

Re: sort + head = weirdness?

Source: <http://unix.derkeiler.com/Newsgroups/comp.unix.shell/2003-10/0094.html>

From: Barry Margolin (barry.margolin_at_level3.com)

Date: 10/02/03

Date: Wed, 01 Oct 2003 23:03:14 GMT

In article <pan.2003.10.01.22.14.55.821407@livebridge.com>, Mark <unixadmin@livebridge.com> wrote:

- >This is really weird. I'm catting a text file, piping it through
- >'sort' and then through 'head' and I'm getting really strange results.
- >
- >This only affects files > 4096 bytes.
- >It's only a problem when piping through 'head'; 'tail' works fine.
- >It's only a problem when piping through a file containing 'sort'; it
- >works fine with 'sort' itself.
- >
- >(The test file is just a text file containing the numbers 1-9 repeated.)
- >('format' is a file containing the single word 'sort'. Eventually this
- >will contain other formatting commands.)
- >
- >So, here's what happens:
- >
- >With a file 4096 bytes or less it works fine.
- ># ls -al testfile -rw-rw-r--
- > 1 root root 4096 Oct 1 14:46 testfile
- >
- ># cat testfile | format | head -1
- >12345
- >
- >If I add one more character to the file, it fails.
- ># ls -al testfile
- >-rw-rw-r-- 1 root root 4097 Oct 1 14:38 testfile
- >
- ># cat testfile | format | head -1
- >123456
- >/usr/local/bin/drs/format: line 1: 8688 Broken pipe sort

The "head" command exits as soon as it has printed the specified number of lines, so the reading end of the pipe is closed. When sort tries to write some more, it gets this error.

The reason it starts at 4096 bytes is because this is stdio's default buffer size. For a small file the entire thing is written in a single buffer, so there's no attempt to write anything after "head" has exited.

comp.unix.shell: Re: sort + head = weirdness?

But for a larger file it gets this error when it tries to flush the second buffer.

```
>If, with the same 4097 byte file I replace 'head' with 'tail', it works  
>fine.  
># cat testfile | format | tail -1  
>123456789
```

Tail has to read all of its input before it can print anything, because that's the only way to know when it has gotten to the last line. So it can't exit until after sort has finished all its writing.

```
>If I replace 'format' with 'sort' it works fine.  
># cat testfile | sort | head -1  
>123456  
>  
>Any idea what's going on here? Is there some buffer somewhere that can  
>only take 4k of data?
```

Interactive shells typically suppress the error message for "broken pipe" because this is such a common situation. But in your case the error is happening in the shell being used to run the "format" script, which is non-interactive.

Check the documentation of the shell (you didn't say what shell you're using for the script) to see if there's any way to suppress this.

```
--  
Barry Margolin, barry.margolin@level3.com  
Level(3), Woburn, MA  
*** DON'T SEND TECHNICAL QUESTIONS DIRECTLY TO ME, post them to newsgroups.  
Please DON'T copy followups to me -- I'll assume it wasn't posted to the group.
```