

Metadata Refactor Plan

Phase 1

- Update the job-id updating so only latest / current job is affected by changes to the job-id
- Add a state to the db for tasks that reached WCT - name TBD
 - This will require changes to the DB structure / status enum
 - This is a pre-failed step but also considered a failed state
 - Hence tasks can be created with the same name / step with an entry in the WCT state
- Update logic to account for failed state rather than WCT hit
- Missing from queue jobs need to be differentiated between WCT fails and other fails

Phase 2

- Add a separate table for duration logging for jobs
 - This should include job_id, queued_time, start_time, end_time, nodes, cores, memory, WCT
 - queued_time, start_time and end_time are populated from the time that the item is added to the MGMTDB queue
 - Missing from queue end_time should be populated from sacct
- Add a metadata collection script
 - Aggregates the time/resources from the DB
 - And grabs relevant associated data from the params file (nx, nt etc)
 - Writes the output to a CSV file
 - Option to specify the total CH used and/or excluded prior failed runs (Getting useful CH used)
 - Create a dataframe and write into a csv file

Rel_name	LF runtime	LF queue time	N_resubmits	LF Cores	LF NX	LF Core Hour	HF.....	Total CH Used
ABC_REL01	1:23			160		(Summation across tasks)		

If multiple data exists, put a range in e.g. 160-240

Phase 3

- Cybershake Progress script update to get CH from DB instead of json files

Notes:

- When calculating total CH used excluding fails
 - Any run with WCT state should be counted too - unless there is a failed task after that - this assumes that any task in the failed state resets the working environment