

# Unify data storage for raw data

## Data outline

The PingER measurements are recorded by the Measurement Agent (MA) into a directory: `/usr/local/share/pinger/data` (or in the unique case of SLAC into `/nfs/slac/g/net/pinger/pinger2/data/`). The file names are by month, i.e. `ping-<YYYY>-<MM>.txt` (e.g. `/usr/local/share/pinger/data/ping-2011-02.txt`). See [PingER data flow at SLAC](#) for more details

When the data is gathered to SLAC from the MAs it is compressed and placed in files of the form:

```
/nfs/slac/g/net/pinger/pingerdata/hep/data/<host>/ping-<YYYY>-<MM>-<DD>.txt.gz
e.g.
/nfs/slac/g/net/pinger/pingerdata/hep/data/pcgiga.cern.ch/ping-2006-09-28.txt.gz
```

The format of the data in both the MA and gathered files can be found in [PingER Monitor node format](#).

To simplify accessing the raw data, we need to have a copy of the gathered data by month, i.e.

```
/nfs/slac/g/net/pinger/pingerdata/hep/data/<host>/ping-<YYYY>-<MM>.txt.gz, e.g. /nfs/slac/g/net/pinger/pingerdata/hep/data/pcgiga.cern.ch/ping-2006-09.txt.gz
```

We might be able to get away without compression. It will simplify the analysis. We will need to experiment to see how much data is created. Each year's worth of data compressed for `pinger.slac.stanford.edu` takes ~ 800MBytes and I am guessing about 4Gbytes uncompressed (guessing compression ratio ~ 5, this needs verifying). For `pinger.cern.ch` (a more typical MA) we see ~ 50MBytes/year compressed or 200MBytes uncompressed. There are ~ 100 MAs or say 20Gbytes. So in total we may need ~25Gbytes for the uncompressed data.

## Task

The task is to take the MA data for each MA for each day of a month and aggregate it into a single file per MA per month. Initially don't bother with compression. Eventually this will run as a daily cronjob, thus it must be robust and if run from a cronjob only produce any stdout if there is a problem.

## Hints

Write in perl. Start from `~cottrell/bin/template.pl` or `~cottrell/sumdir-regex.pl`. Use the perl `opendir` function to get the directory listing. Use `/bin/gunzip` or `bin/zcat` to uncompress (unzip) the files. Use `stat` if you need file size etc. Look at `File::Copy` for copying files. You will need to use the append feature of the perl open with `>>`.

The permissions for the directory are:

```
109cottrell@pinger:~$ls -ld /nfs/slac/g/net/pinger/pingerdata/hep/data/
drwxrwsr-x 210 iepm iepm 6144 Jun 10 10:43 /nfs/slac/g/net/pinger/pingerdata/hep/data/
```

## Script

The script to unify this data is [/afs/slac/package/pinger/unite-monthly.pl](#).

## Storage space for Unified data

Renata created `/pingerdata.archive` on `netfs03` with about 600GB of space. The directory is `/nfs/slac/g/net/pinger/pingerdata/hep/data.unite/`.

The space is available via the automounter. It is owned by user `id pinger` under group `sf`.