DAT470/DIT065 Assignment 1

Mirco Ghadri
mircog@chalmers.se

Ali Alkhaled
alialk@student.chalmers.se

April 15, 2024

Problem 1

(a) Log in to bayes.

Command: ssh mircog@bayes.ita.chalmers.se

(b) Create a directory under your home directory called assignment1 problem1. Make that directory your working directory (change into that directory).

Command:

1. mkdir assignment1_problem1 (create the directory)
2. cd assignment1_problem1 (change into that directory)

(c) Download the Linux kernel version 6.6.17 sources from https://cdn.kernel.org/pub/linux/kernel/v6.x/linux-6.7.5.tar.xz.

Command: wget https://cdn.kernel.org/pub/linux/kernel/v6.x/linux-6.7.5.tar.xz

(d) Extract the contents of the tarball in that directory.

Command: tar -xf linux-6.7.5.tar.xz

(e) Determine the size of the tarball before extraction and the size of the extracted contents in human-readable units.

Command: ls -lh

Output:
drwxrwxr-x 26 mircog mircog 4.0K Feb 16 19:14 linux-6.7.5
-rw-rw-r– 1 mircog mircog 135M Feb 16 19:24 linux-6.7.5.tar.xz

From the output we see that the size of the extracted tarball is 4 KiloBytes, while the size of the tarball is 135 MegaBytes.

(f) The Linux Kernel consists of C code. C code is organized into header files

(file extension .h) and source files (file extension .c). Determine the total number of header and source files in the sources, respectively.

	Source files	Header files
Command	`find linux-6.7.5 -name "*.c" | wc -l`	`find linux-6.7.5 -name "*.h" | wc -l`
Output	33238	24174

Table 1: Number of .c(source) and .h(header) files in linux

(g) Write a one-liner that determines the 10 longest C code files (either a header or a source file) by the number of lines in the file, and stores their names in a file called longest.txt. List the names of the files in your answer.

Command: find linux-6.7.5 -name "*.c" -o -name "*.h" | xargs wc -l | sort -n | grep -v "total" | tail -n 10 | tee longest.txt

Output:

Number of Lines	File Name
82050	linux-6.7.5/drivers/gpu/drm/amd/include/asic_reg/nbio/nbio_4_3_0_sh_mask.h
103385	linux-6.7.5/drivers/gpu/drm/amd/include/asic_reg/dpcs/dpcs_4_2_0_sh_mask.h
103633	linux-6.7.5/drivers/gpu/drm/amd/include/asic_reg/dpcs/dpcs_4_2_2_sh_mask.h
118975	linux-6.7.5/drivers/gpu/drm/amd/include/asic_reg/nbio/nbio_7_0_sh_mask.h
120339	linux-6.7.5/drivers/gpu/drm/amd/include/asic_reg/nbio/nbio_2_3_sh_mask.h
133888	linux-6.7.5/drivers/gpu/drm/amd/include/asic_reg/nbio/nbio_6_1_sh_mask.h
136141	linux-6.7.5/drivers/gpu/drm/amd/include/asic_reg/dpcs/dpcs_4_2_3_sh_mask.h
152483	linux-6.7.5/drivers/gpu/drm/amd/include/asic_reg/nbio/nbio_7_2_0_sh_mask.h
154426	linux-6.7.5/drivers/gpu/drm/amd/include/asic_reg/nbio/nbio_7_7_0_sh_mask.h
222893	linux-6.7.5/drivers/gpu/drm/amd/include/asic_reg/dcn/dcn_3_2_0_sh_mask.h

Table 2: Top 10 .c and .h Files with the Most Lines

(h) Write a one-liner that reads the names of the files from longest.txt, computes the SHA-1 checksum of each file in the list, and outputs the checksums into a file called longest.sha1sum, such that it can be used to verify the correctness of these files. Include the content of longest.sha1sum in your answer.

Command: awk 'print $2' longest.txt | xargs sha1sum | tee longest.sha1sum

Output:
7f9cd0a3b14145afc2a0e8d5cb3ed8404f25f528
8310d3cba875211546ef3f90f128a5e398a977d2
7f08cb654d993778dd3fb8b4aff1818803c21560
d23b6b96a9fb8b633600463d7fb8d7fc29d0fbf7
8b420d3843a8ae686f395a997b214e87a59678e6
10c99459d8a44a2fed00cd1e34c282645762a996
f044c4e1461eb291bffb67298f2eedc37aa97cf1

51aff58e2fd9918b6db9c7ed8040131ec3b2f80f
b42c1c7f15c9616c67c556445bc19e2e2c1d270e
a44869d03eb1a2a631ed51ed11d2ca1205d4f09d

(i) Write a one-liner that determines how many unique files there are that contain the string Linus Torvalds. Note that the string can occur multiple times in a file; only count each file once. Include the number of such files in your answer.

Command: grep -rl "Linus Torvalds" linux-6.7.5 | wc -l **Output:** 571

(j) Write a one-liner that selects all .c files that contain the SPDX-License-Identifier and outputs, for each such file, the string filename license, and output the content to a file called licenses.txt.

Command: find linux-6.7.5 -name *.c | xargs grep "SPDX-License-Identifier" | cut -d ":" -f1,3 | tr -d ":" | tee licenses.txt

(k) Write a one-liner that determines the most common license in licenses.txt. Include the license and the count of its occurrences in your answer.

Command: cut -d " " licenses.txt -f2 — sort — uniq -c — sort -nr — head -n 1

Output: 10794 GPL-2.0

(l) Download the files longest.txt, longest.sha1sum, and licenses.txt from bayes to your own computer.

Command:
scp mircog@bayes.ita.chalmers.se: /assignment1_problem1/longest.txt Desktop
scp mircog@bayes.ita.chalmers.se: /assignment1_problem1/longest.sha1sum Desktop
scp mircog@bayes.ita.chalmers.se: /assignment1_problem1/licenses.txt Desktop

Problem 2

For the full script, see *assignment1_problem2.sh*

To run the script, we wrote `bash assignment1_problem2.sh`

Output:

```
This script is being run by mircog
No LSB modules are available.
The model and clock frequency of the CPU is: Intel(R)
    Xeon(R) Gold 6126 CPU @ 2.60GHz.
The number of physical CPUs, i.e. sockets in use is 2.
Each physical CPU has 12 number of cores.
The number of threads on each core is 2.
```

```
The instruction set architecture of the cpu is x86_64.
The cache line length is 64.
The l1,l2 and l3 cache are divided as such:
L1d cache: 768 KiB (24 instances) L1i cache: 768 KiB (24
    instances) L2 cache: 24 MiB (24 instances) L3 cache:
    38.5 MiB (2 instances)
The total system ram is 791181348 kB.
The total number of gpus 1.
The gpu model is Tesla V100-PCIE-16GB.
The total amount of ram on the GPUs is 16384 MiB.
The filesystem of /data is /dev/mapper/vg_data-lv_data
    and the filesystem of /datainbackup is /dev/mapper/
    vg_data-lv_datainbackup.
The total amount of diskspace in /data is 73T and the
    amount of free space is 67T.
The total amount of diskspace in /datainbackup is 24T and
     the amount of free space is 12T.
The linux kernel version is Linux version 5.15.0-91-
    generic (buildd@lcy02-amd64-045) (gcc (Ubuntu 11.4.0-1
    ubuntu1~22.04) 11.4.0, GNU ld (GNU Binutils for Ubuntu
    ) 2.38) #101-Ubuntu SMP Tue Nov 14 13:30:08 UTC 2023.
The OS distro name and version is:
Distributor ID: Ubuntu Description: Ubuntu 22.04.3 LTS
    Release: 22.04 Codename: jammy
The path to the python executable is /opt/local/anaconda3
    /bin/python3 and the python version is Python 3.9.16.
```