Hi. Followig on from a videoteleconference a few days ago, find here a few
monitoring databases that Tom has dug up from his runs for our main
workflow. This workflow is being replaced by a same-but-different version
that probably has similar characteristics as far as individual tasks go.

<http://www.hawaga.org.uk/tmp/wqdb/>

Two are large production runs from earlier in the year, and one is a more
recent much smaller run. The large runs were made with parsl's
HighThroughputExecutor, not with workqueue.

These databases are made with varying verions of parsl, so the schemas may
not align.

They are sqlite3 files.

For the biggest one, dr3-monitoring.db:

$ sqlite3 dr3-monitoring.db

There are 29594 tasks to run:

sqlite> select count(\*) from task;
29594

and with retries, that was 57305 attempts at executing a task (essentially,
workqueue would see that many distinct workqueue-level task submissions)

sqlite> select count(\*) from try;
57305

There's a lot of dependency structure which is managed by parsl and the
workflow script above it, so there is no burst of submitting all of these at
once.

The most interesting table for you to look at to characterise eg run
durations is probably joinin those two tables together, eg - this is all of
the tasks that actually started running (rather than being known to parsl
but for one reason or another never actually running):

sqlite> select task\_func\_name, task\_try\_time\_running,
sqlite> task\_try\_time\_returned
from try, task where try.task\_id = task.task\_id and try.run\_Id = task.run\_id
and task\_try\_time\_running is not null;

task\_func\_name is the name of the specific task type (which translates to a
specific command line to execute) - it summarises as:

sqlite> select task\_func\_name, count(task\_func\_name) from try, task
sqlite> where
try.task\_id = task.task\_id and try.run\_Id = task.run\_id and
task\_try\_time\_running is not null group by task\_func\_name;

assemble\_coadd|246
check\_ccd\_astrometry|350
deblend\_coadd\_sources|456
detect\_coadd\_sources|243
forced\_phot\_coadd|882
make\_coadd\_temp\_exp|40799
make\_patch\_list\_for\_tract|10
make\_sky\_map|1
make\_tract\_list|2
make\_visit\_file|1
measure\_coadd\_sources|456
merge\_coadd\_detections|76
merge\_coadd\_measurements|76
raft\_list\_for\_visit|368
single\_frame\_driver|2042
sky\_correction|1903
tract2visit\_mapper|352
visits\_for\_tract\_patch\_filter|960

Tom has some made interesting histograms in the past of task duration.