磁盘阵列状态实时监控的通用性解决方案

赵林海[1,2], 李晓风[1], 谭海波[1]

(1. 中国科学院 合肥物质科学研究院信息中心, 安徽 合肥 230031; 2. 中国科学院研究生院, 北京 100049)

摘 要: 针对当前磁盘阵列(redundant array of independent disks, RAID)商家众多,缺乏统一监控工具这一缺陷,引入了一种通用性解决方案。以简单网络管理协议为基础,使用NET-SNMP添加磁盘阵列状态的监控模块,同时使用CACTI对磁盘阵列进行实时监控,并在发生异常时通过邮件和短信报警。该方案选用应用最广泛的软、硬两种RAID5级别进行测试。测试结果表明,该方案可实现各种磁盘阵列的统一监控管理和异常报警。

关键词: 磁盘阵列; CACTI; 网络管理; 监控; 报警

中图法分类号: TP333.3 **文献标识码**: A **文章编号**: 1000-7024 (2011) 02-0517-03

Universal solution for disk array status real-time monitoring

ZHAO Lin-hai[1,2], LI Xiao-feng[1], TAN Hai-bo[1]

(1. Information Center, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China; 2. Graduate University, Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Contrary to the circumstance of numerous RAID (redundant array of independent disks) producers, but lack of a unified monitoring tool, introduce a universal solution. Based on the SNMP (simple network management protocol), append the disk array status monitoring module with NET-SNMP, and then CACTI can perform real-time monitoring for the disk array, give the abnormal alarm through Emails and mobile messages. We choosed SoftRAID5 and HardRAID5, two of the most widely used RAID for the test and verify, the results shown that the solution can take an unified monitoring management and abnormal alarm for various disk arrays.

Key words: RAID; CACTI; SNMP; monitoring; warning

0 引 言

随着企业信息化发展,对数据的存储要求越来越高。虽然RAID(redundant array of independent disks)[1]可以提供较高的存储可靠性,但是如果不能及时发现硬盘故障,就有可能造成磁盘阵列崩溃,给企业带来巨大的损失。当前主流厂商的硬RIAD都配备状态管理工具,但大多数的管理工具只提供远程监控的功能,需要定时地查看硬RAID的状态。虽然有少数的管理工具可以实现主动报警,但是这些工具需要付费才能使用,而且一般只支持Windows操作系统。而对于软RAID[2]的状态,通常只能依赖命令来查看。由于各个厂商的管理工具相互间存在兼容性问题,很难对不同RAID进行统一管理。为此,设计监控模块采集RAID的状态信息,对于不同厂家的RAID产品,只需要简单地修改监控模块即可移植使用。然后利用CACTI[3]查询这些监控模块获得RAID的状态信息,从而实现对各种RAID设备进行统一监控和异常预警,提高管理的效率和数据存储的可靠性。

1 监控系统实现

1.1 相关技术

1.1.1 RAID

RAID技术诞生于1987年,由美国加州大学伯克利分校提出。它是一种把多块独立的磁盘(物理磁盘)按不同方式组合起来形成一个磁盘组(逻辑磁盘),从而提供比单个磁盘更高地存储性能,并实现数据冗余的技术。在这一组磁盘中,数据按照不同的算法分别存储于每块磁盘上从而达到不同的效果这样就形成了不同的RAID级别(RAID LEVEL)。

根据实现方式的不同,RAID可以分为软RAID、硬RAID和半软半硬RAID。而根据可靠性级别划分,RAID的级别有1到6,外加级别10,共7种。对于当前通用的RAID5来说,当有一块磁盘发生故障时,可以利用剩下的数据和相应的奇偶校验信息去恢复被损坏的数据。如果有两块以上磁盘同时发

收稿日期: 2010-02-27; **修订日期**: 2010-04-29。

基金项目: ITER国际高速专用数据网基金项目(2008GB111000)。

作者简介: 赵林海(1985-),男,广西大新人,硕士研究生,研究方向为系统监控与网络管理; 李晓风(1966-),男,安徽砀山人,博士,研究员,研究方向为网络管理与计算机自动控制; 谭海波(1976-),男,安徽泾县人,硕士,副研究员,研究方向为计算机网络应用。

E-mail: zhaolinhai@hfcas.ac.cn

生故障，数据将不能恢复。

1.1.2 CACTI

CACTI是一套基于PHP，MySQL，SNMP[4](simple network management protocol)及RRDTool[5]开发的系统和网络监测图形分析工具，通过轮询的方式从代理设备采集数据，然后使用RRDTool绘制图像。

1.1.3 NET-SNMP[6]

NET-SNMP是开放源代码的SNMP软件，用来开发SNMP应用程序的程序库，它支持SNMP v1，SNMP v2c与SNMP v3，并可以使用IPV4及IPV6。通过NET-SNMP在被监控设备添加自定义的代理模块，即可实现被监控设备性能数据的采集。

CACTI和NET-SNMP在系统和网络监控领域应用非常广泛，除了能够监控网络带宽利用率、CPU、内存及磁盘等基本信息外，还可以通过编写程序代码扩展自定义的监控指标。另外，在CACTI中设置状态参数阈值实现主动预警的功能。这两个软件均为开源软件，可以免费地使用它们来搭建一个功能强大的监控系统。

1.2 系统环境

文章以监控RAID5为例，由于软RAID5和硬RAID5的性能指标不相同，因此监控的时候需要分开考虑。实验系统由两部分组成，分为Agent端和Manager端。Agent端包括一台采用软RAID5的主机和一台采用硬RAID5的主机。Manager端为运行CACTI监控平台的主机，负责采集Agent端的数据。3台主机都基于Linux操作系统，均需要安装NET-SNMP软件。其中，软RAID5由6块通用的SCSI磁盘组成，通过操作系统设置实现。硬RAID5由4块惠普公司的RAID卡组成，磁盘阵列控制器型号为Smart Array E200i。

1.3 系统结构与实现

监控系统分为一个Manager端和多个Agent端，如图1所示。系统是基于SNMP协议实现的，而在SNMP协议通信中，被监控设备的各种状态信息是通过对象标识符OID (object identifiers)[7]来识别。由于软、硬RAID5状态标识差异较大，需要分开考虑。因此，系统中定义了两个私有的OID，其中OID:.1.3.6.1.4.1.1113用于标识软RIAD5状态。OID:.1.3.6.1.4.1.1114用于标识硬RAID5状态，并在这两个OID下添加OID子节点用于标识具体的状态参数。利用NET-SNMP软件工具将定义的私有OID添加到信息库里面，Manager端就可以通过Agent的IP和具体的OID查询RAID5相应的状态信息[8]。

在Agent端，分别编写软、硬RAID5的监控模块。脚本程序周期性地采集RAID5的状态信息，并将具体的状态值赋予对应的OID节点。当它接收到Manager端查询某一OID的请求时，就返回该OID对应的数值给Manager端。

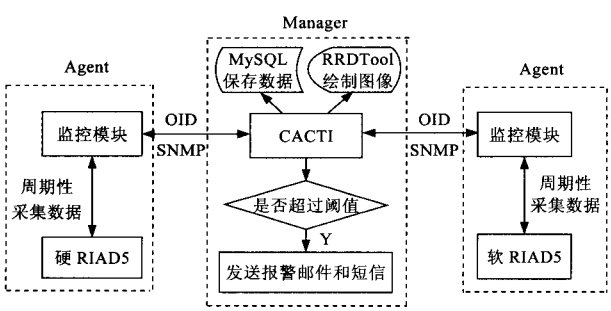

图1 系统结构

在Manager端，CACTI周期性的向各个Agent发送RAID状态查询请求。为了方便管理，系统统一设定RAID5状态阈值为0，当RAID5状态值为0时表示磁盘工作正常，当它大于0时表示有磁盘发生故障。当CACTI采集到数据之后会将数据存储到MySQL，并用RRDTool绘制图像，同时，将采集到的RAID5状态值与设置的阈值进行比较，如果发现其大于阈值，CACTI就会立即发送报警邮件和短信给系统管理员。

当系统中有新的Agent需要添加时，如果是RAID5，只需要将监控模块移植到新的Agent上。如果采用其它级别的RAID，只需要对监控模块做简单地修改即可。所以，系统具有良好的扩展性和移植性。

1.4 RAID状态数据采集

1.4.1 软RAID5状态数据采集

文章使用变量SoftRaid5Stauts标识软RAID5的状态。软RAID5的状态信息可以在文件/proc/mdstat获得[9]。这个文件实时地收集系统软RAID的状态信息，通过它可以判断软RAID5有没有发生故障。文章中的软RAID5有6块SCSI磁盘组成，当前/proc/mdstat内容如下：

```
/*==========================================*/
Personalities: [raid6] [raid5] [raid4]
md0: active raid5 sdf1[5] sde1[4] sdd1[3] sdc1[2] sdb1[1] sda1[0]
4883799040 blocks level 5, 256k chunk, algorithm 2 [6/6]
[UUUUUU]
unused devices: <none>
/*==========================================*/
```

从查询结果可以得到软RAID5的当前状态，其中“[UUUUUU]”的一个字符“U”表示一块磁盘的当前状态为“UP”，即正常状态。从查询结果可知，当前磁盘阵列组的6块磁盘均正常工作，如果其中的“U”变成“_”，则表明该磁盘发生了故障。监控脚本程序会周期性地查看这个信息，并将结果赋予对应的OID，正常时赋予0，异常时赋予1。此时，CACTI监控平台就可以通过OID获得软RAID5的状态值。

1.4.2 硬RAID5状态数据采集

与软RAID5不同，硬RAID5的状态信息不能通过查看文件/proc/mdstat获得，需要安装RAID卡厂商的阵列管理工具。这些阵列管理管理工具可以在厂商的官方网站获得，根据控制器的型号选择相应的版本下载安装。在安装惠普公司的LINUX命令行版阵列配置管理工具后，就可以使用hpacucli命令查看磁盘阵列的当前状态信息。

```
/*==========================================*/
# /usr/sbin/hpacucli ctrl all show status
Smart Array E200 in Slot 4
  Controller Status: OK
  Cache Status: OK
  Battery/Capacitor Status: OK
/*==========================================*/
```

从查询结果可知，当前磁盘阵列使用插槽4，控制器、高速

缓存和电源的状态均正常。监控程序通过调用 system 函数将 hpacucl 命令语句的结果写入文件中，然后再读取文件获得当前磁盘阵列的状态。程序代码如下：

```
/*=============================================*/
system("/usr/sbin/hpacucli ctrl all show status >
RAIDBasicStatusInfo.txt");
system("/usr/sbin/hpacucli ctrl slot = 4 show config detail >
RAIDCardDetailInfo.txt");
/*=============================================*/
```

上面代码中，第 1 行代码的作用是将当前磁盘阵列的基本信息写入 RAIDBasicStatusInfo.txt 文件中，通过它可以查询当前磁盘阵列使用的插槽 slot，以及参数变量 *ControllerStatus*、*CacheStatus* 和 *BatteryCapacitorStatus* 这 3 个基本信息的状态值。第 2 行代码的作用是将插槽 4 当前的详细信息写入 RAIDCardDetailInfo.txt 文件。通过查看这个文件即可获得 RAID 卡当前的详细状态信息，包括电源的数量和每块 RAID 卡的当前状态。硬 RAID 的参数变量及其对应的信息如表 1 所示。监控脚本程序周期性地读取 RAIDBasicStatusInfo.txt 文件和 RAIDCardDetailInfo.txt，将获取到的状态值分别赋予对应的 OID。此时，CACTI 监控平台就可以通过 OID 获得硬 RAID5 的所有状态值。

表 1　状态参数

基本状态		RAID 卡状态	
ControllerStatus	控制器状态	*UpRaidCard*	正常 RAID 卡
CacheStatus	缓存状态	*DownRaidCard*	故障 RAID 卡
BatteryCapacitorStatus	电源状态	*HotBackupRaidCard*	热备 RAID 卡
BatteryCapacitorCount	电源个数		

2　软 RAID5 监控测试

在文中使用变量 SoftRaid5Stauts 标识软 RAID5 的状态，状态分为正常和故障两种情况，其中值为 0 表示正常，值为 1 表示故障。为了测试监控系统是否正常检测软 RAID5 的故障，大约在 9:05 时刻取下任意一块硬盘。从图 2 可以看出，变量 SoftRaid5Stauts 的值在 9:08 之前一直为 0，之后 SoftRaid5Stauts 值变为 1，说明有磁盘发生了故障。之所以有大约 3 分钟的延迟是因为 CACTI 默认采集数据的周期是 5 分钟。系统管理员也可以根据实际情况缩短或者增加 CACTI 采集数据的周期。

为了实现邮件报警，设置变量 SoftRaid5Stauts 的最大阈值为 0。如图 3 所示，绿色表示该参数的当前状态值没有达到阈值，处于正常状态。而红色则表示当前处于报警状态。从图 3 中可看出，当前 SoftRaid5Stauts 值为 1 已经大于最大的阈值 0，处于报警状态。此时，CACTI 就会发送报警邮件给相应的系统管理员，告知某一台主机的软 RAID5 发生了故障。

3　硬 RAID5 监控测试

硬 RAID5 的监控与软 RAID5 在监控原理上是一样的，只是硬 RAID5 需要监控更多的参数指标。它包括基本状态和 RAID 卡状态两部分，可参见表 1。基本状态的监控如图 4 所示，基本状态包含 4 个参数指标。参数变量 ControllerStatus、

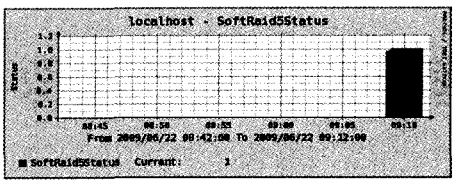

图 2　软 RAID5 状态

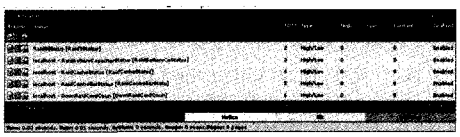

图 3　软 RAID5 报警状态

CacheStatus BatteryCapacitorStatus、BatteryCapacitorCount 分别表示控制器状态、高速缓存状态、电源状态和电源的个数。其中参数变量 ControllerStatus、CacheStatus 和 BatteryCapacitorStatus 需要设定阈值。这些参数变量的最大阈值均为 0。0 表示正常，大于 0 则表示硬 RAID5 基本状态有异常故障。从图 4 可看出大约从 8:56 时至 9:25，控制器电源数为 1，硬 RAID 的基本状态均正常。

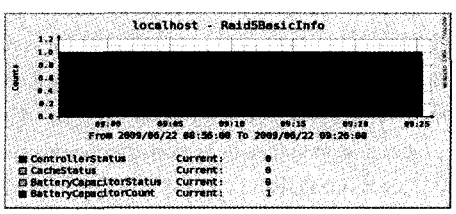

图 4　硬 RAID5 基本状态

RAID 卡的监控状态图如图 5 所示，总共有 4 个参数变量。参数变量 UpRaidCardCount、DownRaidCardCount、HotBackupRaidCardCount 和 TotalRaidCardCount 分别表示正常、故障、热备和阵列组 RAID 卡的数量。其中参数变量 DownRaidCardCount 需要设置状态阈值，最大阈值为 0。当它大于 0 时，CACTI 就会发送报警邮件给管理员。为了测试监控系统是否正常检测硬 RAID5 的故障，在 9:20 时刻将磁盘阵列其中一块正常的 RAID 卡更换成故障的 RAID 卡。大约在 9:23 时刻，DownRaidCardCount 值从 0 变为 1，表明有一块 RAID 卡发生故障了。此时，在硬 RAID5 报警状态中，DownRaidCardCount 已经从正常的绿色状态变成大于阈值的红色报警状态，CACTI 就会发送报警邮件和短信给管理员，通知其处理异常问题。

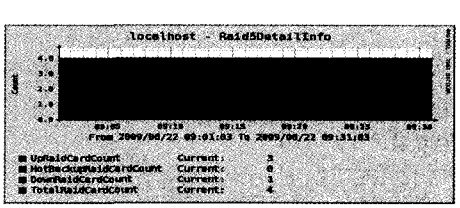

图 5　RAID 卡详细状态

（下转第 523 页）

频图像的压缩，大大降低传输所需带宽要求，有利于提高传输质量。对PC机上重建播放的红外摄像头视频图像的前后两侦作差值运算，即利用减背景法检测出红外摄像头上的热点信息，输出红外图像上热点的具体坐标位置。通过双目摄像机标定的方法来确定USB网络摄像头输出的图像上热点的位置，以得到直观的彩色图像上的热点位置信息。

4 试验结果与分析

既成功交叉编译完新内核、制作新的根文件系统之后，就可以依次将新的内核镜像zImage、新的根文件系统2.6.26_ramdisk.gz通过TFTP协议或者串口烧写到CM-X270核心板上，然后烧写完整的文件系统debian-image.jffs2到CM-X270核心板上，拷贝交叉编译生成的动态模块的文件夹到新的文件系统中。成功移植完操作系统之后，分别将图像采集模块、图像传输模块移植到CM-X270核心板上，然后在CM-X270核心板上运行图像采集和传输进程，在PC机上运行图像处理与显示进程，试验结果表明视频图像能够成功采集并且传输到远程PC机上，PC机上运行的图像处理与显示功能也达到了预期效果。

5 结束语

由于嵌入式系统具有高可靠性、低成本、体积小、功耗小等优点，本文提出了一种基于XScale架构与Linux操作系统的嵌入式图像采集与传输方案，图像采集采用Linux操作系统自带的V4L2视频编程应用程序接口，并不依赖硬件，因此可移植性比较强，只需在宿主机上交叉编译后，即可移植到相应目标板上。并且本系统中传输的图像数据格式是采集得到的原始图像数据，从而避免了在XScale上作复杂的数据格式转换运算，大大地节省了目标板的资源，可广泛应用于嵌入式图像处理领域。系统存在的不足之处，主要有如下几点，虽然Compulab公司提供了Linux-2.6.16及Linux-2.6.24内核版本的无线网卡驱动模块，但是并没有提供源码，短时间内还没有将对应Linux-2.6.26内核版本的无线网卡驱动模块移植到系统中，而是使用USB无线网卡取代XScale自带的无线网卡，完成无线网络传输功能。因此后期要完成的工作之一，就是要将相应无线网卡驱动移植到系统中。并且后期可以考虑采用MPEG-4图像压缩算法来压缩图像，以减少带宽，提高传输质量。

参考文献：

[1] 华清远见嵌入式培训中心.嵌入式Linux系统开发[M].北京:人民邮电出版社,2009.

[2] KarimYaghmour.构建嵌入式Linux系统[M].北京:中国电力出版社,2004.

[3] 科比特,鲁宾尼.Linux设备驱动程序[M].北京:中国电力出版社,2006.

[4] 林文森,李钟慎,洪健.基于ARM嵌入式图像处理系统设计与实现[J].福州大学学报(自然科学版),2008,36:13-16.

[5] 孟超,张曦煌.基于嵌入式系统的图像采集与传输设计[J].计算机工程与设计,2008,29(17):4414-4416.

[6] 宋凯,严丽平,甘岚.嵌入式图像处理系统的设计与实现[J].计算机工程与设计,2009,30(19):4368-4370.

[7] 杨树青,王欢.Linux环境下C编程指南[M].北京:清华大学出版社,2007.

[8] 陈亮.基于Video4Linux2的图像采集程序设计[J].微计算机信息,2009,7:65-67.

[9] 严新忠,陈雨.基于嵌入式ARM的图像采集与传输设计[J].国外电子测量技术,2009,11:57-59.

[10] Richard Stevens W,Bill Fenner,Andrew M Rudoff,et al.UNIX网络编程第1卷:套接口API[M].3版.北京:清华大学出版社,2006.

[11] 冯琪.小型无人直升机基于视觉的导航系统分析与设计[D].广州:华南理工大学,2005:9-10.

[12] 刘瑞祯,于仕琪.OpenCV教程[M].北京:北京航空航天大学出版社,2007.

[13] 冈萨雷斯.数字图像处理[M].2版.北京:电子工业出版社,2006.

(上接第519页)

4 结束语

本文以CACTI和NET-SNMP为平台，设计了一个能够实现对各种磁盘阵列设备进行统一监控和异常报警的方案。测试结果表明该改方案实现简单，效果良好，为磁盘阵列的监控标准提供一定的参考意义。另外，CACTI和NET-SNMP均为开源工具，可以极大降低企业的监控成本。今后将继续深入研究磁盘阵列的监控，进一步细化监控粒度，尤其在磁盘阵列I/O方面的监控，帮助系统管理员更早的发现异常状况，避免磁盘的损坏。

参考文献：

[1] 孙玉霞,陈火炎.RAID技术在iSCSI环境中的应用研究[J].计算机工程与设计,2006,27(24):4632-4634.

[2] 赵德平,史桂颖.基于Linux网络块设备和软RAID技术的网络镜像[J].计算机工程,2007,33(18):87-89.

[3] 岑锐坚.使用Cacti监测系统与网络性能[J].开放系统世界,2006(7):69-72.

[4] 李明江.SNMP简单网络管理协议[M].北京:电子工业出版社,2007:78-162.

[5] 吴刚.RRDTool性能优化的研究与实现[J].襄樊职业技术学院学报,2008,7(4):7-9.

[6] 殷卫红,耿新民.基于SNMP协议的网络管理实现技术[J].微计算机信息,2006(22):69-72.

[7] 区海平,寿国础.基于MIB定义的SNMP分析系统及实现[J].计算机应用,2009,29(1):38-41.

[8] 姜飞,史浩山,徐志燕,等.一种基于SNMP协议的主代理/子代理通信机制[J].计算机工程,2007,33(21):81-83.

[9] Mendel Cooper.Advanced bash-scripting guide,revision 6.0[EB/OL].http://www.tldp.org/LDP/abs/html/,2009.