

## 华为Tecal ES3000产品FAQ

<b>Issue</b>	1.00
<b>Date</b>	2012-11-27



## 华为技术有限公司

华为技术有限公司为客户提供全方位的技术支持，用户可与就近的华为办事处联系，也可直接与公司总部联系。

## 华为技术有限公司

地址：                深圳市龙岗区坂田华为总部办公楼                邮编：518129

网址：                <http://www.huawei.com>

客户服务电话：  0755-28560000  8008302118

客户服务传真：  0755-28560111

客户服务邮箱：  [support@huawei.com](mailto:support@huawei.com)

### 版权所有 © 华为技术有限公司 2006。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

### 商标声明



HUAWEI 和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

### 注意

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。



## 目录

常见问题 .....	4
1 SUSE Linux Enterprise 11 Service Pack 1操作系统下执行Fio、Vdbench等基准测试时出现性能突降、CPU占用率突增的情况 .....	4
2 升级固件或者修改容量后重启(reboot)，系统无法找到SSD卡设备 .....	4
3 Windows Server 2008 R2操作系统下在SSD卡上新建分区、创建文件系统，并在分区间大量拷贝文件过程中卸载驱动，操作挂住，磁盘不可用 .....	4
4 Windows Server 2003 R2操作系统下在图形界面使用默认分区方式新建分区并创建NTFS文件系统，随机读写性能较差 .....	5
5 ES3000 SSD卡在RH2285、RH2288服务器的PCIe x16槽位上性能只有十分之一 .....	5
6 Windows 2003 R2操作系统下建立分区后格式化速度较慢，CPU占用率较高 .....	6
7 Windows 2003 R2/2008 R2操作系统下启动之后找不到SSD设备 .....	6
8 ESX 4.1和ESXi 5.0操作系统下，ES3000规格为2.4TB容量的卡只显示为2TB .....	7
9 ES3000 SSD卡I/O过程中出现性能下降为零的情况 .....	8
10 ES3000 SSD卡如何修改冗余度？ .....	8
11 ES3000 SSD卡可允许的最大坏块率为多少？如何查看SSD卡状态是否正常？ .....	8
12 如何查看ES3000 SSD卡日志？错误类型有哪些？ .....	9
13 如何批量管理SSD卡？ .....	10
14 SSD卡是否支持热插拔？ .....	10
15 VMware环境下VMFS模式和直通模式混用导致服务器挂住 .....	10
16 服务器重启之后找不到部分或全部SSD设备 .....	11
17 ESX 4.1环境中，在虚拟机中起多个线程进行大压力IO应用，存在虚拟机假死现象 ...	11
技术要点 .....	11
1 SSD卡如何实现掉电保护？ .....	11
2 SSD卡采取哪些技术保证数据的可靠性？ .....	11

## 常见问题

### 1 SUSE Linux Enterprise 11 Service Pack 1 操作系统下执行 Fio、Vdbench 等基准测试时出现性能突降、CPU 占用率突增的情况

答复：SLES 11 SP1 默认内核存在BUG。

该BUG由Suse自行修改内核引入：在mm/filemap.c \_\_lock\_page中采用prepare\_to\_wait，而没有采用prepare\_to\_wait\_exclusive，导致wake\_up时所有等待同一page的线程都被wake up，线程进入busywait。特别是在相应的page在buffer中时（对设备进行cache读，例如fdisk -l读分区表，或者dd读），\_\_lock\_page等待sync\_page，cpu一直忙等待，导致占用大量cpu，导致IO性能降低。

解决措施：需要修正内核代码重新编译内核。将mm/filemap.c的\_\_lock\_page、\_\_lock\_page\_killable、\_\_lock\_page\_nosync、unlock\_page等函数恢复成标准2.6.32内核的实现方式，重新编译安装内核即可。

### 2 升级固件或者修改容量后重启(reboot)，系统无法找到 SSD 卡设备

答复：由于升级固件或者修改完容量后重启时，固件会重新加载以使新的设置生效，而加载需要时间，可能加载完毕时，系统扫描PCIE设备的时机已经过去，从而导致进入系统后无法找到SSD卡设备。这时，只需再重启(reboot)一次即可。当然，用户升级完固件或者修改容量后，也可以采取下电一次再上电的方式来使新的固件或者容量设置生效。

### 3 Windows Server 2008 R2 操作系统下在 SSD 卡上新建分区、创建文件系统，并在分区间大量拷贝文件过程中卸载驱动，操作挂住，磁盘不可用

答复：在分区间大量拷贝文件同时卸载驱动，容易造成windows文件系统驱动和卷管理驱动争抢SSD设备的使用权，从而引起系统异常。在系统日志中可以见到大量的文件系统以及卷管理系统异常日志，并引起分区信息丢失。建议用户在拷贝文件或在进行其它I/O操作时不要进行卸载驱动的操作。如果引起此异常，可以通过重启(reboot)系统进行解决，但此时待拷贝的文件可能会丢失。

#### 4 Windows Server 2003 R2 操作系统下在图形界面使用默认分区方式新建分区并创建 NTFS 文件系统，随机读写性能较差

答复: Windows 2003 R2 默认分区方式, 分区offset一直是从63扇区(63X512B=31.5KB)开始, 那么就会造成用户的第一个数据的前4KB会存放在系统"逻辑扇区"的31.5KB~35.5KB间。而SSD的扇区大小(page)一般为4K或4K的整数倍, 这样就会造成后面所有的数据都会落在2个物理扇区(page)之间。在写入的时候就需要进行读-改写操作(对SSD来说读-擦-写), 造成性能的下降。

解决措施: 使用系统自带的命令行工具diskpart进行分区, 步骤如下:

- 1、在虚拟磁盘管理器中找到SSD所在的磁盘序号, 例如是 1;
- 2、用系统管理员权限打开cmd命令行窗口;
- 3、执行diskpart命令;
- 4、在diskpart的命令提示后输入以下命令:

```
DISKPART>sel disk 1
```

- 5、输入以下命令创建起始扇区对齐的分区:

```
DISKPART>create partition primary align=64
```

6、待系统提示创建分区成功之后, 分配一个空白盘符, 也可以在虚拟磁盘管理器中的中进行选择:

```
DISKPART>assign letter=m
```

- 7、退出diskpart:

```
DISKPART>exit
```

- 8、重新进行格式化文件系统的操作,操作过程完成。

#### 5 ES3000 SSD 卡在 RH2285、RH2288 服务器的 PCIe x16 槽位上性能只有十分之一

答复: 此问题为RH服务器BIOS软件的一个BUG, 将在V035的BIOS版本上解决。

如果没有升级到该版本BIOS的条件, 可使用如下方案解决:

- 1、重启系统按ctrl + alt + 1(数字1)进入BIOS配置界面: Setup Utility;
- 2、找到并进入Advanced配置界面;
- 3、找到并进入SandyBridge 110;

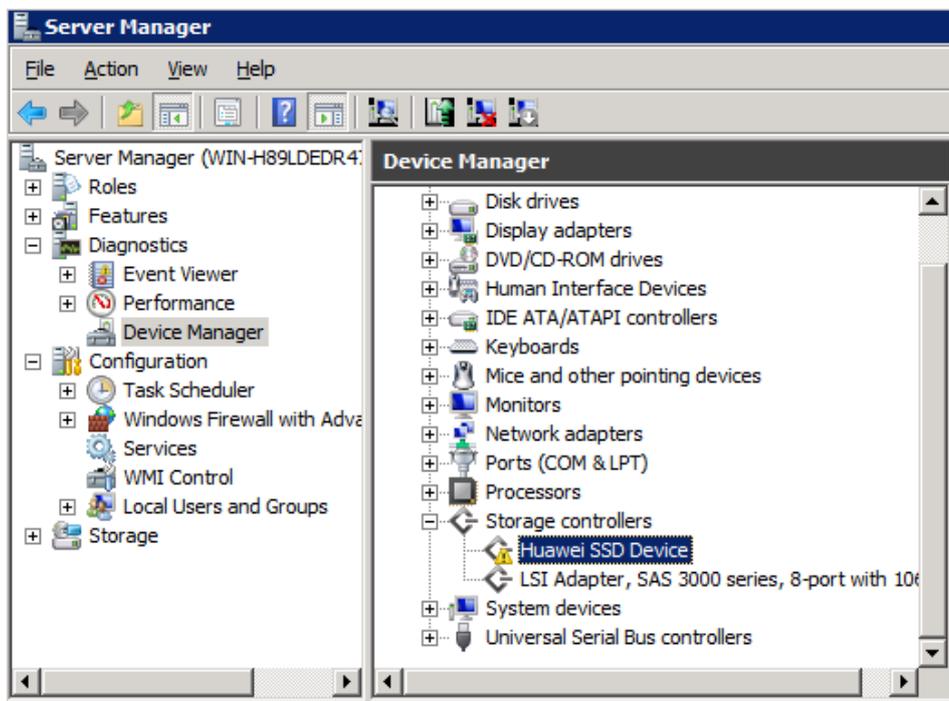
- 4、找到并进入SandyBridge 1100 Configuration（RH2288的16X所在的CPU）或者SandyBridge 1101 Configuration（RH2285的16X所在的CPU）；
- 5、进入之后会看到，“IOU1(110 PCIe Port3) <x16>” 这样一个选项，进入该选项；
- 6、进入之后，选择配置成x8x8；
- 7、保存BIOS配置并退出；
- 8、复位系统，修改完成。

## 6 Windows 2003 R2 操作系统下建立分区后格式化速度较慢，CPU 占用率较高

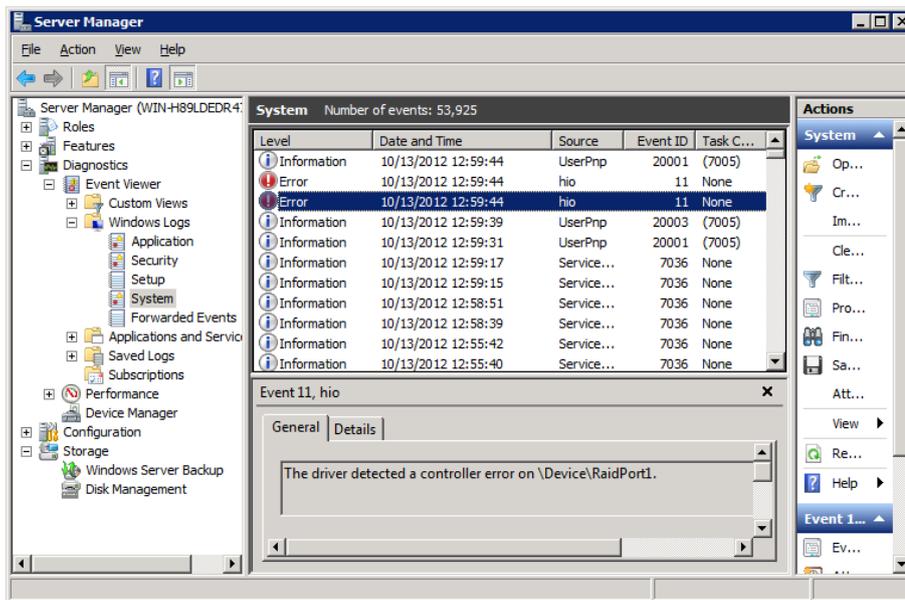
答复： Windows 2003 操作系统下发的格式化请求有特殊性，需要花较长时间进行转换处理，会导致格式化稍长，CPU占用率高，属于正常现象。在业务I/O过程中不会发生此现象。Windows 2008 R2操作系统中无此问题。

## 7 Windows 2003 R2/2008 R2 操作系统下启动之后找不到 SSD 设备

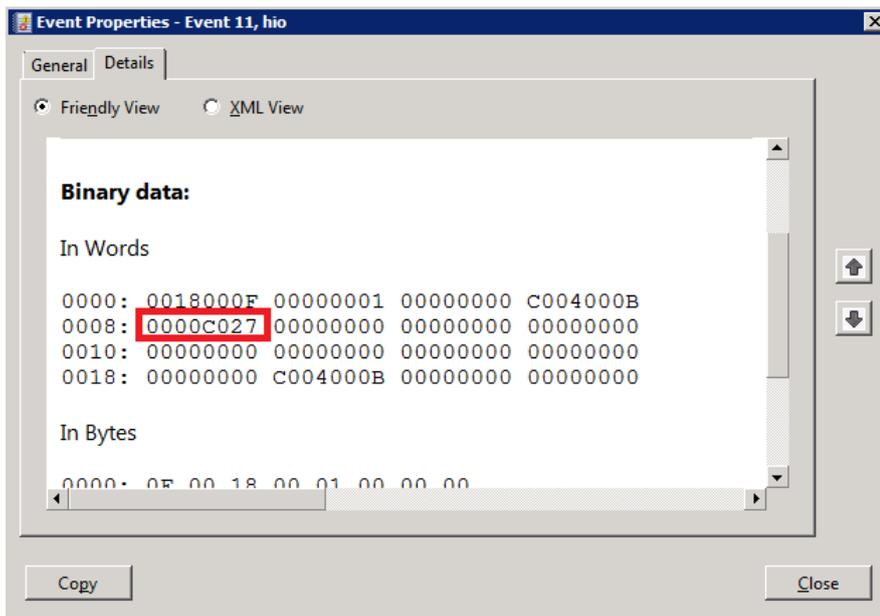
答复：在设备管理器中如下图所示，ES3000设备图标前有一个黄色的感叹号，表示驱动未能正常加载：



如下图所示，打开事件查看器，找到系统日志中source为hio的错误事件：



右键点击该错误事件，选择属性。系统弹出窗口，如下图所示，如果红框位置的错误号为(0x0000c027)或(0x0000c015)，表示由于系统异常，导致DMA内存分配失败，驱动未能正常加载：



解决方案：重启系统。如果重启还未解决问题，可能是服务器兼容性问题，请联系厂家。

## 8 ESX 4.1 和 ESXi 5.0 操作系统下，ES3000 规格为 2.4TB 容量的卡只显示为 2TB

答复：由于VMware ESX、ESXi使用SCSI的READ\_CAPACITY (10-byte)命令去查询设备的容量，该SCSI命令只能得到32-bit的值。因此，通过以下计算方式：

$$(2^{32}) = 4294967296$$

$$4294967296 - 1 = 4294967295$$

$$4294967295 * 512 = 2199023255040 = 2TB$$

所以，设备容量最终显示为2TB。

可参考vmware的官网说明：

[http://kb.vmware.com/selfservice/microsites/search.do?language=en\\_US&cmd=displayKC&externalId=3371739](http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=3371739)

## 9 ES3000 SSD 卡 I/O 过程中出现性能下降为零的情况

答复：查看系统日志，如果出现高温告警，请检查SSD卡是否超温。如果温度过高（芯片温度超过90℃），驱动会暂时降低I/O以保护SSD卡。建议按照服务器调速策略进行调速：

服务器型号	风速要求
Huawei Tecal RH2285 V2	80%
Huawei Tecal RH2288 V2	70%
DELL PowerEdge R720	100%

其它服务器型号请根据实际散热情况进行调速。

## 10 ES3000 SSD 卡如何修改冗余度？

答复：可以通过改小容量来增大SSD卡的冗余度，设置容量之前请停止I/O业务请求并备份SSD卡上的数据，然后通过驱动提供的工具重新设置容量（Linux、VMware环境下使用 hio\_capacity 接口、Windows环境下使用GUI工具）。

注意：修改容量会清除SSD卡里所有数据。

## 11 ES3000 SSD 卡可允许的最大坏块率为多少？如何查看 SSD 卡状态是否正常？

答复：Linux、VMware环境下可通过 hio\_info 命令查看SSD卡状态，Windows环境下可通过驱动提供的GUI工具查看。如果状态信息里出现“Warning”，则SSD卡存在异常，需要及时维修；如果出现“Fatal”，则说明SSD卡已无法正常使用，需要更换。

上述状态信息里包括当前SSD卡的最大坏块率，最大坏块率超过13%会报“Warning”告警，此时会影响SSD卡性能，而超过14%则可能会影响正常I/O，呈“Fatal”状态。

## 12 如何查看 ES3000 SSD 卡日志？错误类型有哪些？

答复：Linux、VMware环境下可通过 hio\_log 命令查看SSD卡日志，Windows环境下可通过驱动提供的GUI工具查看。日志的错误类型如下表所示：

级别	编码	事件	影响	备注
错误	0x0439	初始化扫描PBMT读取全部错误	部分数据的映射关系丢失，最多64MB数据无法找回	该表只列出了错误级别和警告级别的日志及其影响。其中数据出错的前提条件是ECC纠错失败和RAID恢复失败。数据丢失的前提条件
错误	0x043B	初始化扫描首页读取全部错误	部分数据的映射关系丢失，最多64MB数据无法找回	
错误	0x043C	初始化恢复指针，读取数据页错误	4KB的数据映射关系丢失，该数据无法找回	
错误	0x0440	初始化恢复PMT，读数据页错误	4KB的数据映射关系丢失，该数据无法找回	
警告	0x080b	磨损回读出错，Merge回读数据	4KB的数据出错且无法恢复	
错误	0x0850	cache命中数据错误	4KB的数据出错且无法恢复	
错误	0x0851	cache补齐回读出错	4KB的数据出错且无法恢复	
错误	0x0c2b	写失败回读出错	4KB的数据出错且无法恢复	
错误	0x1011	GC和CM读，RAID恢复失败	4KB的数据出错且无法恢复	
错误	0x10f1	IOR读不能恢复（raid不可恢复）	4KB的数据出错且无法恢复	
错误	0x1489	无效页计数超过最大无效页数	部分数据无法正常回收	
错误	0x1407	没有可用块，空闲链表为空	数据无法正常写入	
警告	0x1890	FPGA温度超过阈值（90℃）	FPGA芯片温度过高可能导致SSD工作异常，需要调高风扇转速，降低SSD板块温度	
警告	0x1891	FPGA温度恢复到正常范围	FPGA芯片温度回到正常工作	

		(85°C)	范围	是板卡
警告	0x0852	bus下发异常命令	SSD卡收到异常命令, PCIE或者SERDES等接口链路可能出现误码	异常掉电。
警告	0x107f	flash访问超时	一片flash颗粒有可能损坏, 可能会造成部分数据出错且无法恢复	
警告	0x1483	RAM状态异常	可能造成数据回收异常	
警告	0x1892	超级电容异常告警	异常掉电时可能会有部分数据丢失	

### 13 如何批量管理 SSD 卡？

答复：目前可以在Linux下使用SNMP服务实现SSD卡的批量监控管理，使用方法如下：

- 1、确保OS安装SNMP服务；
- 2、安装hio\_snmp的rpm包；
- 3、修改 /etc/snmp/snmpd.conf 配置文件, 把末尾部分的 trap2sink 127.0.0.1:162 中的IP地址改为实际trap目标地址；
- 4、使用 service snmpd start 启动snmp服务；
- 5、使用 snmpwalk -v2c -c public localhost enterprises.2011.2.269 命令测试snmp服务状态；
- 6、根据 /usr/local/hio\_snmp/HW-HIO-MIB.txt 可以查询相应接口，以开发客户端应用。

### 14 SSD 卡是否支持热插拔？

答复：SSD卡需要插在服务器PCIe插槽中使用，不支持热插拔。

### 15 VMware 环境下 VMFS 模式和直通模式混用导致服务器挂住

答复：当ESX(i)服务器中有多个SSD设备时，建议不要将部分SSD设备配置为VMFS存

储容器、部分SSD设备配置为passthrough模式，否则，在SSD设备进行IO时，进行模式转换，可能会造成ESX(i)服务器挂死。

## 16 服务器重启之后找不到部分或全部 SSD 设备

答复：SSD卡异常掉电之后初始化比较慢，可能出现系统启动完成后驱动还在等待SSD卡初始化的情况，此时是找不到SSD设备的，需要继续等待初始化完成。

## 17 ESX 4.1 环境中，在虚拟机中起多个线程进行大压力 IO 应用，存在虚拟机假死现象

答复：ESX (i) server 对虚拟机多线程的支持不完善，在虚拟机中起较多线程时，CPU会挂住。

# 技术要点

## 1 SSD 卡如何实现掉电保护？

答复：SSD卡是非易失性存储介质，掉电时只有cache中的部分数据未写入flash中。Huawei Tecal ES3000 SSD卡采用超级电容供电，在掉电时提供足够把cache中数据写入nand flash中，达到掉电保护的作用。

## 2 SSD 卡采取哪些技术保证数据的可靠性？

答复：ES3000 SSD卡采用多种先进的算法确保ES3000的高可靠性：

- 1、内嵌的ECC检错/纠错引擎和RAID5引擎（通道间）形成二维的检错/纠错机制，确保数据的高可靠性；
- 2、内置DATA Scrubbing引擎定时检测存储数据，提前预防数据错误的产生；
- 3、通道间使用Dynamic RAID算法，实现通道间的资源共享，确保在芯片坏块过多甚至是多个芯片故障的情况下均能正常工作；



4、内部实现冷热数据分类与管理，配合先进的磨损算法，最大程度上提升回收效率，降低写磨损，从而提升SSD的使用寿命；

5、数据掉电保护的设计确保在异常掉电情况下所有数据均能可靠的写入ES3000，避免数据丢失。