Computing & Operations Summaries

B. Glendenning & M. McKinnon

2nd ngVLA Technical Workshop, Socorro, NM

December 8-9, 2015

NRAO NSF Associated Universities Inc

Computing Architecture Summary

Computing Summary (1)

- ngVLA is a PI observatory, which implies a lot of things about projects, time allocations, data proprietary periods, etc.
- To support efficient operations, M&C systems must be highly reliable/resettable and have good support of remote diagnostics (easier to say than do)
- ngVLA can be a traditional “saves the visibilities” radio telescope
  - Costs very dependent on assumptions (most critically: Moore’s law continues to be in force with 2 year halving timescale)
  - Assume any high-time resolution work is handled with real-time processing
  - Dishes not too small (18m for a 2030 start date)

Computing Summary (2)

- But can we save the resulting pixels?
  - Create a canonical science program and derive parameters like image sizes, required dynamic range/fidelity from it (DRSP in ALMA speak)?
  - Archival policy issues, e.g. do we (e.g.) image the PB for archival research even if the original science does not need it
- The “areas of work” for algorithm R&D is understood, and the current telescopes are good testbeds
  - But there’s a lot of work to be done
  - And some specifications need to flow down from science (e.g., required dynamic range of ngVLA observations (really lower than JVLA?), fractional BW for continuum observations)
- The scale/affordability of computing is not well understood
  - But it won’t be small in any scenario, the question is how scary it will be
  - Parallelization and robust pipelines needed in all circumstances

Computing Summary (3)

- The data management/computing breakdown between what is done by the Observatory vs. the community is an important one
  - Adds complexity, potentially brings resources ($ and facilities)
  - Real world probably requires more data moves than would be ideal (good network connections!)
  - Software portability (containerization?) important
  - Where will researchers do data reduction peculiar for their project
- Large observatories with large construction software staffs require a lot of management, HR work
  - Team culture in a distributed environment requires particular attention

Computing Summary (4)

- Commerical cloud providers (e.g., Amazon AWS) are very interesting
  + No capital investment, spot market pricing, any size you need for a particular time
  - Not clear that it's cost effective (especially vs. national HPC research infrastructure), not simple to use/tune
  – But certainly convenient/effective for "bursty" work, including ngVLA algorithm R&D

Computing Summary (5)

- Science archives are a multiplier for total science output
  - But requires that the archive products are generally (re)usable
- ngVLA calibration will (mostly) not be a computing problem
  - But it should not be neglected, don't replicate ALMA experience with (lack of) calibration plan
  - Many options, how will we pick?
  - What to do about weather (some parts of the array good, others not)

Operations Concept Summary

Science Operations and User Support - I

- Importance of operations plan/concept of operations
- Separate funding for telescope operations & analysis/processing?
  - Distributed science operations?
- Operational model: full dynamic scheduling
  - Account for environmental conditions, etc. to improve observing efficiency. But weather won't be ideal across the entire array
- Retain community by continuing to operate VLA during ngVLA construction
- Engage community in commissioning (RSRO, secondment)
- Distinction between standard/general (demonstrated) and non-standard observations

Science Operations and User Support - II

- Data reduction pipeline available early (construction deliverable)
  - Well-defined calibration plan
- Avoid data poisoning by using pipelines and via access to post-processing cluster
- Implementation of PI-led and survey modes must be driven by scientific impact
- Access and availability of key documents (centralization)
- Define roles and data access control early
- Ownership of science requirements & close work with computing

Array Operations and Maintenance - I

- Understand consequences of separate funding for construction and operations (color of money)
  - Retain construction staff into operations (skill retention)
- Get the basics correct! (infrastructure, power, communications)
- Minimize power costs
  - Green Antenna, alternative energy sources, cryogenics
- Can we afford to operate what we build?
  - Minimize antenna maintenance (visits)
    - LRUs, remote resets, self-lubrication, easy antenna access, note problem areas
- Clearly define maintenance strategy
  - Implications for staff skill level/location and design MTBF
- Importance of support staff proximity to the telescope

Array Operations and Maintenance - II

- Continuous maintenance (vs dedicated array maintenance day)
- Proper estimate of logistical support for distant sites.
  - Location of regional maintenance centers
- Antenna accessibility (avoid road maintenance)
- How to communicate in remote locations?
- Need for well-developed and agreed acceptance procedures/criteria (PA/QA)
- Avoid technical development and construction in parallel
- Landowners not necessarily antagonistic

www.nrao.edu
science.nrao.edu
public.nrao.edu

The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.