Comments

- The DES depends critically on the successful conversion of data-bits to science.
- The overall project data task comprises
  a) The community DECam pipeline
  b) The DES data simulation, reduction (algorithms & implementations) and management (data flow & job control)
  c) The DES science post-processing

  The first and last of these are still largely undefined; the rest of our comments therefore focus on (b).

- Robust science-driven requirements on the data simulation, reduction and management components have not been developed (cf. the hardware).
- The NCSA team has concentrated on the mechanics of the data management rather than the algorithmic core of the data reduction.
- There does not seem to be complete list of required data reduction components and their performance.
- Critical people/parts of the data reduction (e.g. Bertin; Jain, Jarvis, et al.) are not under the control of the PL.
- Some core data reduction steps (e.g. the detection of transients needed for SNe) are omitted from the data reduction.

Comments

- Some of the necessary algorithms (image subtraction; galaxy photometry?) may still require research, not just coding.
- The simulations are currently mostly useful for demonstrating throughput and interfaces; they will need more accurately to replicate data as it will arrive from the DECam to be useful for scientific validation.
- The anticipated data volume is moderately large today (but will not be by 2015), and the computational complexity of the current plan appears to be very modest; as currently described, the data reduction is not a major computational challenge.
- A framework based on distributed data over grid-enabled HPC resources in multiple locations is not *scientifically* necessary, and involves positive and negative trade-offs with alternative options (e.g. working at a single center, running a dedicated cluster).
- The chosen framework is susceptible to the next decade of evolution of grid-HPC hard- and middle-ware that is outside of the control of DES, as well as the (possibly divergent) choices made in the various centers.
- The chosen framework excludes non-grid computing resources at both end of the computing spectrum.
- The use of HPC resources depends on their allocation, typically on a year-by-year basis.

Recommendations

1. Define the specification of the community pipeline.
2. Define the scope of, and requirements on, the science-ready data products.
3. Use these to prepare a requirements document for the simulations, data reduction and data management.
4. Complete a full WBS, including requirements and responsibilities, to include all parts of the pipeline needed to generate science-ready DES data products.
5. Based on this develop a more complete set of milestones augmenting the current series of data challenges.
6. Strengthen the astronomy/image processing group responsible for the core data reduction.
7. Include the data reduction in the end-to-end error budget.
8. Develop a characterization of the computational complexity of each the data reduction tasks (including its scaling, pre-factor, efficiency on base architecture(s) and memory footprint).
9. Identify potential consequences of the chosen framework, and develop plans to minimize possible negative impacts.