Bash, a command line interface for interacting with the operating system, was created in the 1980s. Other popular shells are *zsh* and *fish*.

1 Programming in Bash

1.1 Shebang

The shebang (#!) at the head of a script indicates an interpreter for execution, as in #!/bin/bash. Lines starting with a # (with the exception of shebang) are comments and thus won't be executed.

1.2 Quoting and literals

Single quotes "preserve the literal value of characters enclosed within them. A single quote may not appear between single quotes, even when escaped, but may appear between **double quotes**".

They work similarly, with an exception that the shell expands any variables that appear within them.

1.3 Variables

Variable names are case sensitive. They can contain digits and underscores as well, but a name starting with a digit is not allowed. Example:

```
var="kind"
echo ${var}ness # kindness
```

Special variables:

- 1. \$0: name of the script itself.
- \$1, \$2, \$3, ...: the first, second, etc. argument. shift removes first argument and advances rest of them forward.
- 3. \$* and \$@ denote all the positional parameters.
- 4. \$#: the number of positional parameters
- 5. \$?: exit status of last executed command.
- 6. \$\$: the process ID of the shell.
- 7. \$!: the process ID of last executed command.

To read a line of input, use read shell built-in.

- read reads a line from the stdin:
- n returns after reading n characters,
- n displays a prompt.

1.4 Expansions

"After the command has been split into tokens, these tokens or words are expanded or resolved. There are eight kinds of expansion performed, which we will discuss in the next sections, in the order that they are expanded."

1.4.1 Brace expansion

Brace expansion is used when we need to generate all possible string combinations. Both of the commands produce the same output:

```
echo {I,really,love,dots}.
echo I. really. love. dots.
```

Warning: it does not expand the variables (\$var), which is done later, but supports ranges (sequences) of characters:

```
echo {a..t}
abcdefghijklmnopqrst
```

and (maybe zero paded or with an increment rate) integers, assuming the Bash version is 4 or newer:

```
echo {01..10..1}.~
01. 02. 03. 04. 05. 06. 07. 08. 09. 10.
```

There is a tilde expansion as well. The expressions ~ and ~<user> expand to the home directory of the current (or given) user.

1.4.2 Parameter expansion

 \${var^}, \${var,} convert first character to upper and lowercase. \${var^^}, \${var,,} do the same to all characters.

\${var^}, \${var^^} are undocumented now: they reverse the case.

In case of the array expansion, every expanded element changes case, no matter what.

\${var#pattern} removes the pattern from the beginning of the string, if possible.

It's greedy variant is \${var##pattern}.

\${var%pattern} and **\${var%%pattern}** do the same, but from the end of the string.

Application: extracting parts of a filename.

\${var/pattern/string} performs a single search and replace operation.

\${var//pattern/string} searches for all occurrences of the pattern and replaces them.

- 4. \${#var} returns length of the string.
- 5. \${var:offset:length} skips first offset characters from var and truncates the output to given length. :length may be skipped.

Negative values separated with extra space are accepted.

\${var:-value} uses a default value, if var is empty or unset.

\${var:=value} does the same, but performs an assignment as well.

\${var:+value} uses an alternative value if var isn't empty or unset!

1.4.3 Command substitution

To execute commands in a subshell and then pass their standard output, use \$(commands).

1.4.4 Arithmetic expansion

The arithmetic expression \$((...)) is evaluated and expands to the result. Bash guarantees that the output will be a one-word integer.

1.4.5 Process substitution

This kind of substitution (where input or output of a command appears as a temporary file) is performed simultaneously with the following: arithmetic and parameter expansions, command substitution.

```
<( ... ) # not specified by POSIX! >( ... )
```

1.5 Streams

There are always three default files open:

- 1. stdin (the keyboard, file descriptor 0),
- 2. stdout (the screen, file descriptor 1) and
- stderr (error messages output, file descriptor
 2).

These **streams** can be **redirected**:

- 1. cmd > file redirects to a file (overwrites),
- 2. cmd >> file appends instead,
- m>n (or m>&n) redirects a file descriptor to a file (or another file descriptor),
- 4. &>file redirects stdout and stderr to a file,
- 5. :> file truncates file to zero length,
- 6. | (pipe) serves as a command chaining tool.

1.6 Control flow statements

The one-line constructs && and | | work not like and, or (\land, \lor) , but the *if* – *then* – *else* statement.

1.6.1 Conditionals

Here at least one statement must be specified inside every block, but one can use a single colon (:) as a null statement to avoid rewriting the code.

```
if condition; then
commands
elif second_condition; then
some_commands
else
other_commands
fi

select word in "Bash" "Haskell" "Python"
do
echo "Your language is $word".
done
There is also a case instruction:
```

```
case $language in
  bash)
    echo "Bourne Again Shell!"
;;
  python|haskell)
    echo "Python or Haskell!"
    exit 1
;;
  *)
    echo "Unknown language!"
;; # optional
esac
```

1.6.2 Testing conditions

Remember that test command follows symbolic links (except for the -h test).

1. File tests:

- (a) -e file exists, -s file is nonempty,
- (b) -d directory, -f regular file, -h symlink,
- (c) -b block device, -c character device,
- (d) -p named pipe, -S socket.

2. File permissions:

- (a) -r readable, -w writable, -x executable,
- (b) -u setuid, -g setgid, -k sticky bit.
- 3. String tests: -z empty, -n nonempty.
- 4. Arithmetic tests:
 - (a) -eq =, $-ne \neq$,
 - (b) -lt <, -gt >,
 - (c) $-le \le$, $-ge \ge$.

1.6.3 Loops

```
for var in "the first" "the second"; do
  echo "${var}"
done

for (( i = 1; i <= 10; i++ )); do
  echo "i = ${i}."
done # C-style

while read myline; do
  echo "It says ${myline}"
done < some_file</pre>
```

As Bash Guide for Beginners by M. Garrels says:

- 1. the break statement is used to exit the current loop before its normal ending.
- 2. the continue statement resumes iteration of an enclosing while, until, select or for loop.

2 Emacs shortcuts in Bash

(See http://readline.kablamo.org/emacs.html)

- 1. Ctrl A moves to the start of the line,
- 2. Ctrl E moves to the end of the line,
- 3. Ctrl U deletes to the beginning of the line.
- 4. Ctrl K deletes to the end of the line.
- 5. Ctrl W deletes to the start of the word.
- 6. Ctrl Y pastes text from the clipboard.
- 7. Ctrl L clears the screen.
- 8. Alt Rundoes all changes to the line.
- 9. Ctrl R searches incrementally up the history.
- 10. Ctrl XE invokes an editor to write complex command.

3 Shell style guide

The following notes are meant to be summary of a style guide written by Paul Armstrong and too many more to mention (revision 1.26).

Bash is the only shell scripting language permitted for executables. Bash should only be used for simple wrapper scripts or small utilities.

Executables should have no extension, libraries must have a . sh extension and should not be executable. SUID and SGID are *forbidden* on shell scripts.

All error messages should go to STDERR, a function to print out error messages along with other status information is recommended:

```
err () {
    echo "[$(date +%Y-%m-%d\ %T)]: $@" >&2
}
```

Comments. Start each file with a description of its contents. Any function that is in a library or not both obvious and short, must be commented. Comment tricky, interesting or important parts of code. Use TODO comments for temporary or good enough but not perfect code and short-term solutions.

The following are required for any new code:

- 1. Indent 2 spaces, no tabs.
- 2. Maximum line length is 80 characters.
- 3. Long pipelines should be split one per line.
- 4. Indent case alternatives by 2 spaces.
- 5. Always quote strings containing variables, command substitutions or spaces.

Use quotes rather than filler characters if possible. Use an explicit path when doing wildcard expansion of filenames. Avoid eval. Use process substitution or for loops in preference to piping to while. Finally, [[...]] is preferred over [and test.

Naming conventions. Function and variable names should be lower case, with underscore to separate words. Constants and environment variable names should be all caps, declared at the top of the file. Use readonly or declare -r to ensure they're read only. Declare function specific variables with local. A function called main is required for scripts long enough to contain at least one other function.

4 Text processing: grep, sed, awk

This is an expanded description of three powerful text processing tools: grep, sed and awk.

See also: https://beyondgrep.com/, ack.

4.1 grep – pattern search enginge

The ed command g/re/p was used to globally search a regular expression and print.

- grep prints lines matching a pattern:
- c prints a count of matching lines instead,
- e uses a "regexp" pattern,
- f obtains patterns from a file,
- ignores case disctinctions,
- v inverts the sense of matching,
- w selects only lines with whole words matches,
- n prints line numbers as well,
- A prints "num" lines of trailing content,
- B prints "num" lines of leading content,
- C prints "num" lines of both contents,
- E interprets pattern as an extended regexp,
- P interprets pattern as a Perl regexp,
- R reads all files under each directory.

4.2 sed – stream editor

- **sed** filters and transforms text:
- adds a script to the commands to be executed,
- -i edits files in place,
- -n suppresses auto-printing of pattern space,
- -r accepts extended regular expressions.

The simplest usage is sed 's/foo/bar/g' which substitutes (s) strings globally (g). There are other options, including:

- a appends line before,
- d deletes line,
- i inserts line before,
- p prints line,
- w writes pattern space to a file.

Default delimiter / can be replaced by any other. This is useful when regular expression already contains /. Addresses allow limiting to given line numbers:

- 1. 1-10 first ten lines
- 2. \$ the last line
- 3. 10^2 even lines starting from the 10th.

One can also use regular expressions:

replaces spaces with underscores in lines containing a colon. Negation may be obtained with !s.

4.3 awk - Aho, Weinberger, Kernighan

awk is a language used as a data extraction and reporting tool.

General form of its code:

```
#! /bin/awk
```

```
BEGIN {initialization}
search pattern {actions} # for example:
/word[0-9]/ {gold += $2} # regex
!/word[0-9]/ {counter++} # negation
END {final actions}
```

Awk is weakly typed: variables can be treated either as numeric values or strings, which are not represented as one-dimensional arrays of characters! Important variables include:

- 1. FS: field separator (tab and space by default),
- 2. **OFS**: output field separator,
- 3. **RS**: record separator (new line),
- 4. NR: number of the current record,
- 5. **NF**: number of fields in the current record.

Numerical functions: int, sqrt, exp, log, sin, cos, atan2, rand (pseudo random from [0, 1)), srand (without parameters, uses time of day as a seed).

String/text functions: length, split, sprintf, gsub, sub, index, match, tolower, toupper.

4.4 Regular expressions

- 1. POSIX character classes:
 - (a) [:alnum:] = [a-zA-Z0-9]
 - (b) [:alpha:] = [a-zA-Z]
 - (c) [:ascii:] = $[\x00-\x7F]$
 - (d) $[:blank:] = [\t]$
 - (e) [:cntrl:] = [$\x00-\x1F\x7F$]
 - (e) [:cntrl:] = [\x00-(f) [:digit:] = [0-9]
 - (g) [:graph:] = $[\x21-\x7E]$
 - (h) [:lower:] = [a-z]
 - (i) [:print:] = $[\x20-\x7E]$
 - $(j) [:space:] = [\t\r\n\v\f]$
 - (k) [:word:] = $[A-Za-z0-9_]$
 - (k) [.word.] = [k-2d-20-9_]
- (l) [:xdigit:] = [A-Fa-f0-9]
- 2. Repetitions:
 - (a) *: 0 or more, +: 1 or more, ?: 0 or 1,
 - (b) $\{a, b\}$: at least a, at most b.
- 3. Anchors:
 - (a) ^: start of line,
 - (b) \$: end of line,
 - (c) \<: start of word,
 - (d) \>: end of word.
- 4. Other:
 - (a) one | two: one or two,
 - (b) (one): a group,
 - (c) \$n: *n*th group,
 - (d) [abcd], [a-d]: ranges,
 - (e) [^abcd]: negation (not [abcd]).

Awk regex:

 $/^(ab|c)*[0-9]+/$

Grep regex (with -E flag):

^(ab|c)*[[:digit:]]+

Sed regex:

sed $"s/^{(ab|c)*[[:digit:]]+/1/"}$

5 Unix utilities and shell builtins

File system

- **cat** concatenates and prints files:
- shows all nonprinting characters,
- numbers nonempty output lines,
- s suppresses repeated empty output lines.
- **tac** does the same in reverse.
- rev reverses lines characterwise.
- **nl** numbers lines of files:
- s adds "string" after line number,
- uses "number" columns for line numbers.
- **chgrp** changes group ownership.
- **chmod** changes permissions of a file:

ugoa of the owner, group, other or all users,

- +-= adds, removes or sets selected file mode bits,
- rwx selects file mode bits: read 4/write 2/execute 1.
 - **chown** changes owner of a file.
- umask sets file mode creation mask.
- **touch** changes file timestamps:
- a only the access time,
- m only the modification time,
- uses custom stamp instead of current time,
- does not create files.
- See also: **cksum** (CRC checksums), **md5sum**.
- shasum prints or checks SHA message digests:
- algorithm: 1, 224, 256, 384, 512, 512224 or 512256,
- b reads in binary mode,
- checks SHA sums read from the "files".
- wc prints newline, word and byte counts (lwc):
- m prints the character counts,
- L prints the maximum display width.
- **dd** converts and copies a file:
- 1. if = reads from a file,
- 2. of = writes to a file,
- 3. bs= up to "bytes" bytes at a time,
- 4. count = copies only "n" input blocks.
- **cp** copies files and directories:
- b makes a backup of existing destination files,
- f removes an existing destination file if needed,
- prompts before overwrite,
- n does not overwrite existing files,
- L always follows symlinks in "source",
- P never follows symlinks in "source",
- preserves timestamps, mode, ownership,
- r copies directories recursively,
- makes symbolic links instead,
- hard links files instead,
- t copies all "source" arguments into "directory",
- T treats "destination" as a normal file,
- u copies only newer source files,
- explains what is being done.
- **mv** moves (renames) files:
- b makes a backup of existing destination files,
- prompts before overwriting,
- f does not prompt before overwriting,
- n does not overwrite existing destination files.
- moves all "source" arguments into "directory",
- T treats "destination" as a normal file,
- u moves only newer source files,
- explains what is being done.
- **rm** removes files or directories:
- f never prompts,
- always prompts,
- removes directories and their contents.
- See also: rmdir (directories removal), shred.
- mkdir makes directories (-p: with parents as needed, no error if existing).
- **df** reports file system disk space usage:
- prints size in powers of 1024,
- list inode information instead of block usage,
- limits listing to file systems of given type,
- x limits listing to file systems not of given type,
- T prints file systems types.

- **du** estimates file space usage:
- a writes counts for all files, not just directories,
- c produces a grand total,
- d the depth at which summing should occur,
- h prints sizes in human readable format,
- s diplays only a total,
- X excludes files that match pattern.
- ☐ **file** determines file type.
- in a directory hierarchy.
- 1. Tests:
 - -name base of file name,
 - case insensitive name, -iname
 - -group, -user ownership -perm 755, -perm /u=x permissions
 - -size +5M -1G size between 5MB and 1GB
 - -amin -60 accessed in last hour -cmin, -mmin: created, modified,
 - -mtime +7 modified over a week ago
 - -type d directories only, -type f files only,
- empty files or directories only, -empty 2. Example (deletes files larger than 5 megabytes):
- find / -size +5M -exec rm -f $\{\}\$ \; database can be updated with updatedb.
- **I fsck** checks and repairs a Linux filesystem:
- a automatically repairs (without any question!),
- t specifies the type(s) of filesystem to be checked,
- A tries to check all filesystems in one run, M skips mounted filesystems,
- R skips the root filesystem.
- In makes hard links between files (only in the same file system, not between directories):
- s makes symbolic links instead.
- **Is** lists directory contents:
- a does not ignore entries starting with dot,
- F appends indicator to entries,
- h prints human readable sizes,
- prints the index number of each file, 1 prints permissions, number of hard links,
- owner, group, size, last-modified date as well, r reverses order while sorting,
- R lists subdirectories recursively,
- S sorts by file size (largest first),
- t sorts by modification time (newest first),
- ☐ **tree** lists tree-like contents of directories.
- mount mounts a filesystem.
- **pwd** prints name of current directory.
- **pv** monitors progress of data through a pipe.
- ☐ tar stores and extracts files from a disk archive: c creates a new archive,
- x extracts files.
- t lists the contents of an archive,
- verbosely lists files processed,
- j bzip2 compression,
- z uses zip/gzip (gz compression),
- f uses archive file or device (???),
- k does not replace existing files when extracting.
- **tee** duplicates pipe content:
- a appends to the given files, does not overwrite,
- i ignores interrupts.

5.2 Processes

- **chroot** changes the root directory of the calling process and their children.
- ☐ at schedules commands to be executed once, at a particular time in the future: it accepts times of the form HH:MM, midnight, noon or teatime; MMDD[CC]YY, MM/DD/[CC]YY, DD.MM.[CC]YY or [CC]YY-MM-DD (the specification of a date must follow the specification

- of the time of day). You can also give times like now + 3 hours.
- ☐ **bg** resumes suspended jobs in the background. fg resumes suspended jobs in the foreground.
- ☐ **jobs** lists the active jobs. ☐ **cmd &** runs command in the background.
- crontab maintain individual users' crontab files. See also cron: a daemon that executes scheduled commands.
- **kill** sends a TERM signal to a process.
- ☐ **killall** kills processes by name.
- **ps** reports a snapshot of the current processes:
- e selects all processes,
- f does full-format listing,
- m shows threads (see also: H, -L, -T)
- C selects processes by command name,
- p selects processes by PID,
- u selects processes by EUID or name.
- pstree displays a tree of processes.
- ☐ **nice** changes process priority. pgrep, pkill looks up or signals processes based
- on name and other attributes. ☐ **time** runs programs and summarizes system
- resource usage.
- ☐ **iostat** monitors CPU and disk usage. vmstat monitors memory usage.
- ☐ **top** displays Linux processes.
- ☐ See also: **htop** (Hisham top based on ncurses).

5.3 User environment

- clear clears the terminal screen.
- env runs programs in modified environment.
- exit terminates the calling process.
- ☐ **finger** looks up user information.
- **history** displays the history list. mesg displays messages from other users.
- \square passwd changes user password:
- d deletes (empties) an account's password,
- e expires an account's password, n minimum days to change password,
- warning days before password expire,
- x maximum days a password remains valid.
- **pwgen** generate pronounceable passwords:
- generates hard to memorize passwords,
- includes special characters, n includes numbers,
- **su** changes user ID or becomes superuser.
- **sudo** executes a command as superuser:
- u as a different user.
- ☐ **hostname** shows/sets the host name: i displays the network address.
- uname prints system information:
- a all information, in the following order: s the kernel name
- n the network node hostname,
- r the kernel release,
- v the kernel version.
- m the machine hardware name, p the processor type,
- i the hardware platform, o the operating system.
- **uptime**: how long has system been running? wall writes a message to all users,
- write sends a message to another user. pinky is a lightweight version of finger.
- ☐ **who** shows who is logged on, ☐ w does the same, shows what they are doing,
- ☐ whoami prints effective userid.

5.4	Text processing		yes outputs a string repeatedly until killed.	5.7	Hardware
	$\mathbf{awk},\mathbf{grep}$ and \mathbf{sed} have been described earlier.	5.5	Shell builtins		dmesg prints/controls the kernel ring buffer.
	comm compares two sorted files line by line. shuf generates random permutations: treats each "arg" as an input line,		alias allows a string to be substituted for a word.	_ _ ?	Isblk lists block devices. Isof lists info about files opened by processes. fuser identifies processes using files/sockets.
	treats each number through as an input		cd changes the shell working directory:		lsusb listsq USB devices.
n	line, outputs at most "count" lines,	- □	to the previous directory. echo* displays a line of text:	5.8	For programmers
r	output lines can be repeated (with -n).	e	enables interpretation of backslash escapes,		g++ compiles, assembles and links C++ files:
	sort sorts lines of text files: checks for sorted input,	n	does not output the trailing newline. test checks file types and compares values.	0	writes the build output to a file named
f	folds lower case to upper case characters,		unset unsets a shell variable, removing it from	5.9	Miscellaneous
h	compares general numerical values, compares human readable numbers, sorts via a key,		memory and the shell's exported environment. wait waits for process to change state.		bc is an arbitrary precision calculator language.
n	compares string numerical values,	5.6	Networking		echo 'obase=16;255' bc prints FF,
	reverses the results, stabilizes the sort.		1 6	2.	echo 'ibase=2;obase=A;10' bc prints 2,
	tsort performs topological sort.		ftp is a File Transfer Protocol client.	3.	<pre>scale=10 (after bc -1) sets working preci-</pre>
С	uniq omits repeated lines: prefixes lines by the number of occurences,		wget is a non-interactive network downloader. specifies lists of file suffixes or patterns (when		sion. dc is a reverse-polish desk calculator. One of
	only prints duplicate lines, one for each group,		wildcard characters appear) to accept or reject,		the oldest Unix utilities, predating even the in-
	avoids comparing first fields,		goes to background immediately after startup, continues getting a partially-downloaded file,		vention of the C programming language. cal, ncal displays a calendar.
	ignores differences in case, avoids comparing first characters,		turns on options suitable for mirroring: infi-	е	displays date of Easter,
	compares no more than n characters.		nite recursion and time-stamping,		displays Julian days,
	cut prints selected parts of lines:	np	does not ever ascend to the parent directory when retrieving recursively,		displays the specified month, prints the numbers of the weeks,
c	complement complements the selection, selects only these characters,		identifies as "agent-string" to the HTTP server.		displays a calendar for the specified year,
	uses "delim" instead of Tab for field delimeter,	W	waits the specified number of seconds between the retrievals (see also -random-wait).	3 	displays the previous, current and next month. date prints or set the system date and time.
	selects only these fields, does not print lines not containing delimeters.				seq prints a sequence of numbers:
s	join joins lines of two files on a common field.	Ц	rlogin starts a terminal session (which is not encrypted!) on a remote host.	W	equalizes width by padding with leading zeroes.
	paste merges lines of files.		ssh connects & logs into a (remote) hostname:		sleep delays for a specified amount of time.
d s	reuses characters from "list" instead of tabs, pastes one file at a time, not in parallel.		requests compression of all data, uses private key for authentication,		true , false does nothing, (un)successfully.
	tr translates or deletes characters:	p	selects a port to connect to on the remote host,		
	uses the complement of "set1", deletes characters, does not translate,	Х	enables X11 forwarding.		
	replaces each sequence of a repeated character		dig interrogates DNS name servers.		
	that is listed in the last specified "set" with a		performs a simplified reverse lookup. host is a DNS lookup utility.		
	single occurrence of that character.		nslookup is (probably) deprecated! Use dig		
	diff compares files line by line: outputs in two columns,	_	and host.		
	ignores case differences,		arp manipulates the system ARP cache. ifconfig configures a network interface.		
	ignores all white space.		netstat prints networking subsystem related		
	fmt is a simple optimal text formatter, fold wraps each line to fit in specified width.		info: network connections, interface statistics, routing tables, and so on.		
	head outputs the first (last) part of files:	a	whether sockets is listening or not,		
С	the first "num" bytes,		only TCP connections.		
	the first "num" lines, tail the last "num" bytes:	Ц	ping tests the reachability of a host on an IP network by sending ICMP ECHO_REQUEST:		
	the last "num" bytes,		stops after sending "count" packets,		
	the last "num" lines,	n	numeric output only, avoids symbolic names for host addresses lookup.		
	outputs appended data as the file grows, sleeps for "n" seconds between iterations.		route shows and manipulates the IP routing		
	split splits a file into pieces:		table. traceroute prints a trace of the route that IP		
	generates suffixes of length "n" (default 2), puts "size" bytes per output file,	Ш	packets are travelling to a remote host:		
d	uses numeric (not alphabetic) suffixes,	Ι	uses ICMP ECHO for probes.		
	puts "number" lines/records per output file, generates "chunks" output files.		rsync copies files fast (remote or local):		
	See also: csplit .		in archive mode, equivalent to: preserves group,		
	more pages text too large to fit on one screen,	0	preserves owner (super-user only)		
	allows scrolling down, not up (deprecated!).	_	preserves permissions, preserves modification times,		
	less is an enhanced version of more: monitors the tail of a file which is growing.		copies symlinks as symlinks,		
	vim is an advanced text editor, too complex to		make backups,		
	be explained here. See also emacs or nano .		skip based on checksum, performs a dry run without changes made,		
	xargs builds and executes command lines:	r	resursively,		
0	takes care of filenames with spaces, back-		skip newer files on the receiver, increases verbosity,		
I	slashes. replaces occurrences of "string" with names	z	compresses file data during the transfer,		
	read from standard input.		scp copies files securely.		