，则 GPU 驱动检测到 GPU 卡无法正常访问（读到全 f），并会打印如下信息：

```
METAX.Bxxxx.ATU.ERROR get vm fault status with full f
```

打印信息解读：

日志指示 GPU 驱动访问 GPU 卡的特定 MMIO 寄存器返回全 f 异常，表明 PCIe 通讯已经断开。

可能原因以及建议操作：

- 可能是前面执行业务异常，导致 GPU 通讯断开，可从之前最开始出现错误的地方排查。
- 可能是硬件问题，可以尝试更换 PCIe slot 或 GPU 卡进行测试。

这两种情况都需要联系硬件厂商和服务器厂家分析 PCIe 通讯断开产生的原因。如果是偶尔发生这种情况，可以尝试 warm reset 或重启服务器进行恢复。

5. MXCD 常见错误

表 5-1 MXCD 常见错误列表

编号	错误信息	错误说明
MXCD-EID-1	destroy ce rb failed, logic rb_id:xxx, pipe_id:xxxs	CE Dequeue Error
MXCD-EID-2	device IDs array empty	IOCTL 操作设备失败

5.1 MXCD-EID-1: CE Dequeue Error

问题现象：

在进程主动或强制结束的过程中出现挂起的情况，MXCD 也会打印 destroy ce rb failed 错误信息。

该问题与 4.3 METAX-EID-3: CE Dequeue Timeout 属于同样的问题，典型的日志记录如下：

```
[...] MXCD.Bxxxxx.RINGBUF.ERROR destroy ce rb failed, logic rb_id:x, pipe_id:x
```

打印信息解读：

- rb_id: dequeue 失败的 queue 编号
- pipe_id: dequeue 失败的 queue 所在的 pipe 编号

可能原因以及建议操作：

GPU 内部任务调度失败，报 bug 给硬件厂商，然后尝试 warm reset 或重启服务器进行恢复。

5.2 MXCD-EID-2: IOCTL 操作设备失败

问题现象：

当用户通过 IOCTL 系统调用向 GPU KMD 驱动请求服务，KMD 无法成功完成该请求会报 IOCTL 失败信息，格式如下：

```
[...] MXCD.Bxxxxx.IOCTL.ERROR device IDs array empty
```

可能原因以及建议操作：

IOCTL 失败一般是软件上的参数非法或条件不满足导致的，定位该问题需要结合实际环境配置和报错信息综合判断原因。

6. MXGVM 常见错误

表 6-1 MXGVM 常见错误列表

编号	错误信息	错误说明
MXGVM-EID-1	Failed to mmap xxx	映射 GPU 卡 PCIe BAR 失败
MXGVM-EID-2	Chip preinit, SMPx timeout	GPU 卡初始化由于 SMP 通信超时而失败
MXGVM-EID-3	The vbios version xxx is below the driver's minimum compatible version xxx	GPU 卡 VBIOS 与 GPU 驱动版本不兼容
MXGVM-EID-4	Not support SRIOV	GPU 卡不支持 SRIOV 虚拟化
MXGVM-EID-5	Not enough MMIO resources for SRIOV BARx	GPU 卡 SRIOV BAR 空间不够
MXGVM-EID-6	Failed to properly enable virt, ret=x	开启 GPU 卡 SRIOV 虚拟化失败
MXGVM-EID-7	wait reset status x timeout	GPU 卡 FLR 复位失败
MXGVM-EID-8	get fault status with full f, device may be disconnected	GPU 卡 PCIe 通讯断开
MXGVM-EID-9	die x atul2 x hwip xxxx xpid x vfid x fault perm x va xxxx	GPU 卡 ATU Fault

6.1 MXGVM-EID-1: 映射 GPU 卡 PCIe BAR 失败

问题现象：

在 GPU 卡初始化过程中，映射 GPU 卡 PCIe BAR 空间到 CPU 地址空间失败，会打印如下错误信息：

```
[...] MXGVM.Bxxxx.BASE.ERROR Failed to map mmio
```

可能原因以及建议操作：

服务器的 MMIO 空间不够，确认服务器 BIOS 是否有 MMIO 空间相关配置选项，如果有，建议将 MMIO 空间配置到最大。

6.2 MXGVM-EID-2: GPU 卡初始化由于 SMP 通信超时而失败

问题现象：

在 GPU 卡初始化过程中，GPU 驱动与 GPU 卡 SMP 通信超时，会打印如下错误信息：

```
[...] MXGVM.Bxxxxx.SOC.ERROR Chip preinit, SMPx timeout
```

可能原因以及建议操作：

1. 初步排查，使用 `lspci -s DBDF -x` 命令测试 GPU 卡的 PCIe configure space 能否继续访问（DBDF 为故障卡的 PCIe DBDF）。
 - 如果命令返回全 f 数据，则表示 GPU 卡 PCIe 基本通路有问题，建议按照步骤 2 排查硬件问题。
 - 如果命令返回有效值，则表示 GPU 卡 PCIe 基本通路正常，建议按照步骤 3 排查 GPU VBIOS 软件问题。
2. 排查硬件相关问题，比如 GPU 卡与服务器的接插件的连接是否良好，断电后重新插拔 GPU 卡、更换 PCIe slot 或更换 GPU 卡。
3. 排查 GPU VBIOS 软件问题，通过 mx-report 工具抓取 GPU 卡状态信息，然后报告给硬件厂商。

6.3 MXGVM-EID-3: GPU 卡 VBIOS 与 GPU 驱动版本不兼容

问题现象：

在 GPU 卡初始化过程中，检查到 GPU 卡 VBIOS 固件版本与当前运行驱动版本不兼容，会打印如下错误信息：

```
[...] MXGVM.Bxxxxx.SOC.ERROR The vbios version xxx is below the driver's minimum compatible version xxx
```

可能原因以及建议操作：

GPU 卡的 VBIOS 固件版本与当前 GPU 驱动不兼容，可以尝试升级与 GPU 驱动兼容的 VBIOS 版本。

6.4 MXGVM-EID-4: GPU 卡不支持 SRIOV 虚拟化

问题现象：

在 GPU 卡初始化过程中，检查到 GPU 卡不支持 SRIOV 虚拟化功能时，会打印如下错误信息：

```
[...] MXGVM.Bxxxxx.BASE.ERROR Not support SRIOV
```

可能原因以及建议操作：

GPU 卡不支持 SRIOV 功能，使用 `lspci -s DBDF -vvv` 命令查看 GPU 卡是否有 SRIOV 能力（DBDF 为故障卡的 PCIe DBDF），如果不支持 SRIOV 能力，则需要升级为支持 SRIOV 的 VBIOS 版本。

6.5 MXGVM-EID-5: GPU 卡 SRIOV BAR 空间不够

问题现象:

在 GPU 卡初始化过程中, 检查到 GPU 卡的 SRIOV MMIO 空间不足后, 会打印如下错误信息:

```
[...] MXGVM.Bxxxxx.BASE.ERROR Not enough MMIO resources for SRIOV BARx
```

可能原因以及建议操作:

- 服务器 BIOS 是否开启 SRIOV 相关配置, MMIO 空间是否开启到最大。
- 确认 GPU VBIOS 版本是否与服务器匹配, 可以尝试升级为 SRIOV MMIO 空间需求更小的 VBIOS 版本。

6.6 MXGVM-EID-6: 开启 GPU 卡 SRIOV 虚拟化失败

问题现象:

在 GPU 卡初始化过程中, 在开启 SRIOV 虚拟化功能失败后, 会打印如下错误信息:

```
[...] MXGVM.Bxxxxx0.BASE.ERROR Failed to properly enable virt, ret=x
```

可能原因以及建议操作:

- 服务器 BIOS 是否开启 SRIOV 相关配置, 如果有 SRIOV 配置选项, 则将 SRIOV 配置打开。
- 用 mx-report 工具收集 VBIOS 运行日志, 并报告给硬件厂商。

6.7 MXGVM-EID-7: GPU 卡 FLR 复位失败

问题现象:

在对 GPU 卡 FLR 复位失败后, 会打印如下错误信息:

```
[...] MXGVM.Bxxxxx.MXIOV.ERROR wait reset status x timeout
```

可能原因以及建议操作:

可能是 GPU 硬件或 VBIOS 固件问题, 用 mx-report 工具收集 VBIOS 运行日志, 并报告给硬件厂商。

6.8 MXGVM-EID-8: GPU 卡 PCIe 通讯断开

问题现象:

在 GPU 运行过程中, 如果 PCIe 通讯断开, 则 GPU 驱动检测到 GPU 卡无法正常访问 (读到全 f), 会打印如下错误信息:

```
[...] MXGVM.Bxxxxx.ATU.ERROR get fault status with full f, device may be disconnected
```

可能原因以及建议操作：

- 可能是前面执行业务异常，导致 GPU 通讯断开，可从之前最开始出现错误的地方排查。
- 可能是硬件问题，可以尝试更换 PCIe slot 或 GPU 卡进行测试。

6.9 MXGVM-EID-9: GPU 卡 ATU Fault

问题现象：

在 GPU 运行过程中，GPU 卡 ATU 检测到存在非法 VA 地址访问（ATU Fault），会打印如下错误信息：

```
[...] MXGVM.Bxxxxx.ATU.ERROR die 0 atul2 1 hwip DMAW0 xpid 1 vfid 0 vf 1 fault  
XVM_NO_PERM perm 2 va 0xa06000000
```

此条错误信息每个字段定义如下：

- die: ATU fault 报错的 die ID (0: master, 1: slave)
- atul2: ATU fault 报错的 ATUL2 ID (范围 0-1)
- hwip: ATU fault 报错的硬件 IP
- xpid: ATU fault 报错的 XPID (范围 0-15)
- vfid: ATU fault 报错的 VF 卡 ID
- vf: ATU fault 报错是否为 VF 卡 (0: PF 卡, 1: VF 卡)
- fault: ATU fault 报错的内部原因
- perm: ATU fault 报错的地址页表项权限配置
- va: ATU fault 报错的 VA (虚拟地址) 地址

以上述错误信息为例，表示在 die 0 (master die) 上的 ATUL2 0 (编号 0 的 ATUL2) 上检测到 hwip DMAW0 产生了 ATU fault，同时，此错误是发生在 VF 0 卡上的 XPID 1 地址空间，出错原因是 XVM_NO_PERM，出错 VA 地址为 0xa06000000。

可能原因以及建议操作：

业务程序可能存在问题，排查应用程序中对 GPU 卡的 VA 访问地址是否合法。

声明

版权所有 ©2024 沐曦集成电路（上海）有限公司。保留所有权利。

本文档中呈现的信息属于沐曦集成电路（上海）有限公司和/或其附属公司（以下统称为“沐曦”），非经沐曦事先书面许可，任何实体或个人均不得获得本文档的副本，且无权以任何方式处理本文档，包括但不限于使用、复制、修改、合并、出版、发行、销售或传播本文档的部分或全部。

本文档内容仅供参考，不提供任何形式的、明示或暗示的保证，包括但不限于对适销性、适用于任何目的和/或不侵权的保证。在任何情况下，沐曦均不对因本文档引起的、由本文档造成的、或与之相关的任何索赔、损害或其他责任负责。

沐曦保留自行决定随时更改、修改、添加或删除本文档的部分或全部的权利。沐曦保留最终解释权。

沐曦、MetaX 和其他沐曦图标是沐曦的商标。本文档中提及的所有其他商标和商品名称均为其各自所有者的财产。