LSST Dark Energy Science Collaboration—Data Challenge 2

Post-mortem of IRIS Usage

Background

- DESC is one of eight official LSST science collaborations
- Currently, is largest and most mature w/ 800 members worldwide
- Half of UK PIs members of DESC (equiv. 10% of membership)
- UK astronomers hold various leadership positions in DESC
  - Plus LSST:UK contributes effort (James Perry, Data Wrangler; Joe Zuntz, Pipeline Scientist)
- DESC funded by Dept of Energy in US, including computing time at NERSC (plus time at ANL and SLAC)
  - Also need contributions from international partners – e.g. in UK and France

Motivation

- DESC getting ready for cosmology with LSST
  - LSST will provide high-quality catalogues and images, which DESC can use to compute cosmological parameters, and thus quantify make-up of Dark Energy
- DESC plan to reprocess significant portion of LSST images
  - To measure systematic effects
  - Running LSST software stack
- DESC ramping up capabilities, through series of Data Challenges
- At time of writing, mid-way through Data Challenge 2
  - Which should reach scale of LSST Commissioning (in 2020—2021)

Strategy

- DC2 Run 1.2 (image generation) was scheduled for 2018 Q2 and Q3
  - Two version using competing simulators: Run 1.2i using ImSim; Run 1.2p using PhoSim
  - Plan to run R1.2p at NERSC (on Cori) and R1.2i in IRIS (on GridPP)
- Workflow
  - Input: Observing schedule for telescope; and instance catalogues, seeded with cosmological objects—low compute time, prepared in advance
  - Simulation: ImSim/ PhoSim simulate images produced by camera, from ‘visits’ identified in schedule, of sky map described in instance catalogue
  - Output: Set of images to be fed into LSST software stack for correction and catalogue generation
- Run 1.2i involved 2,001 ‘visits’, equated to 100,000 jobs (though, with newer version of ImSim, could probably reduce to 50,000 jobs)

Strategy, cont’d

- ImSim updated by DESC for Run 1.2i
  - Support for multi-threading, parallelised across sensors in camera
  - Code not stable at beginning of period, due to limitations of testing
- James Perry assigned at 0.5 FTE to set up and complete R1.2i
- GridPP agreed to support work w/ up to 20M core hours
  - Mailing list set up with reps from GridPP and LSST:UK teams
- Outputs from simulations to be transferred back to NERSC for collation with Run 1.2p, and processing

Process and outcome

- LSST VO (at FermiLab) utilised for AuthZ
  - Issues from previous LSST:UK activity resolved
- Used CERN VM client for access to GridPP
  - Issues signing up, bug in web form (CAPTCHA) and unsuitable password being issued
  - GridPP team helped work around issues
- Used Ganga job manager, as recommended by GridPP team
  - Initial testing went well
- Used CVMFS for software deployment
  - GridPP nodes not able to see LSST repository in IN2P3, unresolved
  - Work-around to set up new copy of software on GridPP repository
  - Several updates to ImSim to correct bugs => several rebuilds of stack
  - Reoccurrence of latency issue, with software taking several hours to propagate across grid

Process and Outcomes, cont’d

- ImSim tested on GridPP, using Ganga as workflow manager
  - Initially jobs were killed by system, due to insufficient memory (ImSim run typically needs ~11GB)
  - Therefore, requested 8 cores per job (gives 8×2GB memory)
  - Failed, as LSST VO not authorised for multicore jobs
  - Led to poor efficiency reporting, as jobs only used subset of cores (though all of memory)
  - Instance catalogues uploaded to grid storage using Dirac tool
    - However, bug in Ganga constrained use of sites, based on location of input data
    - Attempt by GridPP team to fix bug did not work
    - Eventually, worked around be replicating same input data to multiple sites
    - N.B Ganga bug has now been fixed

Process and outcomes, cont’d

- Given large number of jobs (~100,000)
  - LSST:UK produced automated scripts for submission, clean-up and re-submission of failed jobs (these are on GitHub, if people are interested)
  - Two iterations produced reasonable process, though manual intervention required to address different failure modes
    - Some modes could be recovered from automatically, though others could not
  - One failure mode related to ‘visits’ that (correctly didn’t produce outputs)
    - Worked around by creating dummy output set
  - Small number of jobs required more than 48 hours to complete
    - So timed out

Process and outcomes, cont’d

- With various iterations of process improvement, able to reach 3,000 jobs per day
- Ganga became limiting factor at this point:
  - Monitoring seems unreliable
  - Responsiveness of UI degrades significantly at O(1,000) jobs
- By September, we had only completed ~25,000 jobs on GridPP
  - Run 1.2i moved onto critical path for DC2
  - DESC team in US were offered substantial time on Theta (large, batch-system based cluster at ANL), which was used to complete runs in just a few days

Conclusions

- GridPP staff went far above and beyond what we could reasonably ask. A big thank-you from LSST:UK for their efforts
- Disappointing that Run 1.2i likely to take ~1 month on GridPP vs. 2—3 days on US system
  - Perhaps Dirac tools would have been better choice than Ganga?
- Because use of GridPP by non-LHC activities is modest, each new activity is likely to require significant effort and expertise to get going
  - New use cases expose weaknesses/ bugs in GridPP infrastructure
  - Grid computing is not a widely held skill in astronomy (whereas batch computing is)

Conclusions, cont’d and Further Work

- Unreasonable to expect activities to make effective use of IRIS without right expertise
  - This is LSST:UK’s third activity on GridPP, but start-up time does not seem to reduce
- IRIS has potential to make big impact on large-scale campaigns such as LSST DESC data challenges
  - though most of features of grid are not required by non-LHC activities?
  - Would be better to provide batch computing instead?

Future Work

- DESC is moving on to DC2 Run 2i/p
  - LSST:UK are prep'ing ImSim for this (issue with scaling for larger instance catalogues with faint objects)
  - IRIS potentially good candidate for some of compute load
- LSST:UK also interested in other options for IRIS
  - Hosting finalised catalogue and image set for DESC DC2, from IN2P3
  - Useful data transfer experiment?
- DESC DC3 scheduled for 2020
  - Massive task—scale of LSST Science Validation

Thank-you again to GridPP team: this presentation is in no way of criticism of them and their commitment to help us