

Parallel and Multi-thread

GNU Parallel

Official: <https://www.gnu.org/software/parallel/>

Download: <http://ftp.gnu.org/gnu/parallel/>

Install

```
# CentOS 7
yum group install "Development Tools"
wget http://ftp.gnu.org/gnu/parallel/parallel-latest.tar.bz2
tar xjf parallel-latest.tar.bz2
cd parallel-*
./configure --prefix=/usr/local
make
make install
```

Don't change the --prefix if you want to use Man to view the manual of the command.

Use case: The common use case I have is to bzip a large file using Parallels. Files with 40 millions rows (8 GB) are compressed to 400MB bz2 files.

```
cat largefile.csv | /usr/local/bin/parallel --pipe -k bzip2 --best > largefile.bz2
```

Use case: my custom shell

One-liner)

```
# ./gen_mm_log_insert.v5.sh <input-file> <output-dir>
cat files.lst | parallel -j3 "./gen_mm_log_insert.v5.sh rows/{} output/"
```

Shell Script)

```
num_processes=3
ls $loaddir/mmsevent* | /usr/local/bin/parallel -j $num_processes "./gen_mm_log_insert.v5.sh {} $outputdir"
```

```
ls $logdir/mmsevent* | /usr/local/bin/parallel -j $num_processes 'a={}; name=${a##*/}; \' \
'./gen_mm_log_insert.v5.sh {} "'$outputdir'" 2>"$logdir/err.${name}.log"
```

With Bash

```
# Multiple files *.lst with lots of file paths such as
# cat 1.lst
# /path/to/file1
# /path/to/file2
# /path/to/file3
#

files="$@"

## An arbitrary limiting factor so that there are some free processes
## in case I want to run something else
num_processes=3
#
echo "Parallel Processing with $num_processes threads (`date "+%F %T`)"
for lst in $files
do
    for f in $(cat $lst | sed '/^#/d')
    do
        ((i=i%num_processes)); ((i++==0)) && wait
        yourshell.sh $f &
    done
    sleep 120
done
```

Revision #11

Created 7 June 2021 20:00:15 by Admin

Updated 9 June 2021 01:40:03 by Admin