

GraceTree: file and data retriever

João Encarnação¹

¹UTCSR

2nd Nov 2016

GraceTree

Problem: numerous files spread over numerous (sub) directories

Objective: quickly find files and retrieve data in them

Procedure: operate at the level of concepts (year, month, file type), not at the level of the file system

Note: this presentation is not an exhaustive description of this tool!

Invocation

Usage: `gracetree.rb -x COMMAND [options]`

`-p, --parameters-file PARFILE`

`-x, --execute COMMAND`

`-r, --root ROOT`

`-S, --sink SINK`

`-t, --filetype FILETYPE`

`-y, --year YEAR`

`-m, --month MONTH`

`-d, --day DAY`

`-j, --jobid JOBID`

`-a, --arc ARC`

`-s, --satellite SAT`

`-g, --pattern PATTERN`

`-c, --[no-]clean-grep`

`-?, --[no-]debug`

`-h, --help`

Defining file types

In the parameter file:

```
estimdir:
  subdir: grace/aux
  prefix: EstimDirs_
  suffix: RL05
aakrep:
  subdir: grace/YEAR/MONTH/conv_orb_1
  prefix: grc_aak_orb
  infix: _20YEAR-MONTH-DAY_RL61_GPSRL62_RL0521eo.JOBID
  suffix: .rep
```

Simple example

```
$ gracetree.rb -x ls -t aakrep -m 10 -d 03
```

```
/corral-tacc/utexas/csr/grace/02/10/conv_orb_1/grc_aak_orb_2002-10-03_RL61_GPSRL62_RL0521eo.1498260194.rep  
/corral-tacc/utexas/csr/grace/03/10/conv_orb_1/grc_aak_orb_2003-10-03_RL61_GPSRL62_RL0521eo.709991244.rep  
/corral-tacc/utexas/csr/grace/04/10/conv_orb_1/grc_aak_orb_2004-10-03_RL61_GPSRL62_RL0521eo.528365271.rep  
/corral-tacc/utexas/csr/grace/05/10/conv_orb_1/grc_aak_orb_2005-10-03_RL61_GPSRL62_RL0521eo.507643286.rep  
/corral-tacc/utexas/csr/grace/06/10/conv_orb_1/grc_aak_orb_2006-10-03_RL61_GPSRL62_RL0521eo.507647283.rep  
/corral-tacc/utexas/csr/grace/07/10/conv_orb_1/grc_aak_orb_2007-10-03_RL61_GPSRL62_RL0521eo.507650279.rep  
/corral-tacc/utexas/csr/grace/07/10/conv_orb_1/grc_aak_orb_2007-10-03_RL61_GPSRL62_RL0521eo.507650370.rep  
/corral-tacc/utexas/csr/grace/08/10/conv_orb_1/grc_aak_orb_2008-10-03_RL61_GPSRL62_RL0521eo.507654283.rep  
/corral-tacc/utexas/csr/grace/09/10/conv_orb_1/grc_aak_orb_2009-10-03_RL61_GPSRL62_RL0521eo.507657282.rep  
/corral-tacc/utexas/csr/grace/10/10/conv_orb_1/grc_aak_orb_2010-10-03_RL61_GPSRL62_RL0521eo.507660279.rep  
/corral-tacc/utexas/csr/grace/11/10/conv_orb_1/grc_aak_orb_2011-10-03_RL61_GPSRL62_RL0521eo.614853101.rep  
/corral-tacc/utexas/csr/grace/13/10/conv_orb_1/grc_aak_orb_2013-10-03_RL61_GPSRL62_RL0521eo.1660223003.rep  
/corral-tacc/utexas/csr/grace/14/10/conv_orb_1/grc_aak_orb_2014-10-03_RL61_GPSRL62_RL0521eo.2307419003.rep
```

Operations

`arc,day,jobid,month,sat,year` : retrieves parts of the filename;

`find,findstr,ls,lsstr` : `ls` or `find` (or show the resp. command strings);

`grep,grepstr,awk,awkstr` : run `grep/awk` (or show the resp. command strings) on the file list (`--pattern PATTERN` mandatory)

Retrieve filename parts

```
$ gracetree.rb -x jobid -t aakrep -m 10 -d 3
```

```
1498260194  
709991244  
528365271  
507643286  
507647283  
507650279  
507650370  
507654283  
507657282  
507660279  
614853101  
1660223003  
2307419003
```

grep data

```
$ gracetree.rb -x grep -t aakrep -m 10 -y 3 -g 'EPOCH1'
```

```
/scratch/00767/byaa676/gracetree/aakrep/03/10/grc_aak_orb_2003-10-01_RL61_GPSRL62_RL0521eo.709991242.rep: EPOCH1 2003. 9. 30.  
/scratch/00767/byaa676/gracetree/aakrep/03/10/grc_aak_orb_2003-10-02_RL61_GPSRL62_RL0521eo.709991243.rep: EPOCH1 2003. 10. 1.  
/scratch/00767/byaa676/gracetree/aakrep/03/10/grc_aak_orb_2003-10-03_RL61_GPSRL62_RL0521eo.709991244.rep: EPOCH1 2003. 10. 2.  
/scratch/00767/byaa676/gracetree/aakrep/03/10/grc_aak_orb_2003-10-04_RL61_GPSRL62_RL0521eo.709991245.rep: EPOCH1 2003. 10. 3.  
/scratch/00767/byaa676/gracetree/aakrep/03/10/grc_aak_orb_2003-10-05_RL61_GPSRL62_RL0521eo.709991246.rep: EPOCH1 2003. 10. 4.  
/scratch/00767/byaa676/gracetree/aakrep/03/10/grc_aak_orb_2003-10-06_RL61_GPSRL62_RL0521eo.709991247.rep: EPOCH1 2003. 10. 5.  
/scratch/00767/byaa676/gracetree/aakrep/03/10/grc_aak_orb_2003-10-07_RL61_GPSRL62_RL0521eo.709991248.rep: EPOCH1 2003. 10. 6.  
...  
/scratch/00767/byaa676/gracetree/aakrep/03/10/grc_aak_orb_2003-10-25_RL61_GPSRL62_RL0521eo.709991266.rep: EPOCH1 2003. 10. 24.  
/scratch/00767/byaa676/gracetree/aakrep/03/10/grc_aak_orb_2003-10-26_RL61_GPSRL62_RL0521eo.709991267.rep: EPOCH1 2003. 10. 25.  
/scratch/00767/byaa676/gracetree/aakrep/03/10/grc_aak_orb_2003-10-27_RL61_GPSRL62_RL0521eo.709991268.rep: EPOCH1 2003. 10. 26.  
/scratch/00767/byaa676/gracetree/aakrep/03/10/grc_aak_orb_2003-10-28_RL61_GPSRL62_RL0521eo.709991269.rep: EPOCH1 2003. 10. 27.  
/scratch/00767/byaa676/gracetree/aakrep/03/10/grc_aak_orb_2003-10-29_RL61_GPSRL62_RL0521eo.709991270.rep: EPOCH1 2003. 10. 28.  
/scratch/00767/byaa676/gracetree/aakrep/03/10/grc_aak_orb_2003-10-30_RL61_GPSRL62_RL0521eo.709991271.rep: EPOCH1 2003. 10. 29.  
/scratch/00767/byaa676/gracetree/aakrep/03/10/grc_aak_orb_2003-10-31_RL61_GPSRL62_RL0521eo.709991272.rep: EPOCH1 2003. 10. 30.
```

Note: 2003/10/* is used instead of */10/3 because a grep/awk operation requires the files to be copied to the --data directory.

grep cleaned data

```
$ gracetree.rb -x grep -t aakrep -m 10 -y 3 -g 'EPOCH1' -c
```

```
2003. 9. 30.  
2003. 10. 1.  
2003. 10. 2.  
2003. 10. 3.  
2003. 10. 4.  
2003. 10. 5.  
2003. 10. 6.  
...  
2003. 10. 24.  
2003. 10. 25.  
2003. 10. 26.  
2003. 10. 27.  
2003. 10. 28.  
2003. 10. 29.  
2003. 10. 30.
```

<released_solution>

```
released_solution:
  subdir: byaa705/grace/grav/RL05_YEAR-MONTH/[a-z]/iter
estim:
  subdir: <released_solution>/solution
  prefix: R.common.
  infix: arcARC.to.arcARC.JOBID
  suffix: .estim
```

Note: the <released_solution> requires dedicated coding and cannot be defined in the parameter file (as well as the estimdir file type).

Simple example (2003/10)

```
$ gracetree.rb -x ls -t estim -m 10 -y 03
```

```
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-10/a/iter/solution/R.common.arc1.to.arc1.717408.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-10/a/iter/solution/R.common.arc2.to.arc2.717408.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-10/a/iter/solution/R.common.arc3.to.arc3.717408.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-10/a/iter/solution/R.common.arc4.to.arc4.717408.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-10/a/iter/solution/R.common.arc5.to.arc5.717408.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-10/a/iter/solution/R.common.arc6.to.arc6.717408.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-10/a/iter/solution/R.common.arc7.to.arc7.717408.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-10/a/iter/solution/R.common.arc8.to.arc8.717408.estim  
...  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-10/a/iter/solution/R.common.arc24.to.arc24.717408.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-10/a/iter/solution/R.common.arc25.to.arc25.717408.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-10/a/iter/solution/R.common.arc26.to.arc26.717408.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-10/a/iter/solution/R.common.arc27.to.arc27.717408.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-10/a/iter/solution/R.common.arc28.to.arc28.717408.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-10/a/iter/solution/R.common.arc29.to.arc29.717408.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-10/a/iter/solution/R.common.arc30.to.arc30.717408.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-10/a/iter/solution/R.common.arc31.to.arc31.717408.estim
```

Simple example (2003/12)

```
$ gracetree.rb -x ls -t estim -m 12 -y 03
```

```
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-12/b/iter/solution/R.common.arc1.to.arc1.728281.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-12/b/iter/solution/R.common.arc2.to.arc2.728281.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-12/b/iter/solution/R.common.arc3.to.arc3.728281.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-12/b/iter/solution/R.common.arc4.to.arc4.728281.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-12/b/iter/solution/R.common.arc5.to.arc5.728281.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-12/b/iter/solution/R.common.arc6.to.arc6.728281.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-12/b/iter/solution/R.common.arc7.to.arc7.728281.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-12/b/iter/solution/R.common.arc8.to.arc8.728281.estim  
...  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-12/b/iter/solution/R.common.arc25.to.arc25.728281.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-12/b/iter/solution/R.common.arc26.to.arc26.728281.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-12/b/iter/solution/R.common.arc27.to.arc27.728281.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-12/b/iter/solution/R.common.arc28.to.arc28.728281.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-12/b/iter/solution/R.common.arc29.to.arc29.728281.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-12/b/iter/solution/R.common.arc30.to.arc30.728281.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-12/b/iter/solution/R.common.arc31.to.arc31.728281.estim  
/corral-tacc/utexas/csr/byaa705/grace/grav/RL05_03-12/b/iter/solution/R.common.arc32.to.arc32.728281.estim
```

JOBID linking

```
input:
  subdir: <released_solution>/arc_R
  prefix: INPUT_
  infix: YEAR-MONTH-DAY_RL61_62_RL05
  suffix: .JOBID.*
```

```
jtk:
  subdir: grace/YEAR/monitor_1
  prefix: grcgrv_jtk
  infix: _20YEAR-MONTH-DAY-
  suffix: RL05.JOBID
  jobid_from: input
```

```
$ gracetree.rb -x ls -t jtk -m 4 -y 02 | wc -l
```

29

```
$ ls /corral-tacc/utexas/csr/grace/02/monitor_1/grcgrv_jtk_2002-04-* | wc -l
```

31

Defining new methods

```
jtkstart:      -t jtk      -x awk -g "{print $12}"
jtkduration:  -t jtk      -x awk -g "{print $13}"
jtkmjd:       -t jtk      -x awk -g "{print $4}"
jtkdate:      -t jtk      -x awk -g "{print $3}"
arclist:      -t estim     -x arc
arcdate:      -x jtkdate+jtkmjd+arclist
inputdate:    -t input     -x year+month+day
```

Using new methods

```
$ gracetree.rb -x arcdate -t jtk -m 4 -y 02
```

```
04/05/02 52369 1
```

```
04/05/02 52369 2
```

```
04/06/02 52370 3
```

```
04/07/02 52371 4
```

```
04/08/02 52372 5
```

```
04/08/02 52372 6
```

```
...
```

```
04/24/02 52388 24
```

```
04/25/02 52389 25
```

```
04/26/02 52390 26
```

```
04/27/02 52391 27
```

```
04/29/02 52393 28
```

```
04/30/02 52394 29
```

Putting it all together

Filetype definition:

grA.ACOX.estim: -t estim -x grep -g "GRC-A ACOX"

```
$ gracetree.rb -x inputdate+arcdate+jtkstart+jtkduration+jobid+grA.ACOX.estim -t estim -m 4 -y 02 -c | column -t
```

2002	04	05	04/05/02	52369	1	0.0	26400.0	1593715	3.774424464092000e-08	-3.585594302740665e-09	3.415865033817934e-08
2002	04	05	04/05/02	52369	2	54600.0	31800.0	1593715	2.881504941231000e-08	-2.675618380916963e-09	2.613943103139303e-08
2002	04	06	04/06/02	52370	3	0.0	86400.0	1593715	2.825316559399000e-08	-1.476094818629190e-09	2.677707077536081e-08
2002	04	07	04/07/02	52371	4	0.0	86400.0	1593715	2.859743466790000e-08	-1.899624227063042e-09	2.669781044083696e-08
2002	04	08	04/08/02	52372	5	0.0	33100.0	1593715	2.300455016988000e-08	-3.530292497868363e-09	1.947425767201164e-08
2002	04	08	04/08/02	52372	6	57700.0	28700.0	1593715	2.738168721344000e-08	-3.277743667500253e-09	2.410394354593975e-08
...											
2002	04	25	04/25/02	52389	25	0.0	53400.0	1593715	2.360124654399000e-08	-2.018281915787644e-09	2.158296462820236e-08
2002	04	26	04/26/02	52390	26	65000.0	21400.0	1593715	1.397107142792000e-08	-2.558863754810568e-09	1.141220767310943e-08
2002	04	27	04/27/02	52391	27	0.0	12000.0	1593715	-1.622743315007000e-08	4.459044962572409e-09	-1.176838818749759e-08
2002	04	29	04/29/02	52393	28	28800.0	57600.0	1593715	2.358236692048000e-08	-1.687372604854361e-09	2.189499431562564e-08
2002	04	30	04/30/02	52394	29	0.0	85400.0	1593715	2.266895844073000e-08	-1.761351335704071e-09	2.090760710502593e-08