Meta Data Collection

Background

Step specific Meta data can be used to estimate run time for future simulations.
Job Meta Data is provide us an overview of the life-cycle of each version.

Currently Meta Data for both sim steps and jobs are scattered, a more centralized and easy to access of storing is preferred.

Story / Deliverable

After a batch of simulation finished, we should store all the useful meta data in a more accessible fashion an visualize them.

Tasks

1) script to collect all sim_step meta data (1d) - Done
2) script to collect all jobs meta data (1d) - Done
3) quantify and plot sim_step meta data (3h)
4) quantify and plot jobs meta data (3h)

Job Meta Data

Params to Collect:

params	Location	Note
Slurm Job ID		to make it unique, may need to be combined with submission time.
RunGroupName		
Submission time		
Start time (of the first task)		
End time (of the last task)		
Wall clock		
Number of cores requested		
Number of nodes requested		
Memory requested		
Partition		
Machine		
System load		Availability unknown

Task Meta Data

Params to Collect

params	LF step	HF step	BB step	Location	Note
Slurm Job ID					
Task ID					Step.Realisation.Timestamp
Run time (in hours)				sim_dir/ch_log	
Cores used				sim_dir/ch_log	
Memory used per Core				sim_dir/LF/srf/Rlog/*.rlog **see note 3 & 4**.	

Nx				sim_dir/ch_log, params_base.py	
Ny				sim_dir/ch_log, params_base.py	
Nz				sim_dir/ch_log, params_base.py	
hh				params_base.py	
nt				sim_dir/ch_log, params.py	
dt				params_base.py	
nsub_stoch				sim_dir/ch_log	
fd_count (station number)				sim_dir/ch_log	
start, end, submission time				???	

Notes:

1. Columns in green at the parameters that is currently known that is affecting Run time.
2. For any parameters that can be retrieved from ch_log, they can be collected by other means; see into proc.sl.template for reference.
3. The Rlog stores the "estimated" memory usage not the actual used.
4. Rlogs for LF are stored in ASCII format. everything should be human readable. the work load of each core is recorded at the start of the log. (size, dt, nt, hh, mem)
5. The location of 'start, end, submission time' is currently unknown and will be collected later.

Data-Fetching

JSON FILE

The JSON file format is used to store the collected metadata.

The python script that writes these JSON files is: https://github.com/ucgmsim/slurm_gm_workflow/blob/master/write_jsons.py

```
usage: write_jsons.py [-h] [-sj] [-sf] run_folder

positional arguments:
  run_folder           path to cybershake run_folder eg'/nesi/nobackup/nesi002
                       13/RunFolder/Cybershake/v18p6_batched/v18p6_exclude_1k_
                       batch_2/Runs/' or '/nesi/nobackup/nesi00213/RunFolder/C
                       ybershake/v18p6_batched/v18p6_exclude_1k_batch_2/Runs/H
                       ollyford'

optional arguments:
  -h, --help           show this help message and exit
  -sj, --single_json   Please add '-sj' to indicate that you only want to
                       output one single_json json file that contains all
                       realizations. Default output one json file for each
                       realization
  -sf, --single_fault  Please add '-sf' to indicate that run_folder path
                       points to a single fault eg, add '-sf' if run_folder is
                       '/nesi/nobackup/nesi00213/RunFolder/Cybershake/v18p6_ba
                       tched/v18p6_exclude_1k_batch_2/Runs/Hollyford'
```

Sample command:

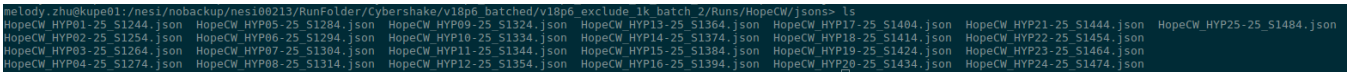
```
# Input path to Runs
$ python write_jsons.py /nesi/nobackup/nesi00213/RunFolder/Cybershake/v18p6_batched/v18p6_exclude_1k_batch_2
/Runs/

# Input path to a single fault, needs '-sf' option
$ python write_jsons.py /nesi/nobackup/nesi00213/RunFolder/Cybershake/v18p6_batched/v18p6_exclude_1k_batch_2
/Runs/HopeCW -sf

# Output a single json file for all realizations, needs '-sj' option
$ python write_jsons.py /nesi/nobackup/nesi00213/RunFolder/Cybershake/v18p6_batched/v18p6_exclude_1k_batch_2
/Runs/HopeCW -sf -sj
```

Sample output:

```
# output json files are located in fault_dir/jsons
$ cd /nesi/nobackup/nesi00213/RunFolder/Cybershake/v18p6_batched/v18p6_exclude_1k_batch_2/Runs/HopeCW/jsons
$ ls
```




```
$ cat HopeCW_HYP25-25_S1484.json
```


```
{
  "LF":
    {
      "nx": "539",
      "ny": "621",
      "nz": "110",
      "run_time": "0.048 hour,
      "cores": "160",
      "nt": "5159",
      "total_memo_usage": "4.8 GB"
      "start_time": "2018-08-20_23:21:54",
      "end_time": "2018-08-20_23:24:45"
    },
  "HF":
    {
      "fd_count": "4164",
      "nsub_stoch": "144",
      "run_time": "0.056 hour",
      "cores": "80",
      "nt": "20636",
      "start_time": "2018-08-20_23:21:54",
      "end_time": "2018-08-20_23:25:15"
      },
  "common":
    {
      "hh": "0.4"
    },
  "BB":
    {
      "cores": "80",
      "fd_count": "4164",
      "dt": "0.005",
      "run_time": "0.015 hour"
      "start_time": "2018-08-20_23:21:55",
      "end_time": "2018-08-20_23:22:49"
    }
}
```

Plots of meta data

run_time