NAME

parallel_alternatives - Alternatives to GNU parallel

DIFFERENCES BETWEEN GNU Parallel AND ALTERNATIVES

There are a lot programs with some of the functionality of GNU parallel. GNU parallel strives to include the best of the functionality without sacrificing ease of use.

parallel has existed since 2002 and as GNU parallel since 2010. A lot of the alternatives have not had the vitality to survive that long, but have come and gone during that time.

GNU parallel is actively maintained with a new release every month since 2010. Most other alternatives are fleeting interests of the developers with irregular releases and only maintained for a few years.

SUMMARY TABLE

The following features are in some of the comparable tools:

Inputs
I1. Arguments can be read from stdin I2. Arguments can be read from a file I3. Arguments can be read from multiple files I4. Arguments can be read from command line I5. Arguments can be read from a table I6. Arguments can be read from the same file using #! (shebang) I7. Line oriented input as default (Quoting of special chars not needed)

Manipulation of input
M1. Composed command M2. Multiple arguments can fill up an execution line M3. Arguments can be put anywhere in the execution line M4. Multiple arguments can be put anywhere in the execution line M5. Arguments can be replaced with context M6. Input can be treated as the complete command line

Outputs
O1. Grouping output so output from different jobs do not mix O2. Send stderr (standard error) to stderr (standard error) O3. Send stdout (standard output) to stdout (standard output) O4. Order of output can be same as order of input O5. Stdout only contains stdout (standard output) from the command O6. Stderr only contains stderr (standard error) from the command O7. Buffering on disk O8. Cleanup of file if killed O9. Test if disk runs full during run O10. Output of a line bigger than 4 GB

Execution
E1. Running jobs in parallel E2. List running jobs E3. Finish running jobs, but do not start new jobs E4. Number of running jobs can depend on number of cpus E5. Finish running jobs, but do not start new jobs after first failure E6. Number of running jobs can be adjusted while running E7. Only spawn new jobs if load is less than a limit

Remote execution
R1. Jobs can be run on remote computers R2. Basefiles can be transferred R3. Argument files can be transferred R4. Result files can be transferred R5. Cleanup of transferred files R6. No config files needed R7. Do not run more than SSHD's MaxStartups can handle R8. Configurable SSH command R9. Retry if connection breaks occasionally

Semaphore
S1. Possibility to work as a mutex S2. Possibility to work as a counting semaphore

Legend - = no x = not applicable ID = yes

As every new version of the programs are not tested the table may be outdated. Please file a bug-report if you find errors (See REPORTING BUGS).

parallel: I1 I2 I3 I4 I5 I6 I7 M1 M2 M3 M4 M5 M6 O1 O2 O3 O4 O5 O6 O7 O8 O9 O10 E1 E2 E3 E4 E5 E6 E7 R1 R2 R3 R4 R5 R6 R7 R8 R9 S1 S2
find -exec: - - - x - - M2 M3 - - - O2 O3 O4 O5 O6 - - - - - - - - - - - - - x x
make -j: - - - - - - O1 O2 O3 - x O6 E1 - - E5 - - - - - - - - -
xjobs, prl, dxargs, mdm/middelman, xapply, paexec, ladon, jobflow, ClusterSSH: TODO - Please file a bug-report if you know what features they support (See REPORTING BUGS).
DIFFERENCES BETWEEN xargs AND GNU Parallel

Summary table (see legend above): I1 I2 - - - - - - M2 M3 - - - - O2 O3 - O5 O6 E1 - - - - - - - - x - - - - - -

xargs offers some of the same possibilities as GNU parallel.

xargs deals badly with special characters (such as space, ",', and "). To see the problem try this:

```
touch important_file
touch 'not important_file'
l s not* | xargs rm
mkdir -p "My brother's 12\" records"
l s | xargs rmdir
touch 'c:\windows\system32\clfs.sys'
echo 'c:\windows\system32\clfs.sys' | xargs ls -l
```

You can specify -0, but many input generators are not optimized for using NUL as separator but are optimized for newline as separator. E.g. awk, ls, echo, tar -v, head (requires using -z), tail (requires using -z), sed (requires using -z), perl (0 and |0 instead of |n), locate (requires using -0), find (requires using -print0), grep (requires using -z or -Z), sort (requires using -z).

GNU parallel's newline separation can be emulated with:

```
cat | xargs -d "\n" -n1 command
```

xargs can run a given number of jobs in parallel, but has no support for running number-of-cpu-cores jobs in parallel.

xargs has no support for grouping the output, therefore output may run together, e.g. the first half of a line is from one process and the last half of the line is from another process. The example Parallel grep cannot be done reliably with xargs because of this. To see this in action try:

```
parallel perl -e '$a="1\"."x10000000\"; print \"a,\"\n"' 
  '->' () { a b c d e f g h
# Serial = no mixing = the wanted result
# 'tr -s a-z' squeezes repeating letters into a single letter
echo a b c d e f g h | xargs -P1 -n1 grep | tr -s a-z
# Compare to 8 jobs in parallel
parallel -kP8 -n1 grep : : a b c d e f g h | tr -s a-z
echo a b c d e f g h | xargs -P8 -n1 grep | tr -s a-z
```

Or try this:

```
slow_seq() {
  echo Count to "$@"
  seq "$@" |
  perl -ne '$|=1; for(split//){ print; select($a,$a,$a,0.100);}'
}
export -f slow_seq
# Serial = no mixing = the wanted result
seq 8 | xargs -n1 -P1 -I {} bash -c 'slow_seq {}'
# Compare to 8 jobs in parallel
seq 8 | parallel -P8 slow_seq {}
seq 8 | xargs -n1 -P8 -I {} bash -c 'slow_seq {}'
```

xargs has no support for keeping the order of the output, therefore if running jobs in parallel using xargs the output of the second job cannot be postponed till the first job is done.
**xargs** has no support for running jobs on remote computers.

**xargs** has no support for context replace, so you will have to create the arguments.

If you use a replace string in **xargs** (-I) you can not force **xargs** to use more than one argument.

Quoting in **xargs** works like -q in GNU parallel. This means composed commands and redirection require using bash -c.

```
ls | parallel "wc {} >\{}.wc"
ls | parallel "echo {}; ls {}|wc"
```

becomes (assuming you have 8 cores and that none of the filenames contain space, " or ").

```
ls | xargs -d "\n" -P8 -I () bash -c "wc {} >\{}.wc"
ls | xargs -d "\n" -P8 -I () bash -c "echo {}; ls {}|wc"
```

https://www.gnu.org/software/findutils/

**DIFFERENCES BETWEEN find -exec AND GNU Parallel**

**find -exec** offers some of the same possibilities as GNU parallel.

**find -exec** only works on files. Processing other input (such as hosts or URLs) will require creating these inputs as files. **find -exec** has no support for running commands in parallel.

https://www.gnu.org/software/findutils/ (Last checked: 2019-01)

**DIFFERENCES BETWEEN make -j AND GNU Parallel**

**make** -j can run jobs in parallel, but requires a crafted Makefile to do this. That results in extra quoting to get filenames containing newlines to work correctly.

**make** -j computes a dependency graph before running jobs. Jobs run by GNU parallel does not depend on each other.

(Very early versions of GNU parallel were coincidentally implemented using make -j).

https://www.gnu.org/software/make/ (Last checked: 2019-01)

**DIFFERENCES BETWEEN ppss AND GNU Parallel**

Summary table (see legend above): I1 I2 - - - - I7 M1 - M3 - - M6 O1 - - x - - E1 E2 ?E3 E4 - - - R1 R2 R3 R4 - - ?R7 ? ? - -

**ppss** is also a tool for running jobs in parallel.

The output of **ppss** is status information and thus not useful for using as input for another command. The output from the jobs are put into files.

The argument replace string ($ITEM) cannot be changed. Arguments must be quoted - thus arguments containing special characters (space "&!") may cause problems. More than one argument is not supported. Filenames containing newlines are not processed correctly. When reading input from a file null cannot be used as a terminator. **ppss** needs to read the whole input file before starting any jobs.

Output and status information is stored in ppss_dir and thus requires cleanup when completed. If the dir is not removed before running **ppss** again it may cause nothing to happen as **ppss** thinks the task is already done. GNU parallel will normally not need cleaning up if running locally and will only need cleaning up if stopped abnormally and running remote (–cleanup may not complete if stopped abnormally). The example **Parallel grep** would require extra postprocessing if written using **ppss**.

For remote systems PPSS requires 3 steps: config, deploy, and start. GNU parallel only requires one step.
EXAMPLES FROM ppss MANUAL

Here are the examples from ppss's manual page with the equivalent using GNU parallel:

1. /ppss.sh standalone -d /path/to/files -c 'gzip'
2. /ppss.sh standalone -d /path/to/files -c 'cp "$ITEM" /destination/dir'
3. /ppss.sh standalone -f list-of-urls.txt -c 'wget -q'
4. /ppss.sh standalone -f list-of-urls.txt -c 'cp "$ITEM" /destination-dir'
5. /ppss.sh standalone -f list-of-urls.txt -c 'wget -q'
6. /ppss.sh standalone -f list-of-urls.txt -c 'wget -q "$ITEM"'

DIFFERENCES BETWEEN pexec AND GNU Parallel

Summary table (see legend above): I1 I2 - I4 I5 - - M1 - M3 - - M6 O1 O2 O3 - O5 O6 E1 - - E4 - E6 - R1 - - - R6 - - - S1 -

pexec is also a tool for running jobs in parallel.

EXAMPLES FROM pexec MANUAL

Here are the examples from pexec's info page with the equivalent using GNU parallel:

1. pexec -o sqrt-%s.dat -p "$\{seq 10\}" -e NUM -n 4 -c -- \"echo "scale=10000;sqrt($NUM)\"\" | bc'
2. seq 10 | parallel -j4 'echo "scale=10000;sqrt\(\{\}\)" | bc > sqrt-\{\}.dat'
3. pexec -p "$\{ls myfiles*.ext\}" -i %s -o %s.sort -- sort
4. ls myfiles*.ext | parallel sort \{ \} "->\{\}.sort"
5. pexec -f image.list -n auto -e B -u star.log -c -- \fistar $B.fits -f100 -F id,x,y,flux -o $B.star'
6. parallel -a image.list \fistar \{\},fits -f100 -F id,x,y,flux -o \{\}.star 2>star.log

https://github.com/louwrentius/PPSS
GNU Parallel alternatives

4 pexec -r *.png -e IMG -c -o - -- `\convert $IMG $IMG\%.png.jpeg ; "echo $IMG: done"`  
4 ls *.png | parallel 'convert {} (.).jpeg; echo {}: done'  
5 pexec -r *.png -i %s -o %s.jpeg -c 'pnmtpng | pnmtjpegl'  
5 ls *.png | parallel 'pnmtpng < {} | pnmtjpegl > {}.jpg'  
6 for p in *.png ; do echo $(p%.png) ; done \ pexec -f -i %s.png -o %s.png -c 'pnmtpng | pnmtjpegl'  
6 ls *.png | parallel 'pnmtpng < {} | pnmtjpegl > {}.jpg'  
7 LIST=$(for p in *.png ; do echo $(p%.png) ; done) pexec -r LIST -i %s.png -o %s.png -c 'pnmtpng | pnmtjpegl'  
7 Is *.png | parallel 'pnmtpng < {} | pnmtjpegl > {}.jpg'  
8 pexec -n 8 -r *.jpg -y unix -e IMG -c \ pexec -j -m blockread -d $IMG | \ jpegtopng | pnmscale 0.5 | pnmtjpegl | \ jpegtopng | \ pexec -j -m blockwrite -s th_\$IMG  
8 Combining GNU parallel and GNU sem.  
8 ls *.jpg | parallel -j8 'sem --id blockread cat {} | jpegtopng | `pnmscale 0.5 | pnmtjpegl | sem --id blockwrite cat > th_{\{}'  
8 If reading and writing is done to the same disk, this may be faster as only one process will be either reading or writing:  
8 ls *.jpg | parallel -j8 'sem --id diskio cat {} | jpegtopng | `pnmscale 0.5 | pnmtjpegl | sem --id diskio cat > th_{\{}'  

https://www.gnu.org/software/pexec/

DIFFERENCES BETWEEN xjobs AND GNU Parallel

xjobs is also a tool for running jobs in parallel. It only supports running jobs on your local computer.

xjobs deals badly with special characters just like xargs. See the section DIFFERENCES BETWEEN xargs AND GNU Parallel.

Here are the examples from xjobs's man page with the equivalent using GNU parallel:

1 ls -1 *.zip | xjobs unzip  
1 ls *.zip | parallel unzip  
2 ls -1 *.zip | xjobs -n unzip  
2 ls *.zip | parallel unzip >/dev/null  
3 find . -name "*.bak" | xjobs gzip  
3 find . -name "*.bak" | parallel gzip  
4 ls -1 *.jar | sed 's\\((.*)\\1 > \1.idx" | xjobs jar tf  
4 ls *.jar | parallel jar tf {} > {} .idx  
5 xjobs -s script  
5 cat script | parallel  
6 mkfifo /var/run/my_named_pipe; xjobs -s /var/run/my_named_pipe & echo unzip 1.zip >> /var/run/my_named_pipe; echo tar cf /backup/myhome.tar /home/me >> /var/run/my_named_pipe  
6 mkfifo /var/run/my_named_pipe; cat /var/run/my_named_pipe | parallel & echo unzip 1.zip >> /var/run/my_named_pipe; echo tar cf /backup/myhome.tar /home/me >> /var/run/my_named_pipe
GNU Parallel alternatives

Differences between prll and GNU Parallel

`prll` is also a tool for running jobs in parallel. It does not support running jobs on remote computers.

`prll` encourages using BASH aliases and BASH functions instead of scripts. GNU `parallel` supports scripts directly, functions if they are exported using `export -f`, and aliases if using `env_parallel`.

`prll` generates a lot of status information on stderr (standard error) which makes it harder to use the stderr (standard error) output of the job directly as input for another program.

Here is the example from `prll`'s man page with the equivalent using GNU `parallel`:

```
prll -s 'mogrify -flip $1' *.jpg
```

```
parallel mogrify -flip ::: *.jpg
```

https://github.com/exzombie/prll (Last checked: 2019-01)

Differences between dxargs and GNU Parallel

`dxargs` is also a tool for running jobs in parallel.

`dxargs` does not deal well with more simultaneous jobs than SSHD's MaxStartups. `dxargs` is only built for remote run jobs, but does not support transferring of files.


Differences between mdm/middleman and GNU Parallel

middleman(mdm) is also a tool for running jobs in parallel.

Here are the shellscripts of https://web.archive.org/web/20110728064735/http://mdm.berlios.de/usage.html ported to GNU `parallel`:

```
1 xapply 'cd %1 && make all' /*
1 parallel 'cd {} && make all' ::: */
2 xapply -f 'diff %1 ../../../version5/%1' manifest | more
2 parallel diff {} ../../../version5/{} < manifest | more
3 xapply -p/dev/null -f 'diff %1 %2' manifest1 checklist1
3 parallel --link diff {1} {2} ::: manifest1 checklist1
4 xapply 'indent' *.c
4 parallel indent ::: *.c
5 find ~ksb/bin -type f ! -perm -111 -print | xapply -f -v 'chmod a+x' -
5 find ~ksb/bin -type f ! -perm -111 -print | parallel -v chmod a+x
```

https://github.com/cklin/mdm (Last checked: 2019-01)

Differences between xapply and GNU Parallel

`xapply` can run jobs in parallel on the local computer.

Here are the examples from `xapply`'s man page with the equivalent using GNU `parallel`:

```
1 xapply 'cd %1 && make all' */
1 parallel 'cd {} && make all' ::: */
2 xapply -f 'diff %1 ../../../version5/%1' manifest | more
2 parallel diff {} ../../../version5/{} < manifest | more
3 xapply -p/dev/null -f 'diff %1 %2' manifest1 checklist1
3 parallel --link diff {1} {2} ::: manifest1 checklist1
4 xapply 'indent' *.c
4 parallel indent ::: *.c
5 find ~ksb/bin -type f ! -perm -111 -print | xapply -f -v 'chmod a+x' -
5 find ~ksb/bin -type f ! -perm -111 -print | parallel -v chmod a+x
```
GNU Parallel alternatives

6 find */ -... | fmt 960 1024 | xapply -f -i /dev/tty 'vi'
6 sh <(find */ -... | parallel -s 1024 echo vi)
6 find */ -... | parallel -s 1024 -Xuj1 vi
7 find ... | xapply -f -5 -i /dev/tty 'vi' - - - - - -
7 sh <(find ... | parallel -n5 echo vi)
7 find ... | parallel -n5 -uj1 vi
8 xapply -fn "" /etc/passwd
8 parallel -k echo < /etc/passwd
9 tr ':' '012' < /etc/passwd | xapply -7 -nf 'chown %1 %6' - - - - - -
9 tr ':' '012' < /etc/passwd | parallel -N7 chown {1} {6}
10 xapply '[ -d %1/RCS ] || echo %1'*
10 parallel '[ -d {}/RCS ] || echo {} :::: *
11 xapply -f '[ -f %1 ] && echo %1' List | ...
11 parallel '[ -f {} ] && echo {}' < List | ...

DIFFERENCES BETWEEN AIX apply AND GNU Parallel

apply can build command lines based on a template and arguments - very much like GNU parallel. apply does not run jobs in parallel. apply does not use an argument separator (like ::::); instead the template must be the first argument.

Here are the examples from IBM's Knowledge Center and the corresponding command using GNU parallel:

1. To obtain results similar to those of the ls command, enter:

   apply echo *
   parallel echo :::: *

2. To compare the file named a1 to the file named b1, and the file named a2 to the file named b2, enter:

   apply -2 cmp a1 b1 a2 b2
   parallel -N2 cmp :::: a1 b1 a2 b2

3. To run the who command five times, enter:

   apply -0 who 1 2 3 4 5
   parallel -N0 who :::: 1 2 3 4 5

4. To link all files in the current directory to the directory /usr/joe, enter:

   apply 'ln %1 /usr/joe' *
   parallel ln {} /usr/joe :::: *

DIFFERENCES BETWEEN paexec AND GNU Parallel

**paexec** can run jobs in parallel on both the local and remote computers.

**paexec** requires commands to print a blank line as the last output. This means you will have to write a wrapper for most programs.

**paexec** has a job dependency facility so a job can depend on another job to be executed successfully. Sort of a poor-man's make.

Here are the examples from **paexec**'s example catalog with the equivalent using GNU parallel:

1. **1_div_X_run**:
   ```
   ../../paexec -s -l -c "`pwd`/1_div_X_cmd" -n +1 <<EOF 
   parallel echo {} | `pwd`/1_div_X_cmd <<EOF
   ```

2. **all_substr_run**:
   ```
   ../../paexec -lp -c "`pwd`/all_substr_cmd" -n +3 <<EOF 
   parallel echo {} | `pwd`/all_substr_cmd <<EOF
   ```

3. **cc_wrapper_run**:
   ```
   ../../paexec -c "env CC=gcc CFLAGS=-O2 `pwd`/cc_wrapper_cmd" \ 
   -n 'host1 host2' \ 
   -t '/usr/bin/ssh -x' <<EOF 
   parallel echo {} | "env CC=gcc CFLAGS=-O2 `pwd`/cc_wrapper_cmd" \ 
   -S host1,host2 <<EOF
   # This is not exactly the same, but avoids the wrapper
   parallel gcc -O2 -c -o {.}.o {} \ 
   -S host1,host2 <<EOF
   ```

4. **toupper_run**:
   ```
   ../../paexec -lp -c "`pwd`/toupper_cmd" -n +10 <<EOF 
   parallel echo {} | `pwd`/toupper_cmd <<EOF
   ```

   # Without the wrapper:
   ```
   parallel echo {} | awk \{print\ toupper\("\$0")\}' <<EOF
   ```

https://github.com/cheusov/paexec

DIFFERENCES BETWEEN map(sitaramc) AND GNU Parallel

**map** sees it as a feature to have less features and in doing so it also handles corner cases incorrectly. A lot of GNU parallel's code is to handle corner cases correctly on every platform, so you will not get a nasty surprise if a user, for example, saves a file called: *My brother's 12" records.txt*

**map**'s example showing how to deal with special characters fails on special characters:

```
echo "The Cure" > My\ brother\'s\ 12"\ records
```

```
ls | \ 
map 'echo \-n \`gzip < "%" | wc \-c`; echo \-n \"*100/\"; wc \-c < "%"\ | bc
```

It works with GNU parallel:

```
ls | \ 
parallel \ 
'echo \-n \`gzip < {} | wc \-c`; echo \-n \"*100/\"; wc \-c < {}\' | bc
```
And you can even get the file name prepended:

```bash
ls | \n   parallel --tag \n   '{echo -n `gzip < {} | wc -c`;*100/; wc -c < {}} | bc''
```

**map** has no support for grouping. So this gives the wrong results without any warnings:

```bash
parallel perl -e '
   $a="1"; 
   x100000000\:print\: $a,"\"\n\"';>}' () \n   ::= a b c d e f
ls -l a b c d e f
parallel -kP4 -nl grep 1 > out.par ::= a b c d e f
map -p 4 'grep 1' a b c d e f > out.map-unbuf
map -p 4 'grep --line-buffered 1' a b c d e f > out.map-linebuf
map -p 1 'grep --line-buffered 1' a b c d e f > out.map-serial
ls -l out*
md5sum out*
```

The documentation shows a workaround, but not only does that mix stdout (standard output) with stderr (standard error) it also fails completely for certain jobs (and may even be considered less readable):

```bash
parallel echo -n () ::= 1 2 3

map -p 4 'echo -n % 2>/devnull: sed -e "s/\^/\$/;;:" 1 2 3 | \n   sort | cut -f2- -d:
```

**maps** replacement strings (% %D %B %E) can be simulated in GNU **parallel** by putting this in `~/.parallel/config`:

```bash
--rpl '%'
--rpl '%D $=Q{::dirname($$_)}';
--rpl '%B s:.*/:::s:./[.\[^/\.\]]+$::;
--rpl '%E s:.*\..::*'
```

**map** does not have an argument separator on the command line, but uses the first argument as command. This makes quoting harder which again may affect readability. Compare:

```bash
map -p 2 'perl -ne " for i in /$S+/$S+/$S/ and print $ARGV,"\n"
```

```bash
parallel -q perl -ne " for i in /$S+/$S+/$S/ and print $ARGV,"\n"
```

**map** can do multiple arguments with context replace, but not without context replace:

```bash
parallel --xargs echo 'BEGIN{()}'END' ::= 1 2 3
```

```bash
map "echo 'BEGIN{"$"}END'' 1 2 3
```

**map** requires Perl v5.10.0 making it harder to use on old systems.

**map** has no way of using % in the command (GNU **parallel** has -I to specify another replacement string than {}). By design **map** is option incompatible with **xargs**, it does not have remote job execution, a structured way of saving results, multiple input sources, progress indicator, configurable record delimiter (only field delimiter), logging of jobs run with possibility to resume, keeping the output in the same order as input, --pipe processing, and dynamically timeouts.
GNU Parallel alternatives

DIFFERENCES BETWEEN ladon AND GNU Parallel

ladon can run multiple jobs on files in parallel.

ladon only works on files and the only way to specify files is using a quoted glob string (such as "./*.jpg"). It is not possible to list the files manually.

As replacement strings it uses FULLPATH DIRNAME BASENAME EXT RELDIR RELPATH

These can be simulated using GNU parallel by putting this in ~/.parallel/config:

```bash
--rpl 'FULLPATH $_=Q($_);chomp($_=qx\{readlink -f $_\});'
--rpl 'DIRNAME $_=Q(:.dirname($_));chomp($_=qx\{readlink -f $_\});'
--rpl 'BASENAME s::*::;s:.[^./]*::;
--rpl 'EXT s:*::'
--rpl 'RELDIR $_=Q($_);chomp(($_,$c)=qx\{readlink -f $_;pwd\});
  s:\Q$c/\E::;$_=::dirname($_);'
--rpl 'RELPATH $_=Q($_);chomp(($_,$c)=qx\{readlink -f $_;pwd\});
  s:\Q$c/\E::;'
```

ladon deals badly with filenames containing " and newline, and it fails for output larger than 200k:

```bash
ladon "*" -- seq 36000 | wc
```

EXAMPLES FROM ladon MANUAL

It is assumed that the '--rpl's above are put in ~/.parallel/config and that it is run under a shell that supports ""/"" globbing (such as zsh):

1 ladon ""*.txt"" -- echo RELPATH
2 parallel echo RELPATH ::: ""*.txt
3 ladon "~/Documents/""*.pdf"" -- shasum FULLPATH >hashes.txt
4 parallel shasum FULLPATH ::: ~/Documents/""*.pdf >hashes.txt
5 ladon -m thumbs/RELDIR ""*/.jpg"" -- convert FULLPATH -thumbnail 100x100^ -gravity center -extent 100x100 thumbs/RELPATH
6 parallel mkdir -p thumbs/RELDIR; convert FULLPATH -thumbnail 100x100^ -gravity center -extent 100x100 thumbs/RELPATH ::: ""*/.jpg
7 ladon "~/Music/""*.wav"" -- lame -V 2 FULLPATH DIRNAME/BASENAME.mp3
8 parallel lame -V 2 FULLPATH DIRNAME/BASENAME.mp3 ::: ~/Music/""*.wav

https://github.com/danielgtaylor/ladon (Last checked: 2019-01)

DIFFERENCES BETWEEN jobflow AND GNU Parallel

jobflow can run multiple jobs in parallel.

Just like xargs output from jobflow jobs running in parallel mix together by default. jobflow can buffer into files (placed in /run/shm), but these are not cleaned up if jobflow dies unexpectedly (e.g. by Ctrl-C). If the total output is big (in the order of RAM+swap) it can cause the system to slow to a crawl and eventually run out of memory.

jobflow gives no error if the command is unknown, and like xargs redirection and composed commands require wrapping with bash -c.

Input lines can at most be 4096 bytes. You can at most have 16 '{'}s in the command template. More than that either crashes the program or simple does not execute the command.
**GNU Parallel alternatives**

**jobflow** has no equivalent for `--pipe`, or `--sshlogin`.

**jobflow** makes it possible to set resource limits on the running jobs. This can be emulated by GNU parallel using `bash`'s `ulimit`:

```
jobflow -limits=mem=100M,cpu=3,fsize=20M,nofiles=300 myjob
```

```
parallel 'ulimit -v 102400 -t 3 -f 204800 -n 300 myjob'
```

**EXAMPLES FROM jobflow README**

1. `cat things.list | jobflow -threads=8 -exec ./mytask {}`
2. `cat things.list | parallel -j8 ./mytask {}`
3. `seq 100 | jobflow -threads=100 -exec echo {}`
4. `seq 100 | parallel -j100 echo {}`
5. `cat urls.txt | jobflow -threads=32 -exec wget {}`
6. `cat urls.txt | parallel -j32 wget {}`
7. `find . -name "*.bmp" | jobflow -threads=8 -exec bmp2jpeg {.}.bmp {.}.jpg`
8. `find . -name "*.bmp" | parallel -j8 bmp2jpeg {.}.bmp {.}.jpg`

https://github.com/rofl0r/jobflow

**DIFFERENCES BETWEEN gargs AND GNU Parallel**

**gargs** can run multiple jobs in parallel.

Older versions cache output in memory. This causes it to be extremely slow when the output is larger than the physical RAM, and can cause the system to run out of memory.

See more details on this in `man parallel_design`.

Newer versions cache output in files, but leave files in $TMPDIR if it is killed.

Output to stderr (standard error) is changed if the command fails.

Here are the two examples from gargs website.

1. `seq 12 -1 1 | gargs -p 4 -n 3 "sleep {0}; echo {1} {2}"`
2. `seq 12 -1 1 | parallel -P 4 -n 3 "sleep {0}; echo {1} {2}"`

```
cat t.txt | gargs --sep "\s+" -p 2 "echo \{0\}:-\{1\}" full-line:\\n```

```
cat t.txt | parallel --colsep "\s+" -P 2 "echo \{1\}:-\{2\}" full-line:\\n```

https://github.com/brentp/gargs

**DIFFERENCES BETWEEN orgalorg AND GNU Parallel**

**orgalorg** can run the same job on multiple machines. This is related to `--onall` and `--nonall`.

**orgalorg** supports entering the SSH password - provided it is the same for all servers. GNU parallel advocates using `ssh-agent` instead, but it is possible to emulate `orgalorg`'s behavior by setting SSHPASS and by using `--ssh "sshpass ssh"`.

To make the emulation easier, make a simple alias:

```
alias par_emul="parallel -j0 --ssh 'sshpass ssh' --nonall --tag --lb"
```
If you want to supply a password run:

```
SSHPASS=`ssh-askpass`
```

or set the password directly:

```
SSHPASS=P4$$w0rd!
```

If the above is set up you can then do:

```
orgalorg -o frontend1 -o frontend2 -p -C uptime
par_emul -S frontend1 -S frontend2 uptime

orgalorg -o frontend1 -o frontend2 -p -C top -bid 1
par_emul -S frontend1 -S frontend2 top -bid 1

orgalorg -o frontend1 -o frontend2 -p -er /tmp -n \n'md5sum /tmp/bigfile' -S bigfile
par_emul -S frontend1 -S frontend2 --basefile bigfile \n--workdir /tmp md5sum /tmp/bigfile
```

**orgalorg** has a progress indicator for the transferring of a file. **GNU parallel** does not.

https://github.com/reconquest/orgalorg

**DIFFERENCES BETWEEN Rust parallel AND GNU Parallel**

Rust parallel focuses on speed. It is almost as fast as **xargs**. It implements a few features from **GNU parallel**, but lacks many functions. All these fail:

- # Read arguments from file
  parallel -a file echo

- # Changing the delimiter
  parallel -d _ echo ::: a_b_c_

These do something different from GNU parallel

- # -q to protect quoted $ and space
  parallel -q perl -e '$a=shift; print "$a"x10000000' ::: a b c

- # Generation of combination of inputs
  parallel echo {1} {2} ::: red green blue ::: S M L XL XXL

- # {= perl expression =} replacement string
  parallel echo '({= s/new/old/ =}') ::: my.new your.new

- # --pipe
  seq 100000 | parallel --pipe wc

- # linked arguments
  parallel echo ::: S M L :::: sml med lrg :::: R G B ::::: red grn blu

- # Run different shell dialects
  zsh -c 'parallel echo \=() ::: zsh && true'
  csh -c 'parallel echo \$\{\} ::: shell && true'
  bash -c 'parallel echo \$\((\()) ::: pwd && true'

- # Rust parallel does not start before the last argument is read
  (seq 10; sleep 5; echo 2) | time parallel -j2 'sleep 2; echo'
  tail -f /var/log/syslog | parallel echo

Most of the examples from the book GNU Parallel 2018 do not work, thus Rust parallel is not close to being a compatible replacement.
Rust parallel has no remote facilities.

It uses /tmp/parallel for tmp files and does not clean up if terminated abruptly. If another user on the system uses Rust parallel, then /tmp/parallel will have the wrong permissions and Rust parallel will fail. A malicious user can setup the right permissions and symlink the output file to one of the user's files and next time the user uses Rust parallel it will overwrite this file.

```
attacker$ mkdir /tmp/parallel
attacker$ chmod a+rwX /tmp/parallel
# Symlink to the file the attacker wants to zero out
attacker$ ln -s ~victim/.important-file /tmp/parallel/stderr_1
victim$ seq 1000 | parallel echo
# This file is now overwritten with stderr from 'echo'
victim$ cat ~victim/.important-file
```

If /tmp/parallel runs full during the run, Rust parallel does not report this, but finishes with success - thereby risking data loss.

https://github.com/mmstick/parallel

DIFFERENCES BETWEEN Rush AND GNU Parallel

**rush** (https://github.com/shenwei356/rush) is written in Go and based on gargs.

Just like GNU parallel rush buffers in temporary files. But opposite GNU parallel rush does not clean up, if the process dies abnormally.

rush has some string manipulations that can be emulated by putting this into ~/.parallel/config (/ is used instead of %, and % is used instead of ^ as that is closer to bash's ${var%postfix}):

```
--rpl '\{:\} s:(.[^/]+)*$$:\n--rpl '\{:%(^[^\{]+)?\} s:$$1(.[^/]+)*$$\n--rpl '\{/%(^[^\{]*?)\} s:.$$1(.[^/]+)*$$1:\n--rpl '\{/\} s:.*$$1(.[^/]+)*$$2:\n--rpl '\{@(.*?)\} /$$1/ and \_\_=$1;'
```

Here are the examples from rush's website with the equivalent command in GNU parallel.

EXAMPLES

1. Simple run, quoting is not necessary

```
$ seq 1 3 | rush echo {}

$ seq 1 3 | parallel echo {}
```

2. Read data from file (`-i')

```
$ rush echo {} -i data1.txt -i data2.txt

$ cat data1.txt data2.txt | parallel echo {}
```

3. Keep output order (`-k')

```
$ seq 1 3 | rush 'echo {}' -k

$ seq 1 3 | parallel -k echo {}
```

4. Timeout (`-t')
$ time seq 1 | rush 'sleep 2; echo {}' -t 1

$ time seq 1 | parallel --timeout 1 'sleep 2; echo {}'

5. Retry (`-r')

$ seq 1 | rush 'python unexisted_script.py' -r 1

$ seq 1 | parallel --retries 2 'python unexisted_script.py'

Use -u to see it is really run twice:

$ seq 1 | parallel -u --retries 2 'python unexisted_script.py'

6. Dirname (`/`) and basename (`%{}`) and remove custom suffix (`^suffix`)

$ echo dir/file_1.txt.gz | rush 'echo / %{} ^_1.txt.gz'

$ echo dir/file_1.txt.gz | parallel --plus echo // (/) (%^_1.txt.gz)

7. Get basename, and remove last (`.`) or any (`:`) extension

$ echo dir.d/file.txt.gz | rush 'echo . {}: %{.} %{:.} '

$ echo dir.d/file.txt.gz | parallel 'echo . {}: (/) (/:.)'

8. Job ID, combine fields index and other replacement strings

$ echo 12 file.txt dir/s_1.fq.gz | rush 'echo job #{}: {2} {2.:} {3%:^_1} '

$ echo 12 file.txt dir/s_1.fq.gz | parallel --colsep ' " ' echo job #{}: {2} {2.:} {3:/^_1} '

9. Capture submatch using regular expression (`@regexp')

$ echo read_1.fq.gz | rush 'echo @(.+)_\d'

$ echo read_1.fq.gz | parallel 'echo @(.+)_\d'

10. Custom field delimiter (`-d')

$ echo a=b=c | rush 'echo (1) (2) (3)' -d =

$ echo a=b=c | parallel -d = echo (1) (2) (3)

11. Send multi-lines to every command (`-n')

$ seq 5 | rush -n 2 -k 'echo "\n\n"'; echo'

$ seq 5 | parallel -n 2 -k \'echo {=-1 $=_join"\n",@arg[1..$#arg]} =}; echo'

---

GNU Parallel alternatives
GNU Parallel alternatives

$ seq 5 | rush -n 2 -k 'echo "()"; echo' -J ''

$ seq 5 | parallel -n 2 -k 'echo {}; echo'

12. Custom record delimiter (`-D`), note that empty records are not used.

$ echo a b c d | rush -D " " -k 'echo {}'

$ echo a b c d | parallel -n " " -k 'echo {}

Cannot be done by GNU Parallel

$ cat fasta.fa
>seq1
tag
>seq2
cat
gat
>seq3
attac
a
cat

$ cat fasta.fa | rush -D ">
  'echo FASTA record (#): name: {1} sequence: {2}' -k -d "\n"
# rush fails to join the multiline sequences

$ cat fasta.fa | (read -n1 ignore_first_char; parallel -d '=>' --colsep '\n' echo FASTA record {#}: 
        name: {1} sequence: '({=2 $_=_join"",@arg[2..$#arg]=}')
)

13. Assign value to variable, like `awk -v` (`-v`)

$ seq 1 | rush 'echo Hello, {fname} {lname}!' -v fname=Wei -v lname=Shen

$ seq 1 | parallel --NO 
        'fname=Wei; lname=Shen; echo Hello, {fname} {lname}!' 

$ for var in a b; do 
  $ seq 1 3 | rush -k -v var=$var 'echo var: {var}, data: {}'; 
$ done

In GNU parallel you would typically do:

$ seq 1 3 | parallel -k echo var: {1}, data: {2} :::: a b :::: -

If you really want the var:

$ seq 1 3 |
GNU Parallel alternatives

parallel -k var={1} ';echo var: $var, data: {}' ::: a b :::: -

If you really want the for-loop:

$ for var in a b; do
>    export var;
>    seq 1 3 | parallel -k 'echo var: $var, data: {}';
> done

Contrary to rush this also works if the value is complex like:

My brother's 12" records

14. Preset variable (`-v`), avoid repeatedly writing verbose replacement strings

# naive way
$ echo read_1.fq.gz | rush 'echo {:_1} {:_1}_2.fq.gz'

$ echo read_1.fq.gz | parallel 'echo {:_1} {:_1}_2.fq.gz'

# macro + removing suffix
$ echo read_1.fq.gz |
    rush -v p='{:_1}' 'echo {p} {p}_2.fq.gz'

$ echo read_1.fq.gz |
    parallel 'p={:_1}; echo $p$_2.fq.gz'

# macro + regular expression
$ echo read_1.fq.gz | rush -v p='@(.+?)_\d)' 'echo {p} {p}_2.fq.gz'

$ echo read_1.fq.gz | parallel 'p=@(.+?)_\d); echo $p$_2.fq.gz'

Contrary to rush GNU parallel works with complex values:

echo "My brother's 12"read_1.fq.gz" |
    parallel 'p=@(.+?)_\d); echo $p$_2.fq.gz'

15. Interrupt jobs by `Ctrl-C`, rush will stop unfinished commands and exit.

$ seq 1 20 | rush 'sleep 1; echo {}'
^C

$ seq 1 20 | parallel 'sleep 1; echo {}'
^C

16. Continue/resume jobs (`-c`). When some jobs failed (by execution failure, timeout, or canceling by user with `Ctrl + C`), please switch flag `-c/--continue` on and run again, so that rush can save successful commands and ignore them in NEXT run.

$ seq 1 3 | rush 'sleep {}; echo {}' -t 3 -c
$ cat successful_cmds.rush
$ seq 1 3 | rush 'sleep {}; echo {}' -t 3 -c

$ seq 1 3 | parallel --joblog mylog --timeout 2 \ 'sleep {}; echo {}'
GNU Parallel alternatives

```
$ cat mylog
$ seq 1 3 | parallel --joblog mylog --retry-failed 'sleep {}; echo ()'

Multi-line jobs:

$ seq 1 3 | rush 'sleep {}; echo ();
  echo finish ()' -t 3 -c -C finished.rush
$ cat finished.rush
$ seq 1 3 | rush 'sleep {}; echo ();
  echo finish ()' -t 3 -c -C finished.rush

$ seq 1 3 |
  parallel --joblog mylog --timeout 2 'sleep {}; echo ();
  echo finish ()'
$ cat mylog
$ seq 1 3 |
  parallel --joblog mylog --retry-failed 'sleep {}; echo ();
  echo finish ()'

17. A comprehensive example: downloading 1K+ pages given by three URL list files using `phantomjs save_page.js` (some page contents are dynamically generated by Javascript, so `wget` does not work). Here I set max jobs number (’-j’) as 20, each job has a max running time (-t) of 60 seconds and 3 retry changes (-r). Continue flag -c is also switched on, so we can continue unfinished jobs. Luckily, it’s accomplished in one run :) 

$ for f in $(seq 2014 2016); do 
  /bin/rm -rf $f; mkdir -p $f; 
  cat $f.html.txt | rush -v d=$f -d = 
    'phantomjs save_page.js "{1}" > {d}/(3).html' 
    -j 20 -t 60 -r 3 -c; 
done

GNU parallel can append to an existing joblog with ‘+’:

$ for f in $(seq 2014 2016); do 
  /bin/rm -rf $f; mkdir -p $f;
  cat $f.html.txt | parallel -j20 --timeout 60 --retries 4 --joblog +mylog \ 
  --colsep = \ 
    phantomjs save_page.js {1}={2}={3} '>' $f/(3).html
done

18. A bioinformatics example: mapping with `bwa`, and processing result with `samtools`:

$ ref=ref/xxx.fa
$ threads=25
$ ls -d raw.cluster.clean.mapping/* 
  | rush -v ref=$ref -v j=$threads -v p='{}({})' 
  'bwa mem -t [j] -M -a {ref} (p)_1.fq.gz (p)_2.fq.gz >{p}.sam;\ samtools view -bS (p).sam > (p).bam; 
  samtools sort -T (p).tmp -@ (j) (p).bam -o (p).sorted.bam; 
  samtools index (p).sorted.bam; 
  samtools flagstat (p).sorted.bam > (p).sorted.bam.flagstat; \ 
  /bin/rm (p).bam (p).sam;'
    -j 2 --verbose -c -C mapping.rush
```
GNU **parallel** would use a function:

```bash
$ ref=ref/xxx.fa
$ export ref
$ thr=25
$ export thr
$bwa_sam() {
    p="$1"
    bam="$p".bam
    sam="$p".sam
    sortbam="$p".sorted.bam
    bwa mem -t $thr -M -a $ref ${p}_1.fq.gz ${p}_2.fq.gz > "$sam"
    samtools view -bS "$sam" > "$bam"
    samtools sort -T ${p}.tmp -@ $thr "$bam" -o "$sortbam"
    samtools index "$sortbam"
    samtools flagstat "$sortbam" > "$sortbam".flagstat
    /bin/rm "$bam" "$sam"
}
$ export -f bwa_sam
$ ls -d raw.cluster.clean.mapping/* | parallel --j 2 --verbose --joblog mylog bwa_sam
```

**Other rush features**

**rush** has:

* **awk -v** like custom defined variables (-v)

  With GNU **parallel** you would simply set a shell variable:

  ```bash
  parallel 'v={}; echo "$v"' :::: foo
echo foo | rush -v v={} 'echo (v)"
  ```

  Also **rush** does not like special chars. So these do not work:

  ```bash
echo does not work | rush -v v="" 'echo (v)"
echo "My brother's 12" records" | rush -v v={} 'echo (v)"
  ```

  Whereas the corresponding GNU **parallel** version works:

  ```bash
  parallel 'v=""'; echo "$v"' :::: works
  parallel 'v={}; echo "$v"' :::: "My brother's 12" records"
  ```

* Exit on first error(s) (-e)

  This is called --halt now,fail=1 (or shorter: --halt 2) when used with GNU **parallel**.

* Settable records sending to every command (-n, default 1)

  This is also called -n in GNU **parallel**.

* Practical replacement strings

  `{}` remove any extension

    With GNU **parallel** this can be emulated by:

    ```bash
    parallel --plus echo '{/\.*}'} :::: foo.ext.bar.gz
    ```

  `{^suffix}`, remove suffix

    With GNU **parallel** this can be emulated by:

    ```bash
    parallel --plus echo '{%.bar.gz}'} :::: foo.ext.bar.gz
    ```
(@regexp), capture submatch using regular expression

With GNU parallel this can be emulated by:

```bash
parallel --rpl '(@(.*)) /$1/ and $=_$1;' \n   echo '@d_(.*)\.gz' :::: l_foo.gz
```

{%.}, {%.}, basename without extension

With GNU parallel this can be emulated by:

```bash
parallel echo '={s::*/::;s/\../*/ =}' :::: dir/foo.bar.gz
```

And if you need it often, you define a --rpl in $HOME/parallel/config:

```bash
--rpl '{.} s::*/::;s/\../*/'
--rpl '{.} s::*/::;s/\../*/'
```

Then you can use them as:

```bash
parallel echo {.} {.} :::: dir/foo.bar.gz
```

* Preset variable (macro)

E.g.

```bash
echo foosuffix | rush -v p='^suffix' 'echo {p}_new_suffix'
```

With GNU parallel this can be emulated by:

```bash
echo foosuffix |
parallel --plus 'p=${suffix}; echo ${p}_new_suffix'
```

Opposite rush GNU parallel works fine if the input contains double space, ’ and ":

```bash
echo "1'6" foosuffix |
parallel --plus 'p=${suffix}; echo "$p"_new_suffix'
```

* Commands of multi-lines

While you can use multi-lined commands in GNU parallel, to improve readability GNU parallel discourages the use of multi-line commands. In most cases it can be written as a function:

```bash
seq 1 3 |
parallel --timeout 2 --joblog my.log 'sleep {}; echo {}; \n   echo finish {}'
```

Could be written as:

```bash
doit() {
   sleep "$1"
   echo "$1"
   echo finish "$1"
}
export -f doit
seq 1 3 | parallel --timeout 2 --joblog my.log doit
```

The failed commands can be resumed with:

```bash
seq 1 3 |
parallel --resume-failed --joblog my.log 'sleep {}; echo {};\n   echo finish {}'
```

https://github.com/shenwei356/rush
DIFFERENCES BETWEEN ClusterSSH AND GNU Parallel

ClusterSSH solves a different problem than GNU parallel.

ClusterSSH opens a terminal window for each computer and using a master window you can run the same command on all the computers. This is typically used for administrating several computers that are almost identical.

GNU parallel runs the same (or different) commands with different arguments in parallel possibly using remote computers to help computing. If more than one computer is listed in -S GNU parallel may only use one of these (e.g. if there are 8 jobs to be run and one computer has 8 cores).

GNU parallel can be used as a poor-man's version of ClusterSSH:

```
parallel --nonall -S server-a,server-b do_stufoo bar
```

https://github.com/duncs/clusterssh

DIFFERENCES BETWEEN coshell AND GNU Parallel

coshell only accepts full commands on standard input. Any quoting needs to be done by the user.

Commands are run in sh so any bash/tcsh/zsh specific syntax will not work.

Output can be buffered by using -d. Output is buffered in memory, so big output can cause swapping and therefore be terrible slow or even cause out of memory.

https://github.com/gdm85/coshell (Last checked: 2019-01)

DIFFERENCES BETWEEN spread AND GNU Parallel

spread runs commands on all directories.

It can be emulated with GNU parallel using this Bash function:

```
spread() {
  _cmds() {
    perl -e '$"=" && \";print ":ARGV\"\";cd \"{}\" \"\"' 
    parallel ${_cmds "$@"} | | echo exit status $? :: : */
  }

  This works except for the --exclude option.

  (Last checked: 2017-11)
```

DIFFERENCES BETWEEN pyargs AND GNU Parallel

pyargs deals badly with input containing spaces. It buffers stdout, but not stderr. It buffers in RAM. {} does not work as replacement string. It does not support running functions.

pyargs does not support composed commands if run with --lines, and fails on pyargs traceroute gnu.org fsf.org.

Examples

```
seq 5 | pyargs -P50 -L seq
seq 5 | parallel -P50 --lb seq

seq 5 | pyargs --mark -L seq
seq 5 | parallel -P50 --lb --tagstring OUTPUT'[= $_=$job->replaced()]' seq
# Similar, but not precisely the same
seq 5 | parallel -P50 --lb --tag seq
```
GNU Parallel alternatives

GNU Parallel alternatives

```bash
seq 5 | pyargs -P50 --mark command
# Somewhat longer with GNU Parallel due to the special
# --mark formatting
cmd="$(echo "command" | parallel --shellquote)"
wrap_cmd() {
  echo "MARK $cmd $0="
  eval $cmd $0
  echo "OUTPUT END[$cmd $0]"
}
(seq 5 | env_parallel -P2 wrap_cmd) 3>&1
# Similar, but not exactly the same
seq 5 | parallel -t --tag command

(seq 5 | env_parallel -P2 wrap_cmd) 3>&1
# Similar, but not exactly the same
seq 5 | parallel -t --tag command

https://github.com/robertblackwell/pyargs (Last checked: 2019-01)

DIFFERENCES BETWEEN concurrently AND GNU Parallel

concurrently runs jobs in parallel.
The output is prepended with the job number, and may be incomplete:

```
$ concurrently 'seq 100000' | (sleep 3;wc -l)
7165
```

When pretty printing it caches output in memory. Output mixes by using test MIX below whether or not output is cached.

There seems to be no way of making a template command and have concurrently fill that with different args. The full commands must be given on the command line.

There is also no way of controlling how many jobs should be run in parallel at a time - i.e. "number of jobslots". Instead all jobs are simply started in parallel.

https://github.com/kimmobrunfeldt/concurrently (Last checked: 2019-01)

DIFFERENCES BETWEEN map(soveran) AND GNU Parallel

map does not run jobs in parallel by default. The README suggests using:

```
... | map t 'sleep $t && say done &'
```

But this fails if more jobs are run in parallel than the number of available processes. Since there is no support for parallelization in map itself, the output also mixes:

```
seq 10 | map i 'echo start-$i && sleep 0.$i && echo end-$i &'
```

The major difference is that GNU parallel is built for parallelization and map is not. So GNU parallel has lots of ways of dealing with the issues that parallelization raises:

- Keep the number of processes manageable
- Make sure output does not mix
- Make Ctrl-C kill all running processes
Here are the 5 examples converted to GNU Parallel:

1$ ls *.c | map f 'foo $f'
1$ ls *.c | parallel foo

2$ ls *.c | map f 'foo $f; bar $f'
2$ ls *.c | parallel 'foo {}; bar {}'

3$ cat urls | map u 'curl -O $u'
3$ cat urls | parallel curl -O

4$ printf "1
1
1
" | map t 'sleep $t && say done'
4$ printf "1
1
1
" | parallel 'sleep () && say done' ::: 1 1 1

5$ printf "1
1
1
" | map t 'sleep $t && say done &'
5$ printf "1
1
1
" | parallel -j0 'sleep () && say done' ::: 1 1 1

https://github.com/soveran/map (Last checked: 2019-01)

DIFFERENCES BETWEEN loop AND GNU Parallel

**loop** mixes stdout and stderr:

```
loop 'ls /no-such-file' >/dev/null
```

**loop**'s replacement string $ITEM does not quote strings:

```
echo 'two spaces' | loop 'echo $ITEM'
```

**loop** cannot run functions:

```
myfunc() { echo joe; }
export -f myfunc
loop 'myfunc this fails'
```

Some of the examples from https://github.com/Miserlou/Loop/ can be emulated with GNU parallel:

```
# A couple of functions will make the code easier to read
$ loopy() {
    yes | parallel -uN0 -j1 "$@"
}
$ export -f loopy
$ time_out() {
    parallel -uN0 -q --timeout "$@" ::: 1
}
$ match() {
    perl -0777 -ne 'grep /"$1"/,$_ and print or exit 1'
}
$ export -f match

$ loop 'ls' --every 10s
$ loopy --delay 10s ls

$ loop 'touch $COUNT.txt' --count-by 5
```
$ loopy touch '($= $=seq()*5 =)'.txt

$ loop --until-contains 200 --\n  ./get_response_code.sh --site mysite.biz
$ loopy --halt now,success=1 \n  ./get_response_code.sh --site mysite.biz | match 200'

$ loop './poke_server' --for-duration 8h
$ time_out 8h loopy ./poke_server

$ loop './poke_server' --until-success
$ loopy --halt now,success=1 ./poke_server

$ cat files_to_create.txt | loop 'touch $ITEM'
$ cat files_to_create.txt | parallel touch {}

$ loop 'ls' --for-duration 10min --summary
# --joblog is somewhat more verbose than --summary
$ time_out 10m loopy --joblog verbose ./poke_server; cat my.log

$ loop 'echo hello'
$ loopy echo hello

$ loop 'echo $COUNT'
# GNU Parallel counts from 1
$ loopy echo {}
# Counting from 0 can be forced
$ loopy echo '($= $=seq()-1 =)'

$ loop 'echo $COUNT' --count-by 2
$ loopy echo '($= $=2*(seq()-1) =)'

$ loop 'echo $COUNT' --count-by 2 --offset 10
$ loopy echo '($= $=10+2*(seq()-1) =)'

$ loop 'echo $COUNT' --count-by 1.1
# GNU Parallel rounds 3.3000000000000003 to 3.3
$ loopy echo '($= $=1.1*(seq()-1) =)'

$ loop 'echo $COUNT $ACTUALCOUNT' --count-by 2
$ loopy echo '($= $=2*(seq()-1) =) {}

$ loop 'echo $COUNT' --num 3 --summary
# --joblog is somewhat more verbose than --summary
$ seq 3 | parallel --joblog verbose my.log echo; cat my.log

$ loop 'ls -foobartz' --num 3 --summary
# --joblog is somewhat more verbose than --summary
$ seq 3 | parallel --joblog verbose my.log -N0 ls -foobartz; cat my.log

$ loop 'echo $COUNT' --count-by 2 --num 50 --only-last
# Can be emulated by running 2 jobs
$ seq 49 | parallel echo '{= $_=2*(seq()-1) =}' >/dev/null
$ echo 50 | parallel echo '{= $_=2*(seq()-1) =}'

$ loop 'date' --every 5s
$ loopy --delay 5s date

$ loop 'date' --for-duration 8s --every 2s
$ time_out 8s loopy --delay 2s date

$ loop 'date -u' --until-time '2018-05-25 20:50:00' --every 5s
$ seconds=$('date -d 2018-05-25T20:50:00 +%s` - `date +%s`)
$ time_out $seconds loopy --delay 5s date -u

$ loop 'echo $RANDOM' --until-contains "666"
$ loopy --halt now,success=1 'echo $RANDOM | match 666'

$ loop 'if (( RANDOM % 2 )); then
  (echo "TRUE"; true);
else
  (echo "FALSE"; false);
fi' --until-success
$ loopy --halt now,success=1 'if (( RANDOM % 2 )); then
  (echo "TRUE"; true);
else
  (echo "FALSE"; false);
fi'

$ loop 'if (( RANDOM % 2 )); then
  (echo "TRUE"; true);
else
  (echo "FALSE"; false);
fi' --until-error
$ loopy --halt now,fail=1 'if (( RANDOM % 2 )); then
  (echo "TRUE"; true);
else
  (echo "FALSE"; false);
fi'

$ loop 'date' --until-match "(\d{4})"
$ loopy --halt now,success=1 'date | match [0-9][0-9][0-9][0-9][0-9][0-9]'

$ loop 'echo $ITEM' --for red,green,blue
$ parallel echo ::: red green blue

$ cat /tmp/my-list-of-files-to-create.txt | loop 'touch $ITEM'
$ cat /tmp/my-list-of-files-to-create.txt | parallel touch

$ ls | loop 'cp $ITEM $ITEM.bak'; ls
$ ls | parallel cp {} {} .bak; ls

$ loop 'echo $ITEM | tr a-z A-Z' -i
$ parallel 'echo {} | tr a-z A-Z'
# Or more efficiently:
GNU Parallel alternatives

$ parallel --pipe tr a-z A-Z

$ loop 'echo $ITEM' --for "$ls" $ parallel echo {} ::: "$ls"

$ ls | loop './my_program $ITEM' --until-success; $ ls | parallel --halt now,success=1 ./my_program {}

$ ls | loop './my_program $ITEM' --until-fail; $ ls | parallel --halt now,fail=1 ./my_program {}

$ ./deploy.sh; loop 'echo $ITEM' --for "`ls`" $ ls | loop './my_program $ITEM' --until-success; $ ls | parallel --halt now,success=1 ./my_program {} $ ls | loop './my_program $ITEM' --until-fail; $ ls | parallel --halt now,fail=1 ./my_program {} $ ./deploy.sh; loop 'echo $ITEM' --for "`ls`" $ ls | parallel echo {} ::: "`ls`" $ ls | loop './my_program $ITEM' --until-success; $ ls | parallel --halt now,success=1 ./my_program {} $ ls | loop './my_program $ITEM' --until-fail; $ ls | parallel --halt now,fail=1 ./my_program {}

$ ./deploy.sh; loopy --delay 5s --halt now,success=1 "`ls`" $ ./deploy.sh; loopy --delay 5s --halt now,success=1 "`ls`" $ ./deploy.sh; loopy --delay 5s --halt now,success=1 "`ls`" $ ls | loop 'curl -sw "%{http_code}" http://coolwebsite.biz' \ --every 5s --until-contains 200; ./announce_to_slack.sh $ ./deploy.sh; loopy --delay 5s --halt now,success=1 \ 'curl -sw "%{http_code}" http://coolwebsite.biz | match 200'; ./announce_to_slack.sh

$ ./soothing_is_food.sh $ ls | loop 'curl -sw "%{http_code}" http://coolwebsite.biz' \ --every 5s --until-contains 200; ./announce_to_slack.sh

$ ls | loop 'ping -c 1 mysite.com' --until-success; ./do_next_thing $ loopy --halt now,success=1 ping -c 1 mysite.com; ./do_next_thing

$ ./create_big_file -o my_big_file.bin; loop 'ls' --until-contains 'my_big_file.bin'; ./upload_big_file my_big_file.bin # inotifywait is a better tool to detect file system changes. # It can even make sure the file is complete # so you are not uploading an incomplete file $ inotifywait -q -m -r -e MOVED_TO -e CLOSE_WRITE --format %w%f . | grep my_big_file.bin

$ ls | loop 'cp $ITEM $ITEM.bak' $ ls | parallel cp {} {}.bak

$ ./deploy.sh; loop 'curl -sw "%{http_code}" http://coolwebsite.biz' \ --every 5s --until-contains 200; ./announce_to_slack.sh

https://github.com/Miserlou/Loop/ (Last checked: 2018-10)

DIFFERENCES BETWEEN lorikeet AND GNU Parallel

lorikeet can run jobs in parallel. It does this based on a dependency graph described in a file, so this is similar to make. https://github.com/cetra3/lorikeet (Last checked: 2018-10)

DIFFERENCES BETWEEN spp AND GNU Parallel

spp can run jobs in parallel. spp does not use a command template to generate the jobs, but requires jobs to be in a file. Output from the jobs mix. https://github.com/john01dav/spp (Last checked: 2019-01)

DIFFERENCES BETWEEN paral AND GNU Parallel

paral prints a lot of status information and stores the output from the commands run into files. This means it cannot be used the middle of a pipe like this
GNU Parallel alternatives

Instead it puts the output into files named like `out_#_command.out.log`. To get a very similar behaviour with GNU parallel use `--results 'out_#_\{=s/[^a-z_0-9]/\}g;g/s+/g=\}.log' --eta`

parallel only takes arguments on the command line and each argument should be a full command. Thus it does not use command templates.

This limits how many jobs it can run in total, because they all need to fit on a single command line.

parallel has no support for running jobs remotely.

The examples from README.markdown and the corresponding command run with GNU parallel (--results 'out_#_\{=s/[^a-z_0-9]/g;g/s+/g=\}.log' --eta is omitted from the GNU parallel command):

```
parallel "command 1" "command 2 --flag" "command arg1 arg2"
```

```
parallel "sleep 1 && echo c1" "sleep 2 && echo c2" 
  "sleep 3 && echo c3" "sleep 4 && echo c4" "sleep 5 && echo c5"
```

# Or shorter:

```
parallel "sleep () && echo c{1}" ::: (1..5)
```

```
parallel -n=0 "sleep 5 && echo c5" "sleep 4 && echo c4" 
  "sleep 3 && echo c3" "sleep 2 && echo c2" "sleep 1 && echo c1"
```

# Or shorter:

```
parallel -j0 "sleep () && echo c{}" ::: 5 4 3 2 1
```

```
parallel -n=1 "sleep 5 && echo c5" "sleep 4 && echo c4" 
  "sleep 3 && echo c3" "sleep 2 && echo c2" "sleep 1 && echo c1"
```

# Or shorter:

```
parallel -j1 "sleep () && echo c{}" ::: 5 4 3 2 1
```

```
parallel -n=2 "sleep 5 && echo c5" "sleep 4 && echo c4" 
  "sleep 3 && echo c3" "sleep 2 && echo c2" "sleep 1 && echo c1"
```

# Or shorter:

```
parallel -j2 "sleep () && echo c{}" ::: 5 4 3 2 1
```

```
parallel -n=5 "sleep 5 && echo c5" "sleep 4 && echo c4" 
  "sleep 3 && echo c3" "sleep 2 && echo c2" "sleep 1 && echo c1"
```

# Or shorter:

```
parallel -j5 "sleep () && echo c{}" ::: 5 4 3 2 1
```

```
parallel -n=1 "echo a && sleep 0.5 && echo b && sleep 0.5 && 
  echo c && sleep 0.5 && echo d && sleep 0.5 && 
  echo e && sleep 0.5 && echo f && sleep 0.5 && 
  echo g && sleep 0.5 && echo h"
```

```
parallel ::: "echo a && sleep 0.5 && echo b && sleep 0.5 && 
  echo c && sleep 0.5 && echo d && sleep 0.5 && 
  echo e && sleep 0.5 && echo f && sleep 0.5 && 
  echo g && sleep 0.5 && echo h"
```

https://github.com/amattn/paral (Last checked: 2019-01)
DIFFERENCES BETWEEN concurr AND GNU Parallel

concurr is built to run jobs in parallel using a client/server model.

The examples from README.md:

```bash
concurr 'echo job {#} on slot %: {}' : arg1 arg2 arg3 arg4
parallel 'echo job {#} on slot %: {}' ::: arg1 arg2 arg3 arg4

concurr 'echo job {#} on slot %: {}' ::: file1 file2 file3
parallel 'echo job {#} on slot %: {}' ::: file1 file2 file3

concurr 'echo {}' < input_file
parallel 'echo {}' < input_file

cat file | concurr 'echo {}'
cat file | parallel 'echo {}'

concurr deals badly empty input files and with output larger than 64 KB.
https://github.com/mmstick/concurr (Last checked: 2019-01)

DIFFERENCES BETWEEN lesser-parallel AND GNU Parallel

lesser-parallel is the inspiration for parallel --embed. Both lesser-parallel and parallel --embed define bash functions that can be included as part of a bash script to run jobs in parallel.

lesser-parallel implements a few of the replacement strings, but hardly any options, whereas parallel --embed gives you the full GNU parallel experience.

https://github.com/kou1okada/lesser-parallel (Last checked: 2019-01)

DIFFERENCES BETWEEN npm-parallel AND GNU Parallel

npm-parallel can run npm tasks in parallel.

There are no examples and very little documentation, so it is hard to compare to GNU parallel.

https://github.com/spion/npm-parallel (Last checked: 2019-01)

DIFFERENCES BETWEEN machma AND GNU Parallel

machma runs tasks in parallel. It gives time stamped output. It buffers in RAM. The examples from README.md:

```
# Put shorthand for timestamp in config for the examples
echo '--rpl "
    "\'(time) $_=::strftime("%Y-%m-%d %H:%M:%S",localtime())'" \n    > ~/.parallel/machma
echo '--line-buffer --tagstring "(\#) {time} ()"' >> ~/.parallel/machma

find . -iname '*.jpg' |
machma -- mogrify -resize 1200x1200 -filter Lanczos {}
find . -iname '*.jpg' |
parallel --bar -Jmachma mogrify -resize 1200x1200 -filter Lanczos {}

cat /tmp/ips | machma -p 2 -- ping -c 2 -q {}
cat /tmp/ips | parallel -j2 -Jmachma ping -c 2 -q {}

cat /tmp/ips |
machma -- sh -c 'ping -c 2 -q $0 > /dev/null && echo alive' {}
```
cat /tmp/ips | parallel -Jmachma 'ping -c 2 -q {} > /dev/null & echo alive'

find . -iname '*.jpg' |
machma --timeout 5s -- mogrify -resize 1200x1200 -filter Lanczos {}
find . -iname '*.jpg' |
parallel --timeout 5s --bar mogrify -resize 1200x1200 \
   -filter Lanczos {}

find . -iname '*.jpg' -print0 |
machma --null -- mogrify -resize 1200x1200 -filter Lanczos {}
find . -iname '*.jpg' -print0 |
parallel --null --bar mogrify -resize 1200x1200 -filter Lanczos {}

https://github.com/fd0/machma (Last checked: 2019-06)

DIFFERENCES BETWEEN interlace AND GNU Parallel

Summary table (see legend above): - I2 I3 I4 - - M1 - M3 -- M6 - O2 O3 - - - x x E1 E2 - - - - - - - - - - - - - - - - - - - - -

interlace is built for network analysis to run network tools in parallel.

interface does not buffer output, so output from different jobs mixes.

The overhead for each target is O(n*n), so with 1000 targets it becomes very slow with an overhead in the order of 500ms/target.

Using prips most of the examples from https://github.com/codingo/Interlace can be run with GNU parallel:

Blocker

commands.txt:
   mkdir -p _output_/ _target_/scans/
   _blocker_
   nmap _target_ -oA _output_/ _target_/scans/_target_-nmap
interlace -tL ./targets.txt -cL commands.txt -o $output

parallel -a targets.txt \
   mkdir -p $output/{} /scans/; nmap {} -oA $output/{} /scans/{} -nmap

Blocks

commands.txt:
   _block:nmap_
   mkdir -p _target_/output/scans/
   nmap _target_ -oN _target_/output/scans/_target_-nmap
   _block:nmap_
   nikto --host _target_
interlace -tL ./targets.txt -cL commands.txt

   _nmap() {
      mkdir -p $1/output/scans/
      nmap $1 -oN $1/output/scans/$1-nmap
   }
   export -f _nmap
parallel :::: _nmap "nikto --host" :::: targets.txt

---
Run Nikto Over Multiple Sites

```
interlace -tL ./targets.txt -threads 5 \
  -c "nikto --host _target_ > ./_target_-nikto.txt" -v

parallel -a targets.txt -P5 nikto --host () \> ./{}/-nikto.txt
```

Run Nikto Over Multiple Sites and Ports

```
interlace -tL ./targets.txt -threads 5 -c \
  "nikto --host _target_:_port_ > ./_target_:_port_-nikto.txt" \
  -p 80,443 -v

parallel -P5 nikto --host (1):(2) \> .{1}-{2}-nikto.txt \ 
  ::: targets.txt ::: 80 443
```

Run a List of Commands against Target Hosts

```
commands.txt:
nikto --host _target_:_port_ > _output_//_target_-nikto.txt 
sslscan _target_:_port_ > _output_//_target_-sslscan.txt 
testssl.sh _target_:_port_ > _output_//_target_-testssl.txt 
interlace -t example.com -o ~/Engagements/example/ \
  -cL ./commands.txt -p 80,443

parallel --results ~/Engagements/example/{2}:{3}{1} {1} {2}:{3} \
  ::: "nikto --host" sslscan testssl.sh ::: example.com ::: 80 443
```

CIDR notation with an application that doesn't support it

```
interlace -t 192.168.12.0/24 -c "vhostscan _target_ \
  -oN _output_//_target_-vhosts.txt" -o ~/scans/ -threads 50

prrps 192.168.12.0/24 | 
  parallel -P50 vhostscan () -oN ~/scans/{}-vhosts.txt
```

Glob notation with an application that doesn't support it

```
interlace -t 192.168.12.* -c "vhostscan _target_ \
  -oN _output_//_target_-vhosts.txt" -o ~/scans/ -threads 50

# Glob is not supported in prrps 
prrps 192.168.12.0/24 | 
  parallel -P50 vhostscan () -oN ~/scans/{}-vhosts.txt
```

Dash (-) notation with an application that doesn't support it

```
interlace -t 192.168.12.1-15 -c \
"vhostscan _target_ -oN _output_//_target_-vhosts.txt" \
  -o ~/scans/ -threads 50

# Dash notation is not supported in prrps 
prrps 192.168.12.1 192.168.12.15 | 
  parallel -P50 vhostscan () -oN ~/scans/{}-vhosts.txt
Threading Support for an application that doesn't support it

```bash
interlace -tL ./target-list.txt -c \
  "vhostscan -t _target_ -oN _output_//_target_-vhosts.txt" \
-0 ~/scans/ -threads 50

```.

```
cat ./target-list.txt |
  parallel -P50 vhostscan -t {} -oN ~/scans/{}-vhosts.txt

``` alternatively

```
./vhosts-commands.txt:
  vhostscan -t $target -oN _output_//_target_-vhosts.txt
interlace -cL ./vhosts-commands.txt -tL ./target-list.txt \
-threads 50 -o ~/scans

```.

```
./vhosts-commands.txt:
  vhostscan -t "$1" -oN "$2"
parallel -P50 ./vhosts-commands.txt {} -oN ~/scans/{}-vhosts.txt \
::: ./target-list.txt
```

### Exclusions

```
interlace -t 192.168.12.0/24 -e 192.168.12.0/26 -c \
  "vhostscan _target_ -oN _output_//_target_-vhosts.txt" \
-0 ~/scans/ -threads 50

```.

```
prips 192.168.12.0/24 | grep -xv -Ff <(prips 192.168.12.0/26) | 
  parallel -P50 vhostscan () -oN ~/scans/{}-vhosts.txt

```.

### Run Nikto Using Multiple Proxies

```
interlace -tL ./targets.txt -pL ./proxies.txt -threads 5 -c \
  "nikto --host _target_: _port_ -useproxy _proxy_ > \ 
  _/target_/port_/nikto.txt" -p 80,443 -v

```.

```
parallel -j5 \
  "nikto --host (1):(2) -useproxy (3) > ./1-(2)-nikto.txt" \
  ::: ./targets.txt :::: 80 443 :::: ./proxies.txt
```

https://github.com/codingo/Interlace (Last checked: 2019-09)

**DIFFERENCES BETWEEN otonvm Parallel AND GNU Parallel**

I have been unable to get the code to run at all. It seems unfinished.

https://github.com/otonvm/Parallel (Last checked: 2019-02)

**DIFFERENCES BETWEEN k-bx par AND GNU Parallel**

`par` requires Haskell to work. This limits the number of platforms this can work on.

`par` does line buffering in memory. The memory usage is 3x the longest line (compared to 1x for `parallel --lb`). Commands must be given as arguments. There is no template.

These are the examples from https://github.com/k-bx/par with the corresponding GNU parallel command.

```
par "echo foo; sleep 1; echo foo; sleep 1; echo foo"
```

```
"echo bar; sleep 1; echo bar; sleep 1; echo bar" && echo "success"
par --lb ::: "echo foo; sleep 1; echo foo; sleep 1; echo foo" \
"echo bar; sleep 1; echo bar; sleep 1; echo bar" && echo "success"

par "echo foo; sleep 1; foofoo" \
"echo bar; sleep 1; foofoo" &&
par "PARPREFIX=[fooechoer] echo foo" "PARPREFIX=[bar] echo bar"
parallel --lb --halt 1 ::: "echo foo; sleep 1; echo foo; sleep 1; echo foo"

PARPREFIX=[fooechoer]
parallel --lb --colsep , --tagstring {1} {2} \
::: 
PARPREFIX=[bar], echo bar"

par --succeed "foo" "bar" && echo 'wow'
parallel "foo" "bar"; true && echo 'wow'

https://github.com/k-bx/par (Last checked: 2019-02)

DIFFERENCES BETWEEN parallelshell AND GNU Parallel
parallelshell does not allow for composed commands:

# This does not work
parallelshell 'echo foo;echo bar' 'echo baz;echo quuz'

Instead you have to wrap that in a shell:

parallelshell 'sh -c "echo foo;echo bar"' 'sh -c "echo baz;echo quuz"'

It buffers output in RAM. All commands must be given on the command line and all commands are started in parallel at the same time. This will cause the system to freeze if there are so many jobs that there is not enough memory to run them all at the same time.

https://github.com/keithamus/parallelshell (Last checked: 2019-02)


DIFFERENCES BETWEEN shell-executor AND GNU Parallel
shell-executor does not allow for composed commands:

# This does not work
sx 'echo foo;echo bar' 'echo baz;echo quuz'

Instead you have to wrap that in a shell:

sx 'sh -c "echo foo;echo bar"' 'sh -c "echo baz;echo quuz"

It buffers output in RAM. All commands must be given on the command line and all commands are started in parallel at the same time. This will cause the system to freeze if there are so many jobs that there is not enough memory to run them all at the same time.


DIFFERENCES BETWEEN non-GNU par AND GNU Parallel

par buffers in memory to avoid mixing of jobs. It takes 1s per 1 million output lines.

par needs to have all commands before starting the first job. The jobs are read from stdin (standard input) so any quoting will have to be done by the user.
Stdout (standard output) is prepended with o:. Stderr (standard error) is sent to stdout (standard output) and prepended with e:.

For short jobs with little output par is 20% faster than GNU parallel and 60% slower than xargs.


DIFFERENCES BETWEEN fd AND GNU Parallel

fd does not support composed commands, so commands must be wrapped in sh -c.

It buffers output in RAM.

It only takes file names from the filesystem as input (similar to find).

https://github.com/sharkdp/fd (Last checked: 2019-02)

DIFFERENCES BETWEEN lateral AND GNU Parallel

lateral is very similar to sem: It takes a single command and runs it in the background. The design means that output from parallel running jobs may mix. If it dies unexpectedly it leaves a socket in ~/.lateral/socket.PID.

lateral deals badly with too long command lines. This makes the lateral server crash:

    lateral run echo 'seq 100000| head -c 1000k'

Any options will be read by lateral so this does not work (lateral interprets the -l):

    lateral run ls -l

Composed commands do not work:

    lateral run pwd ';' ls

Functions do not work:

    myfunc() { echo a; }
    export -f myfunc
    lateral run myfunc

Running emacs in the terminal causes the parent shell to die:

    echo '#!/bin/bash' > mycmd
    echo emacs -nw >> mycmd
    chmod +x mycmd
    lateral start
    lateral run ./mycmd

Here are the examples from https://github.com/akramer/lateral with the corresponding GNU sem and GNU parallel commands:

    1$ lateral start
    1$ for i in $(cat /tmp/names); do
    1$   lateral run -- some_command $i
    1$ done
    1$ lateral wait
    1$
    1$ for i in $(cat /tmp/names); do
    1$   sem some_command $i
    1$ done
1$ sem --wait
1$
1$ parallel some_command ::: /tmp/names

2$ lateral start
2$ for i in $(seq 1 100); do
2$ lateral run -- my_slow_command < workfile$i > /tmp/logfile$i
2$ done
2$ lateral wait
2$
2$ for i in $(seq 1 100); do
2$ sem my_slow_command < workfile$i > /tmp/logfile$i
2$ done
2$ sem --wait
2$
2$ parallel 'my_slow_command < workfile{} > /tmp/logfile{}' \
::: (1..100)

3$ lateral start -p 0 # yup, it will just queue tasks
3$ for i in $(seq 1 100); do
3$ lateral run -- command_still_outputs_but_wont_spam inputfile$i
3$ done
3$ # command output spam can commence
3$ lateral config -p 10; lateral wait
3$
3$ for i in $(seq 1 100); do
3$ echo "command inputfile$i" >> joblist
3$ done
3$ parallel -j 10 :::: joblist
3$
3$ echo 1 > /tmp/njobs
3$ parallel -j /tmp/njobs command inputfile{} \ 
::: (1..100) &
3$ echo 10 >/tmp/njobs
3$ wait

https://github.com/akramer/lateral (Last checked: 2019-03)

DIFFERENCES BETWEEN with-this AND GNU Parallel
The examples from https://github.com/amritb/with-this.git and the corresponding GNU parallel command:

```bash
with -v "$(cat myurls.txt)" "curl -L this"
parallel curl -L ::: myurls.txt

with -v "$(cat myregions.txt)" \
"aws --region=this ec2 describe-instance-status"
parallel aws --region={} ec2 describe-instance-status \ 
::: myregions.txt

with -v "$({ls})" "kubectl --kubeconfig=this get pods"
ls | parallel kubectl --kubeconfig={} get pods

with -v "$({ls | grep config})" "kubectl --kubeconfig=this get pods"
ls | grep config | parallel kubectl --kubeconfig={} get pods
```
with -v "$(echo \{1..10\})" "echo 123"
parallel -N0 echo 123 ::: \{1..10\}

Stderr is merged with stdout. **with-this** buffers in RAM. It uses 3x the output size, so you cannot have output larger than 1/3rd the amount of RAM. The input values cannot contain spaces. Composed commands do not work.

**with-this** gives some additional information, so the output has to be cleaned before piping it to the next command.

https://github.com/amritb/with-this.git (Last checked: 2019-03)

---

**DIFFERENCES BETWEEN Tollef’s parallel (moreutils) AND GNU Parallel**

Summary table (see legend above):

- - - - I4 - - I7 - - M3 - - M6 - O2 O3 - O5 O6 - x x E1 - - - - E7 - x x

---

**EXAMPLES FROM Tollef’s parallel MANUAL**

Tollef parallel sh -c "echo hi; sleep 2; echo bye" -- 1 2 3

GNU parallel "echo hi; sleep 2; echo bye" ::: 1 2 3

Tollef parallel -j 3 ufraw -o processed - *.NEF

GNU parallel -j 3 ufraw -o processed ::: *.NEF

Tollef parallel -j 3 -- ls df "echo hi"

GNU parallel -j 3 ::: ls df "echo hi"

(Last checked: 2019-08)

---

**DIFFERENCES BETWEEN rargs AND GNU Parallel**

Summary table (see legend above): I1 - - - - I4 - - I7 - - M3 - - M6 - O2 O3 - O5 O6 - O8 - E1 - - E4 - - - - - - - - -

---

rargs has elegant ways of doing named regexp capture and field ranges.

With GNU parallel you can use --rpl to get a similar functionality as regexp capture gives, and use join and @arg to get the field ranges. But the syntax is longer. This:

```
--rpl \'(r((-?\d+)?)\.)\.(\d+)\) $a=join"$opt::colsep",@arg[$$1..$$2]'
```

would make it possible to use:

```
{1r3..6}
```

for field 3..6.

For full support of \{n..m:s\} including negative numbers use a dynamic replacement string like this:

```
 PARALLEL=--rpl\'\'(r((-?\d+)?))\.[(\d+)]((\d+))?\)\) $a = defined $$2 ? $$2 < 0 ? 1+$#arg+$#2 : $$2 : 1; $b = defined $$4 ? $$4 < 0 ? 1+$#arg+$#4 : $$4 : $$4 + 1; $s = defined $$6 ? $$7 : ""; $s = join $$s, @arg[$$a..$$b]\' \n```

You can then do:

```
head /etc/passwd | parallel --colsep : echo \{1r3..\} \{3=\(1r\..3\) 4..=\(1r4..\) 2..4=\(1r2..4\) 3..3=\(1r3..3\) ..3:=-\(1r3..3\) ..3:=/\(1r3..3\) /-1=\(-1\) -5=\(-5\) -6=\(-6\) -3..=\(1r3..\) 
```
EXAMPLES FROM rargs MANUAL

```bash
ls *.bak | rargs -p "{.*}\.bak" mv {0} {1}
ls *.bak | parallel mv {0} {1}
```

cat download-list.csv | rargs -p "(?P<url>.*),(?P<filename>.*),\s+" wget {url} -O {filename}
cat download-list.csv | parallel --csv wget {1} -O {filename} # or use regexps:
cat download-list.csv | parallel --rpl '{url} s/,.*//' --rpl '{filename} s/.*?,//' wget {url} -O {filename}
cat /etc/passwd | rargs -d: echo -e 'id: "{1}\tname: "{5}\trest: "{6..::}"'
cat /etc/passwd | parallel -q --colsep ':' echo -e 'id: "{1}\tname: "{5}\trest: "{=6
$_=join":",@arg[6..$#arg]=\}"'

https://github.com/lotabout/rargs (Last checked: 2020-01)

DIFFERENCES BETWEEN threader AND GNU Parallel

Summary table (see legend above): I1 - - - - - M1 - M3 - - M6 O1 - O3 - O5 - - N/A N/A E1 - - E4 - - - - - - - - - -

Newline separates arguments, but newline at the end of file is treated as an empty argument. So this runs 2 jobs:

```bash
echo two_jobs | threader -run 'echo "$THREADID"'
```

threader ignores stderr, so any output to stderr is lost. threader buffers in RAM, so output bigger than the machine's virtual memory will cause the machine to crash.

https://github.com/voodooEntity/threader (Last checked: 2020-04)

DIFFERENCES BETWEEN runp AND GNU Parallel

Summary table (see legend above): I1 I2 - - - - - M1 - (M3) - - M6 O1 O2 O3 - O5 O6 - N/A N/A - E1 - - - - - - - - - -

(M3): You can add a prefix and a postfix to the input, so it means you can only insert the argument on the command line once.

runp runs 10 jobs in parallel by default. runp blocks if output of a command is > 64 Kbytes. Quoting of input is needed. It adds output to stderr (this can be prevented with -q)

Examples as GNU Parallel

```bash
base='https://images-api.nasa.gov/search'
query='jupiter'
desc='planet'
type='image'
url="$base?q=$query&description=$desc&media_type=$type"

# Download the images in parallel using runp
curl -s $url | jq -r .collection.items[].href | |
runp -p 'curl -s' | jq -r [] | grep large | 
runp -p 'curl -s -L -O'
```
time curl -s $url | jq -r .collection.items[].href | \ 
  runp -g 1 -q -p 'curl -s' | jq -r .[] | grep large | \ 
  runp -g 1 -q -p 'curl -s -L -O'

# Download the images in parallel
curl -s $url | jq -r .collection.items[].href | \ 
  parallel curl -s | jq -r .[] | grep large | \ 
  parallel curl -s -L -O

time curl -s $url | jq -r .collection.items[].href | \ 
  parallel -j 1 curl -s | jq -r .[] | grep large | \ 
  parallel -j 1 curl -s -L -O

Run some test commands (read from file)

# Create a file containing commands to run in parallel.
cat << EOF > /tmp/test-commands.txt
sleep 5
sleep 3
blah    # this will fail
ls $PWD   # PWD shell variable is used here
EOF

# Run commands from the file.
runp /tmp/test-commands.txt > /dev/null

parallel -a /tmp/test-commands.txt > /dev/null

Ping several hosts and see packet loss (read from stdin)

# First copy this line and press Enter
runp -p 'ping -c 5 -W 2' -s ' | grep loss'
localhost
1.1.1.1
8.8.8.8
# Press Enter and Ctrl-D when done entering the hosts

# First copy this line and press Enter
parallel ping -c 5 -W 2 () ' | grep loss'
localhost
1.1.1.1
8.8.8.8
# Press Enter and Ctrl-D when done entering the hosts

Get directories' sizes (read from stdin)

echo -e "$HOME\n/etc\n/tmp" | runp -q -p 'sudo du -sh'

echo -e "$HOME\n/etc\n/tmp" | parallel sudo du -sh
# or:
parallel sudo du -sh ::: "$HOME" /etc /tmp

Compress files

find . -iname '*.txt' | runp -p 'gzip --best'
find . -iname '*.txt' | parallel gzip --best

Measure HTTP request + response time

```bash
export CURL="\"curl -w \"time_total: %{time_total}\n\"\"
CURL=$CURL -o /dev/null -s https://golang.org/
perl -wE 'for (1..10) { say $ENV\{CURL\} }' | runp -q # Make 10 requests

perl -wE 'for (1..10) { say $ENV\{CURL\} }' | parallel
# or:
parallel -N0 "$CURL" ::: {1..10}
```

Find open TCP ports

```bash
cat << EOF > /tmp/host-port.txt
localhost 22
localhost 80
localhost 81
127.0.0.1 443
127.0.0.1 444
scanme.nmap.org 22
scanme.nmap.org 23
scanme.nmap.org 443
EOF

cat /tmp/host-port.txt | \
    runp -q -p \"netcat -v -w2 -z\" 2>&1 | egrep \'(succeeded!|open)\$\''

# --colsep is needed to split the line

cat /tmp/host-port.txt | \
    parallel --colsep \' \"netcat -v -w2 -z\" 2>&1 \| egrep \'(succeeded!|open)\$\''
# or use uq for unquoted:

cat /tmp/host-port.txt | \
    parallel netcat -v -w2 -z \{=uq=\} 2>&1 \| egrep \'(succeeded!|open)\$\''
```

https://github.com/jreisinger/runp (Last checked: 2020-04)

Todo

- https://pypi.org/project/papply/
- https://github.com/JeiKeiLim/simple_distribute_job
- https://github.com/reggi/pkgrun
- https://github.com/benoror/better-npm-run - not obvious how to use
- https://github.com/bahmutov/with-package
- https://github.com/xuchenCN/go-pssh
- https://github.com/flesler/parallel
- https://github.com/Julian/Verge
- https://github.com/ExpectationMax/simple_gpu_scheduler
- simple_gpu_scheduler --gpus 0 1 2 < gpu_commands.txt parallel -j3 --shuf CUDA_VISIBLE_DEVICES='\={1 $:\_slot()-1 =}{=uq:=}\}' < gpu_commands.txt
simple_hypersearch "python3 train_dnn.py --lr {lr} --batch_size {bs}" -p lr 0.001 0.0005 0.0001 -p bs 32 64 128 | simple_gpu_scheduler --gpus 0,1,2
parallel --header : --shuf -j3 -v CUDA_VISIBLE_DEVICES='(=1 $_=slot()-1 =)' python3 train_dnn.py --lr {lr} --batch_size {bs} ::: lr 0.001 0.0005 0.0001 ::: bs 32 64 128

simple_hypersearch "python3 train_dnn.py --lr {lr} --batch_size {bs}" --n-samples 5 -p lr 0.001 0.0005 0.0001 -p bs 32 64 128 |
parallel --header : --shuf CUDA_VISIBLE_DEVICES='(=1 $_=slot()-1; seq() > 5 and skip() =)'' python3 train_dnn.py --lr {lr} --batch_size {bs} ::: lr 0.001 0.0005 0.0001 ::: bs 32 64 128

touch gpu.queue
tail -f -n 0 gpu.queue | simple_gpu_scheduler --gpus 0,1,2 &
echo "my_command_with | and stuff > logfile" >> gpu.queue

touch gpu.queue
tail -f -n 0 gpu.queue | parallel -j3 CUDA_VISIBLE_DEVICES='(=1 $_=slot()-1 =) (=ug;=)'' &
  # Needed to fill job slots once
  seq 3 | parallel echo true >> gpu.queue
  # Add jobs
  echo "my_command_with | and stuff > logfile" >> gpu.queue
  # Needed to flush output from completed jobs
  seq 3 | parallel echo true >> gpu.queue

**TESTING OTHER TOOLS**

There are certain issues that are very common on parallelizing tools. Here are a few stress tests. Be warned: If the tool is badly coded it may overload your machine.

**MIX: Output mixes**

Output from 2 jobs should not mix. If the output is not used, this does not matter; but if the output is used then it is important that you do not get half a line from one job followed by half a line from another job.

If the tool does not buffer, output will most likely mix now and then.

This test stresses whether output mixes.

```bash
#!/bin/bash

paralleltool="parallel -j0"

cat <<-EOF > mycommand
#!/bin/bash

# If a, b, c, d, e, and f mix: Very bad
perl -e 'print STDOUT "a"x3000_000,"'"
perl -e 'print STDERR "b"x3000_000,"'"
perl -e 'print STDOUT "c"x3000_000,"'"
perl -e 'print STDOUT "d"x3000_000,"'"
perl -e 'print STDOUT "e"x3000_000,"'"
perl -e 'print STDOUT "f"x3000_000,"'"
```
echo
echo >&2
EOF
chmod +x mycommand

# Run 30 jobs in parallel
seq 30 |
   $paralleltool ./mycommand > (tr -s abcdef 2> >(tr -s abcdef >&2))

# 'a c e' and 'b d f' should always stay together
# and there should only be a single line per job

STDERRMERGE: Stderr is merged with stdout
Output from stdout and stderr should not be merged, but kept separated.
This test shows whether stdout is mixed with stderr.

#!/bin/bash
paralleltool="parallel -j0"

# Run one job
# echo |
   $paralleltool ./mycommand > stdout 2> stderr
cat stdout
cat stderr

RAM: Output limited by RAM
Some tools cache output in RAM. This makes them extremely slow if the output is bigger than
physical memory and crash if the output is bigger than the virtual memory.

#!/bin/bash
paralleltool="parallel -j0"

cat <<'EOF' > mycommand
#!/bin/bash

# Generate 1 GB output
yes "`perl -e 'print "c\"x30_000\"'`" | head -c 1G
EOF
chmod +x mycommand
GNU Parallel alternatives

```bash
#!/bin/bash
paralleltool="parallel -j0"

# This should be a dir with less than 100 GB free space
smalldisk=/tmp/shm/parallel
TMPDIR="$smalldisk"
export TMPDIR

max_output() {
    # Force worst case scenario:
    # Make GNU Parallel only check once per second
    sleep 10
    # Generate 100 GB to fill $TMPDIR
    # Adjust if /tmp is bigger than 100 GB
    yes | head -c 100G >$TMPDIR/$$
    # Generate 10 MB output that will not be buffered due to full disk
    perl -e 'print "X"x10_000_000' | head -c 10M
    echo This part is missing from incomplete output
    sleep 2
    rm $TMPDIR/$$
    echo Final output
}

export -f max_output
seq 10 | $paralleltool max_output | tr -s X
```
SPCCHAR: Dealing badly with special file names.

It is not uncommon for users to create files like:

My brother's 12" *** record (costs $$$$).jpg

Some tools break on this.

#!/bin/bash

paralleltool=parallel

touch "My brother's 12" *** record (costs \$\$\$).jpg"
ls My*jpg | $paralleltool ls -l

COMPOSED: Composed commands do not work

Some tools require you to wrap composed commands into bash -c.

echo bar | $paralleltool echo foo';' echo {}

ONEREP: Only one replacement string allowed

Some tools can only insert the argument once.

echo bar | $paralleltool echo () foo {}

INPUTSIZE: Length of input should not be limited

Some tools limit the length of the input lines artificially with no good reason. GNU parallel does not:

perl -e 'print "foo."."x"x100_000_000' | parallel echo {}

GNU parallel limits the command to run to 128 KB due to execve(1):

perl -e 'print "x"x131_000' | parallel echo {} | wc

NUMWORDS: Speed depends on number of words

Some tools become very slow if output lines have many words.

#!/bin/bash

paralleltool=parallel

cat <<-EOF > mycommand
#!/bin/bash

# 10 MB of lines with 1000 words
yes "`seq 1000`" | head -c 10M
EOF
chmod +x mycommand

# Run 30 jobs in parallel
seq 30 | time $paralleltool -j0 ./mycommand > /dev/null
4GB: Output with a line > 4GB should be OK

#!/bin/bash

paralleltool="parallel -j0"

cat <<-EOF > mycommand
#!/bin/bash
perl -e '
$a="a"x1000_000; for(1..5000) { print \$a }'
EOF
chmod +x mycommand

# Run 1 job
seq 1 | $paralleltool ./mycommand | LC_ALL=C wc

AUTHOR

When using GNU parallel for a publication please cite:


This helps funding further development; and it won't cost you a cent. If you pay 10000 EUR you should feel free to use GNU Parallel without citing.

Copyright (C) 2007-10-18 Ole Tange, http://ole.tange.dk

Copyright (C) 2008-2010 Ole Tange, http://ole.tange.dk

Copyright (C) 2010-2020 Ole Tange, http://ole.tange.dk and Free Software Foundation, Inc.

Parts of the manual concerning xargs compatibility is inspired by the manual of xargs from GNU findutils 4.4.2.

LICENSE

This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 3 of the License, or at your option any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see <http://www.gnu.org/licenses/>.

Documentation license I

Permission is granted to copy, distribute and/or modify this documentation under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, with no Front-Cover Texts, and with no Back-Cover Texts. A copy of the license is included in the file fdl.txt.

Documentation license II

You are free:

to Share

to copy, distribute and transmit the work

to Remix
to adapt the work

Under the following conditions:

**Attribution**
You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

**Share Alike**
If you alter, transform, or build upon this work, you may distribute the resulting work only under the same, similar or a compatible license.

With the understanding that:

**Waiver**
Any of the above conditions can be waived if you get permission from the copyright holder.

**Public Domain**
Where the work or any of its elements is in the public domain under applicable law, that status is in no way affected by the license.

**Other Rights**
In no way are any of the following rights affected by the license:

- Your fair dealing or fair use rights, or other applicable copyright exceptions and limitations;
- The author's moral rights;
- Rights other persons may have either in the work itself or in how the work is used, such as publicity or privacy rights.

**Notice**
For any reuse or distribution, you must make clear to others the license terms of this work.

A copy of the full license is included in the file as cc-by-sa.txt.

**DEPENDENCIES**
GNU parallel uses Perl, and the Perl modules Getopt::Long, IPC::Open3, Symbol, IO::File, POSIX, and File::Temp. For remote usage it also uses rsync with ssh.

**SEE ALSO**
find(1), xargs(1), make(1), pexec(1), ppss(1), xjobs(1), prll(1), dxargs(1), mdm(1)