

[PATCH] update Documentation/iostats.txt for 2.6

Source: <http://linux.derkeiler.com/Mailing-Lists/Kernel/2003-09/7586.html>

From: Rick Lindsley (ricklind_at_us.ibm.com)

Date: 09/30/03

To: linux-kernel@vger.kernel.org

Date: Tue, 30 Sep 2003 00:26:26 -0700

One typo corrected, and the references to 2.5 are minimized and mostly changed to 2.6.

```
diff -rup linux-2.6.0-t6/Documentation/iostats.txt linux-2.6.0-t6-doc/Documentation/iostats.txt
--- linux-2.6.0-t6/Documentation/iostats.txt Mon May 26 18:00:37 2003
+++ linux-2.6.0-t6-doc/Documentation/iostats.txt Tue Sep 30 00:23:07 2003
@@ -1,22 +1,22 @@
I/O statistics fields
-----
```

-Last modified 5/15/03

+Last modified Sep 30, 2003

-In 2.4.20 (and some versions before, with patches), and 2.5.45,
-more extensive disk statistics were introduced to help measure disk
+Since 2.4.20 (and some versions before, with patches), and 2.5.45,
+more extensive disk statistics have been introduced to help measure disk
activity. Tools such as sar and iostat typically interpret these and do
the work for you, but in case you are interested in creating your own
tools, the fields are explained here.

-In most versions of the 2.4 patch, the information is found as additional
-fields in /proc/partitions. In 2.5, the same information is found in
-two places: one is in the file /proc/diskstats (appears in 2.5.69 and
-beyond), and the other is within the sysfs file system, which must be
-mounted in order to obtain the information. Throughout this document
-we'll assume that sysfs is mounted on /sys, although of course it may
-be mounted anywhere. In 2.5, both /proc/diskstats and sysfs use the
-same source for the information and so should not differ.
+In 2.4 now, the information is found as additional fields in
+/proc/partitions. In 2.6, the same information is found in two
+places: one is in the file /proc/diskstats, and the other is within
+the sysfs file system, which must be mounted in order to obtain
+the information. Throughout this document we'll assume that sysfs
+is mounted on /sys, although of course it may be mounted anywhere.
+Both /proc/diskstats and sysfs use the same source for the information
+and so should not differ.

Linux–Kernel: [PATCH] update Documentation/iostats.txt for 2.6

Here are examples of these different formats:

@@ –25,15 +25,15 @@ Here are examples of these different for
3 1 9221278 hda1 35486 0 35496 38030 0 0 0 0 38030 38030

–2.5 sysfs:

+2.6 sysfs:

446216 784926 9550688 4382310 424847 312726 5922052 19310380 0 3376340 23705160
35486 38030 38030 38030

–2.5 diskstats:

+2.6 diskstats:

3 0 hda 446216 784926 9550688 4382310 424847 312726 5922052 19310380 0 3376340 23705160
3 1 hda1 35486 38030 38030 38030

–On 2.4 you might execute "grep 'hda ' /proc/partitions". On 2.5, you have
+On 2.4 you might execute "grep 'hda ' /proc/partitions". On 2.6, you have
a choice of "cat /sys/block/hda/stat" or "grep 'hda ' /proc/diskstats".
The advantage of one over the other is that the sysfs choice works well
if you are watching a known, small set of disks. /proc/diskstats may
@@ –43,7 +43,7 @@ each snapshot of your disk statistics.

In 2.4, the statistics fields are those after the device name. In
the above example, the first field of statistics would be 446216.

–By contrast, in 2.5 if you look at /sys/block/hda/stat, you'll

+By contrast, in 2.6 if you look at /sys/block/hda/stat, you'll

find just the eleven fields, beginning with 446216. If you look at
/proc/diskstats, the eleven fields will be preceded by the major and
minor device numbers, and device name. Each of these formats provide
@@ –93,35 +93,35 @@ Field 11 — weighted # of milliseconds s

To avoid introducing performance bottlenecks, no locks are held while
modifying these counters. This implies that minor inaccuracies may be
introduced when changes collide, so (for instance) adding up all the
–read I/Os issued per partition should equal those made to the disks
–... but due to the lack of locking it may only be very close.

+read I/Os issued per partition should equal those made to the disks ...

+but due to the lack of locking it may only be very close.

–In release 2.5.65 the 2.5 counters were made per–cpu, which made the lack
–of locking almost a non–issue. When the statistics are read, the per–cpu
–counters are summed (possibly overflowing the unsigned 32–bit variable
–they are summed to) and the result given to the user. There is no
–convenient user interface for accessing the per–cpu counters themselves.

+In 2.6, there are counters for each cpu, which made the lack of locking
+almost a non–issue. When the statistics are read, the per–cpu counters
+are summed (possibly overflowing the unsigned 32–bit variable they are
+summed to) and the result given to the user. There is no convenient
+user interface for accessing the per–cpu counters themselves.

Disks vs Partitions

- There were significant changes between 2.4 and 2.5 in the I/O subsystem.
- +There were significant changes between 2.4 and 2.6 in the I/O subsystem. As a result, some statistic information disappeared. The translation from a disk address relative to a partition to the disk address relative to the host disk happens much earlier. All merges and timings now happen at the disk level rather than at both the disk and partition level as
- in 2.4. Consequently, you'll see a different statistics output on 2.5 for
- +in 2.4. Consequently, you'll see a different statistics output on 2.6 for partitions from that for disks. There are only *four* fields available
- for partitions on 2.5 machines. This is reflected in the examples above.
- +for partitions on 2.6 machines. This is reflected in the examples above.

Field 1 -- # of reads issued

This is the total number of reads issued to this partition.

Field 2 -- # of sectors read

This is the total number of sectors requested to be read from this partition.

-Field 3 -- # of reads issued

+Field 3 -- # of writes issued

This is the total number of writes issued to this partition.

-Field 4 -- # of sectors read

+Field 4 -- # of sectors written

This is the total number of sectors requested to be written to this partition.

@@ -135,14 +135,16 @@ a subtle distinction that is probably un
Additional notes

- In 2.5, sysfs is not mounted by default. Here's the line you'll want
- to add to your /etc/fstab:
- +In 2.6, sysfs is not mounted by default. If your distribution of
- +Linux hasn't added it already, here's the line you'll want to add to
- +your /etc/fstab:

```
none /sys sysfs defaults 0 0
```

- In 2.5, at the same time that disk statistics appeared in sysfs, they were
- removed from /proc/stat. In 2.4, they appear in both /proc/partitions
- and /proc/stat.
- +In 2.6, all disk statistics were removed from /proc/stat. In 2.4, they
- +appear in both /proc/partitions and /proc/stat, although the ones in
- +/proc/stat take a very different format from those in /proc/partitions
- +(see proc(5), if your system has it.)

-- ricklind@us.ibm.com

-

Linux-Kernel: [PATCH] update Documentation/iostats.txt for 2.6

To unsubscribe from this list: send the line "unsubscribe linux-kernel" in the body of a message to majordomo@vger.kernel.org
More majordomo info at <http://vger.kernel.org/majordomo-info.html>
Please read the FAQ at <http://www.tux.org/lkml/>