

comic courtesy of [xkcd.com](http://xkcd.com/)

Create a file and populate it with content:

|  |
| --- |
| **grep “lin” /usr/share/dict/words > lin\_words.txt** |

Create a tarball that contains lin\_words.txt:

|  |
| --- |
| **tar -cvf lin\_words.tar lin\_words.txt** |

And a tarball that is gzipped:

|  |
| --- |
| **tar -czvf lin\_words.tar.gz lin\_words.txt** |

And one that is bzipped:

|  |
| --- |
| **tar -cjvf lin\_words.tar.bz2 lin\_words.txt** |

Check the file types:

|  |
| --- |
| **file lin\*** |

lin\_words.tar: POSIX tar archive (GNU)  
lin\_words.tar.bz2: bzip2 compressed data, block size = 900k  
lin\_words.tar.gz: gzip compressed data, last modified: Sun Feb 11 18:00:12 2018, from Unix  
lin\_words.txt: UTF-8 Unicode text

Let’s take a look at our files:

|  |
| --- |
| **ls -l lin\*** |

-rw-rw-r-- 1 dmc1208 dmc1208 122880 Feb 11 17:59 lin\_words.tar  
-rw-rw-r-- 1 dmc1208 dmc1208 34450 Feb 11 18:00 lin\_words.tar.bz2  
-rw-rw-r-- 1 dmc1208 dmc1208 31521 Feb 11 18:00 lin\_words.tar.gz  
-rw-rw-r-- 1 dmc1208 dmc1208 120953 Feb 11 17:59 lin\_words.txt

Notes:

1. The .tar file is larger than the original because of overhead. Tar does no compression to the files.
2. Gzip and bzip are close to the same size, it is personal preference as to which one to use (gzip is more common).

Let’s create some more files in a directory:

|  |
| --- |
| **mkdir ~/in\_words**  **tar xzvf ~/lin\_words.tar.gz -C ~/in\_words/**  **grep “bin” /usr/share/dict/words > ~/in\_words/bin\_words.txt**  **grep “din” /usr/share/dict/words > ~/in\_words/din\_words.txt**  **grep “ein” /usr/share/dict/words > ~/in\_words/ein\_words.txt**  **grep “sin” /usr/share/dict/words > ~/in\_words/sin\_words.txt**  **grep “tin” /usr/share/dict/words > ~/in\_words/tin\_words.txt**  **du -h ~/in\_words**  **ls -l ~/in\_words** |

384K in\_words/

-rw-rw-r-- 1 dmc1208 dmc1208 17436 Feb 11 18:09 bin\_words.txt  
-rw-rw-r-- 1 dmc1208 dmc1208 48681 Feb 11 18:09 din\_words.txt  
-rw-rw-r-- 1 dmc1208 dmc1208 24168 Feb 11 18:09 ein\_words.txt  
-rw-rw-r-- 1 dmc1208 dmc1208 120953 Feb 11 17:59 lin\_words.txt  
-rw-rw-r-- 1 dmc1208 dmc1208 40933 Feb 11 18:10 sin\_words.txt  
-rw-rw-r-- 1 dmc1208 dmc1208 129359 Feb 11 18:10 tin\_words.txt

Note:

1. The -C option for tar is used to specify a target location other than your present working directory.
2. The -h option for du is for ‘human-readable’ output (bytes instead of blocks)

Let’s gzip the files:

|  |
| --- |
| **gzip -r in\_words/**  **du -h in\_words/**  **ls -l in\_words/** |

116K in\_words/

-rw-rw-r-- 1 dmc1208 dmc1208 4744 Feb 11 18:09 bin\_words.txt.gz  
-rw-rw-r-- 1 dmc1208 dmc1208 12955 Febg 11 18:09 din\_words.txt.gz  
-rw-rw-r-- 1 dmc1208 dmc1208 6429 Feb 11 18:09 ein\_words.txt.gz  
-rw-rw-r-- 1 dmc1208 dmc1208 31377 Feb 11 17:59 lin\_words.txt.gz  
-rw-rw-r-- 1 dmc1208 dmc1208 11119 Feb 11 18:10 sin\_words.txt.gz  
-rw-rw-r-- 1 dmc1208 dmc1208 33501 Feb 11 18:10 tin\_words.txt.gz

Notes:

1. Compare the file sizes of the gzip’d files to the standard text files.
2. Notice that it gzips each file but doesn’t put them into a single file. That is what tar does.
3. Each file has been replaced by its gzipped content. The original doesn’t exist.
4. The -r option for gzip tells it to recurse through all files and directories.

Decompress the gzip’d files and bzip them:

|  |
| --- |
| **gzip -dr in\_words/**  **du -h in\_words/**  **ls -l in\_words/**  **bzip2 in\_words/\***  **du -h in\_words/**  **ls -l in\_words/** |

84K in\_words/

-rw-rw-r-- 1 dmc1208 dmc1208 4744 Feb 11 18:09 bin\_words.txt.bz2  
-rw-rw-r-- 1 dmc1208 dmc1208 12955 Feb 11 18:09 din\_words.txt.bz2  
-rw-rw-r-- 1 dmc1208 dmc1208 6429 Feb 11 18:09 ein\_words.txt.bz2  
-rw-rw-r-- 1 dmc1208 dmc1208 31377 Feb 11 17:59 lin\_words.txt.bz2  
-rw-rw-r-- 1 dmc1208 dmc1208 11119 Feb 11 18:10 sin\_words.txt.bz2  
-rw-rw-r-- 1 dmc1208 dmc1208 33501 Feb 11 18:10 tin\_words.txt.bz2

Notes:

1. The -d option in gzip decompresses the files and renames them back to original.
2. Compare the bzip’d file size to the gzip’d. Which seems more efficient with text files?
3. Each file has been replaced, just like gzip does.

Decompress the bzip’d files and create tarballs that are gzip’d and bzip’d:

|  |
| --- |
| **bzip2 -d in\_words/\***  **du -h in\_words/**  **tar czvf inwords.tar.gz in\_words/**  **tar cjvf inwords.tar.bz2 in\_words/**  **ls -l inw\*** |

-rw-rw-r--. 1 linuxuser linuxuser 75117 Feb 8 14:40 inwords.tar.bz2  
-rw-rw-r--. 1 linuxuser linuxuser 70853 Feb 8 14:39 inwords.tar.gz

List the contents of each tarball:

|  |
| --- |
| **tar tzvf inwords.tar.gz**  **tar tjvf inwords.tar.bz2** |

drwxrwxr-x linuxuser/linuxuser 0 2014-02-08 14:38 in\_words/  
-rw-rw-r-- linuxuser/linuxuser 39873 2014-02-08 14:00 in\_words/sin\_words.txt  
-rw-rw-r-- linuxuser/linuxuser 12551 2014-02-08 13:52 in\_words/bin\_words.txt  
-rw-rw-r-- linuxuser/linuxuser 83249 2014-02-08 13:37 in\_words/din\_words.txt  
-rw-rw-r-- linuxuser/linuxuser 83249 2014-02-08 13:37 in\_words/ein\_words.txt  
-rw-rw-r-- linuxuser/linuxuser 83249 2014-02-08 13:37 in\_words/lin\_words.txt  
-rw-rw-r-- linuxuser/linuxuser 105097 2014-02-08 13:59 in\_words/tin\_words.txt

Notes:

1. The -t option displays the contents of the tar file.
2. The -v option means verbosely (shows all file information like ls -l)..

Let’s create a new location and un-tar the files:

|  |
| --- |
| **mkdir new\_words**  **tar xzvf inwords.tar.gz -C new\_words/**  **du -sh new\_words/**  **ls -l new\_words** |