
Subject: Re: the last line of a large file

Posted by [Carsten Lechte](#) on Fri, 10 Aug 2007 15:00:33 GMT

[View Forum Message](#) <> [Reply to Message](#)

Conor wrote:

> lol! Really! What in the world is the point of putting the number of
> lines at the end of the file?

One legitimate reason would be that sometimes you only know how much data you have until after you have processed it all, especially if the data sets are so large that you only ever have a small subset in RAM.

A legitimate example are zip archives, where the table of contents is written to the end of the file, because the the compressed sizes of the archive members cannot be known in advance, and it would double the running time to determine the compressed size beforehand, it would furthermore use twice the disk space to re-write the file with the contents in front, it would be impossible to keep the whole archive in RAM before writing it, and finally, one may not be able leave space for the contents table at the beginning of the file, to be filled in later, because one would have to know how long the table will be beforehand...

Of course, this does not mean that the original poster's data has a legitimate reason for being organised like this.

For the original poster's problem, one idea is to get the file size in bytes, skip to position `file_size-1000`, read that small chunk and parse it for the desired metadata. This might even be faster than actually counting the lines with `FILE_LINES`, but it is probably only worth it if the metadata contains more useful information than just the number of lines in the file.

chl
