

The Computing perspective

Whats work is done by LSST on GridPP DIRAC?

- ➊ (input) Copy seeded simulations of night sky from NERSC
- ➋ (input) Currently workflow bottleneck is in copying/registering ≈ 6 Million files
- ➌ (input) Input per job is small and transferred to the compute by jobs.
- ➍ (compute) 1 core simulates one sensor of telescope response
- ➎ (compute) Simulation of sensor response takes ≈ 3 hr
- ➏ (output) Potentially need all output files to be available for post-processing to be possible

Some Numbers

What resources are LSST using?

- Main bottleneck is working how to get input data (already generated) into UK and registered in GridPP-DIRAC.
- Running an average of 4k jobs/day, each takes 3hr on 4CPU & 8Gb RAM.
- Simulating 1yr of telescope response takes 60days on GridPP.
Aiming to get this down to 20 through software improvements during future challenges.
- Each job produces 1 tar file output with multiple simulated results
- Total output of LSST UK work would be $\approx 50T$ of data across $\approx 2M$ files.

Some Observations

Some observations of the LSST workflow:

- DIRAC API/CLI-tools meets all of their needs. Sometimes not immediately obvious to non-HEP people how to use more advanced features or how to discover them.
- Managing permissions/voms details required a bit of effort to understand/setup, but works.
- Experts always seem happy to help on mailing lists. This has been extremely useful.
- Job management/re-submission almost completely done manually through DIRAC web interface.
- Tracking large numbers of production jobs and output is currently done via text files.