Title
Improving Data Mobility & Management for International Cosmology: Summary Report of the CrossConnects 2015 Workshop

Authors
Borrill, Julian
Dart, Eli
Gore, Brooklin
et al.

Publication Date
2015-10-02

CrossConnects
WORKSHOP SERIES

Where Science
& Technology Meet

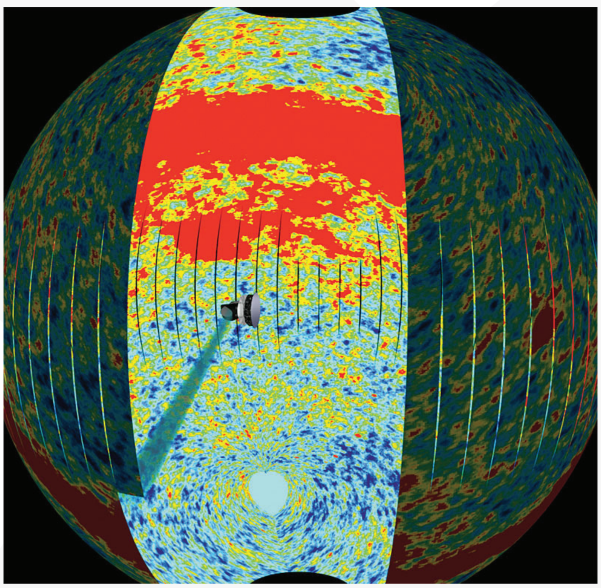
ESnet
ENERGY SCIENCES NETWORK

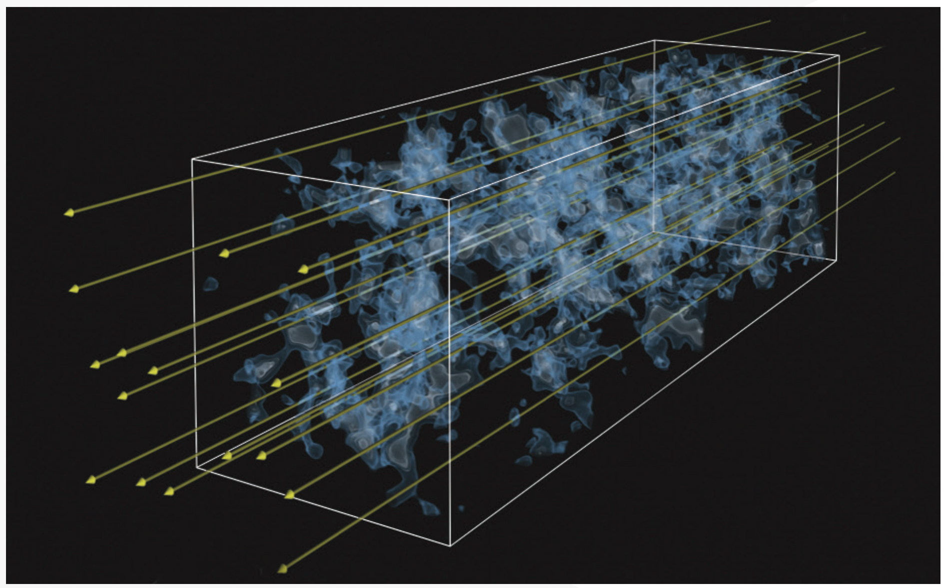
INTERNET2

Summary Report of the CrossConnects 2015 Workshop

Improving Data Mobility & Management for International Cosmology

Julian Borrill, Eli Dart, Brooklin Gore, Salman Habib, Steven T. Myers,
Peter Nugent, Don Petravick, Rollin Thomas





BERKELEY LAB

U.S. DEPARTMENT OF ENERGY
Office of Science

16-SN-2198

Summary Report of the CrossConnects 2015 Workshop on Improving Data Mobility & Management for International Cosmology

Julian Borrill[1], Eli Dart[1], Brooklin Gore[1], Salman Habib[2], Steven T. Myers[3], Peter Nugent[1], Don Petravick[4], Rollin Thomas[1]

1) Lawrence Berkeley National Laboratory, Berkeley CA.
2) Argonne National Laboratory, Argonne, IL
3) National Radio Astronomy Observatory, Socorro, NM
4) National Center for Supercomputing Applications, Urbana, IL

Acknowledgements

ESnet is funded by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research. Vince Dattoria is the ESnet Program Manager.

ESnet is operated by Lawrence Berkeley National Laboratory, which is operated by the University of California for the U.S. Department of Energy under contract DE-AC02-05CH11231. This work was supported by the Directors of the Office of Science, Office of Advanced Scientific Computing Research, Facilities Division.

This is LBNL report LBNL-1001456

Disclaimer

Executive Summary

In February 2015 the third workshop in the CrossConnects series, with a focus on *Improving Data Mobility & Management for International Cosmology,* was held at Lawrence Berkeley National Laboratory. Scientists from fields including astrophysics, cosmology, and astronomy collaborated with experts in computing and networking to outline strategic opportunities for enhancing scientific productivity and effectively managing the ever-increasing scale of scientific data. While each field has unique details which depend on the instruments employed, the type and scale of the data, and the structure of scientific collaborations, several important themes emerged from the workshop discussions.

One important theme is the need for workflows or analysis pipelines that are scalable, robust, and reliable. Gone are the days when individual humans could effectively process and manage data sets without computers, and the only way to allow scientists to continue to be productive is to use computers to manage the placement, processing, archival, and publication of the data. This has implications for the software groups that write and maintain the components of these pipelines, as well as for the observational facilities and computing centers that acquire and process the data. There is work to be done in building and enhancing the tools that enable data analysis workflows and pipelines, and the potential for significant improvement in scientific productivity if these efforts are successful.

Another key theme to emerge at the workshop is the synergy between simulation and observation. It is unlikely that a major sky survey or satellite mission will be undertaken without a corresponding simulation component. These two complementary disciplines, and therefore the facilities they use, will both be key players in all major projects going forward.

A third theme is the movement of data at scale between facilities, and the tools to effectively manage the movement of data without placing undue burdens on the scientists. Infrastructure providers (including computing facilities, networking facilities, and observational facilities) must continue to enhance their capabilities, and work together to ensure smooth interoperability in support of large-scale data movement.

Further findings, as well as a set of recommendations, are contained in their respective sections in this report.

Motivation

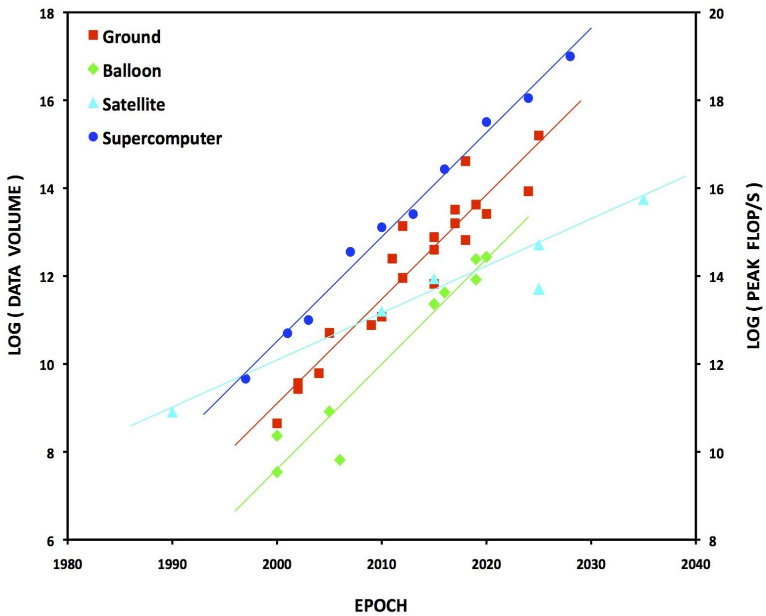
As is the case in many fields today, astronomy and cosmology are experiencing sustained, dramatic increases in data rates and volumes. The scientific promise of these data sets is unprecedented, but only if they can be effectively managed and analyzed. The details of what is needed and which solutions are best undertaken vary by area and project. This section provides background for several scientific areas represented at the workshop, and provides context for the findings and recommendations that follow.

Radio Astrophysics and Cosmology

Radio Astronomy is entering an era where new observational facilities consisting of large numbers (10^2 - 10^5) of antennas combined together as interferometric arrays are producing petabytes of data in a year. Near term and future arrays such as the Square Kilometre Array (SKA) and its precursors now under construction will present Exascale data challenges. The key science drivers for these facilities include those that will be high-impact in cosmology research, including detection and measurement of the cosmic dawn of structure formation and the Epoch of Reionization (EoR), tomography of the Universe through the measurement of weak lensing in the radio, and many others. Realization of these goals will depend on the synchronous networking of these arrays of receptors, collation and processing of data streams, imaging and and analysis in time-series of Terapixel and larger image spectral cubes, and distribution to and use by an internationally distributed network of astronomers and physicists. Computing and software at all layers in the workflow must be able to interface and talk to the others. Current state-of-the art radio arrays such as the NRAO Karl G. Jansky Very Large Array (VLA), and the Atacama Large Millimeter-submillimeter Array (ALMA) are paving the way forward from the Petascale to the Exascale domain of the SKA. There are urgent needs in automated pipeline processing, internationally distributed archives, processing, and standardized fast network transfers, all of which share common ground with other cases highlighted at this workshop.

Cosmic Microwave Background

The Cosmic Microwave Background (CMB) radiation carries the imprint of the entire history of the Universe as tiny fluctuations in its temperature and polarization on various angular scales. The quest to detect fainter and fainter signals in the CMB has required gathering larger and larger datasets, resulting in the exponential growth of CMB datasets for the last 25 years that is expected to continue for at least the next 20.



Exponential growth in the number of samples gathered by ground-based (red), balloon-borne (green) and satellite (cyan) CMB experiments, and in high performance computing capability (blue, using the theoretical peak performance of the NERSC flagship supercomputer as a proxy). Suborbital experiments perfectly track Moore's Law, while the longer lead-time for satellite missions results in their correspondingly smaller exponent

The most ambitious target on the horizon is the CMB-S4 experiment, recommended to DOE and NSF by the recent Particle Physics Project Prioritization Process (P5), which will field hundreds of thousands of detectors on multiple ground-based telescopes at multiple sites around the world for several years to gather O(10^{15}) samples - 1000 times the data volume of the current *Planck* satellite mission, and sufficient to extract the faintest CMB signal, corresponding to the imprint in the CMB polarization of gravity waves generated during the inflationary epoch. The ensuing data challenges will range from moving the data from remote sites with limited bandwidth (including the South Pole and the Atacama Desert) to the need to reduce the entire cross-correlated dataset in one step on the largest supercomputers available in the mid-2020s.

Observation

Over the next decade observational programs from ground and space-based telescopes will increase their data taking by over an order of magnitude per experiment. PTF -> ZTF will see 5 times the areal coverage and 2 times more exposure while DES -> LSST will see 3 times the areal coverage and a factor of 4 times as many exposures per night. While transient discoveries dominate the real-time turnaround of these programs, there will be many astrophysicists who will want to look at cadences on weekly or monthly timescales with these data sets - especially after discovering a new object. Their queries will be focused on answering the question, "Did this new discovery come out of nowhere, or has it been just below our single image detection threshold for a while?" and "Do we know anything else about this part of the sky?" This will require pulling over vast quantities of data at that location in the sky, coupled with outside data from other surveys, to provide both historical information as well as data at other frequencies and resolutions. Currently in PTF these types of contextual queries produce, for just a few objects, as much I/O as the pipeline itself during a night.

Simulations

The exponentially rapid increase in the quantity and quality of data from cosmological surveys has led to remarkable scientific discoveries; given the continuing trends in extending the reach of observations, with future missions such as DESI, LSST, Euclid, WFIRST, and CMB-S4, more such results may be confidently expected. The impressive success of the current cosmological model, however, implies that new discoveries will have to be teased out as subtle features in complex datasets. For this reason, accurate simulations and robust modeling will be essential components of next-generation cosmology. The required simulations will have to incorporate as much detailed physics and knowledge of observational probes as is needed to predict the results of surveys and to carry out sophisticated scientific inference programs based on the observations. The simulations already generate datasets that are significantly larger than those from all the observations combined, and this trend shows all signs of accelerating as computer power is growing at a rate that easily outstrips our current ability to collect data from the sky. The current rate of data production -- which is limited only by systemic chokepoints and not by computational limits -- presents many challenges including in situ data reduction and analysis, extreme-scale data management, interactive large-scale analysis and data exploration, automated large-scale data transfer protocols, hierarchical storage solutions, and community access to massive datasets and analysis tools. There is an urgent need to address these problems, many of which are intimately connected to exploiting powerful networking resources.

Over the next five years, the estimated level of simulation activity will include campaigns such as the ongoing Mira-Titan Universe simulation suite, which covers ~100 large simulations run over a period of three years, to massive individual runs including hydrodynamics and a variety of subgrid models (gas cooling, star formation, astrophysical feedback mechanisms, etc.). The large individual runs will create ~10 PB of data per simulation in roughly a month of wall clock time. (This is a subset of the expected output of ~100 PB total in Level I -- raw data -- from all simulations.) Given current policies at supercomputing centers, moving this data set to the project's archival storage (a mixture of disk and tape) will have to be done on a period of roughly three months, which translates to a hard requirement of moving data at the rate of ~1 PB/week.

Additionally, there will be large “science level” data sets at Level II and Level III, which will be moved across supercomputer centers and storage sites as well as to sites with sufficient local analysis capabilities. The total amount of data at this level is estimated to be ~50 PB. This number is also large and reflects the fact that post-processing can yield many different outputs depending on the choice of modeling parameters, even for a single base simulation. Although individual data transfers here will not be as large as the one discussed above, they too can be at the ~1 PB level. Given that these datasets will be transferred essentially as part of an analysis campaign, strategies for maximizing overall throughput will be essential. These will include pre-staging of transfers, co-scheduling of analysis with the transfers, and the ability to interact with a sufficient fraction of the data set during the analysis process -- a loosely coupled workflow. Such a requirement translates to several TB/hour.

Findings and Recommendations

The discussion at the workshop covered a variety of topics of both immediate and strategic importance in astronomy and cosmology. Computing and data, the effective use of computing in astronomy and cosmology, and the relationship between data scale, computing, networking and storage were all part of the discussion, with the science drivers as the foundation. Several key findings are outlined below, as well as some recommendations for beneficial action.

Findings

1. A project to achieve routine petascale data movement between major computational facilities was discussed at the meeting. Salman Habib will drive it from the science side, and Eli Dart will drive it from the networking side. The goal would be to achieve and maintain the ability to move a one-petabyte data set between facilities in a week. ANL, BNL, NCSA, NERSC, and SDSC were discussed as initial participating facilities. Testing has begun between ANL and NERSC, and NCSA, OLCF, and SDSC will participate, to be joined at a later stage by BNL and FNAL.
2. Some of the data portals used by astronomy and cosmology projects use older tools and technologies for data transfer (e.g. FTP, HTTP, rsync), and moving large volumes of data (e.g. tens of terabytes) is difficult. An effort to modernize these data portals and augment them with high-performance data transfer tools would benefit large-scale users of the data. This will become more and more important as time goes on and data sets continue to grow in scale.
3. In addition to large scale bulk access to data, there is a need for portals to provide programmatic (e.g. via well-understood APIs) access to data. This is important for workflow systems, since different groups have a different idea of what a “workflow system” is and what it does. The key requirement is to make it easy for science groups to get access to as much data as they need, using a programming language they understand, in a way that is productive for the work they are trying to do.
4. The use of cloud storage for irreplaceable (e.g. observational) data presents some strategic challenges. As an example, one institution had difficulties getting a cloud provider to write a long-term contract for access to data stored in their cloud, including the ability to get the data back out of the provider’s cloud infrastructure in the event that the storage contract were not renewed. Custodial responsibility for irreplaceable data is

a serious matter, and outsourcing the storage for such data in a responsible way is challenging. This has strategic implications - if custodial responsibility for irreplaceable data sets is to remain with the scientific community, then there is a need to adequately fund data storage and curation efforts within the science community. As data sets increase in scale, those costs will rise.

5. It is invariably the case that data will need to be moved in order to make the data accessible in the execution context of analysis software. No matter how hard we try to move computation to the data, it is not always possible. Running simulations, moving the data, storing the data, and running the code to analyze the data all have non-negligible costs. Since surveys involve integrating simulation data sets with observational data sets (at the petabyte scale), additional analysis is needed to determine a workable method of balancing the costs involved while ensuring that the required capabilities are available to science collaborations.
6. In order to be reliably integrated into an automated workflow, data transfers need to be reliable in the sense that they must reliably provide actionable status information in a form that an automated workflow can consume. Providing this is challenging with tools which mask or hide errors. If there were a data transfer toolset that could reliably provide callbacks to a workflow manager, this would be quite valuable. Note that data transfer tools need to provide appropriate information to workflow systems which are capable of monitoring errors and service faults in a context other that the transfer tool itself, so that faults and errors can be handled in the appropriate context (and escalated as appropriate).
7. The notion of a data processing workflow or “pipeline” came up several times. It may be helpful to enumerate common design patterns for pipelines which process data, so that different projects can benefit from shared experience and common components.
8. Observational Cosmology projects are intrinsically multi-site, and often multi-agency. Cosmology projects are often embedded in projects having other stakeholder sciences (if one distinguishes cosmology from general astronomy). Facile use of networks is needed to satisfy the stakeholders, because of the shared, distributed (for example, from observation to comparing processed data with completed simulations) nature of skills needed to do the science. As projects become larger and data sets become more useful to more constituents, the ability to provide data products to scientists outside the core project which collected the original data will become increasingly important. It is likely that funding agencies will need to be involved in this discussion.

Recommendations

1. ESnet should collaborate with the major computing centers in support of the petascale data movement workflow (see finding 1).
2. Globus and the DES data management group at NCSA could benefit from exploring the suitability of Globus for use in automated workflows (see finding 6).

3. It was clear from the discussions that the effective use of computing has become critical to the success of large-scale astronomy and cosmology projects. Based on the discussions at the workshop, computer scientists should be included in the project at the beginning, and the computer scientists should learn as much as they can about the scientific goals of the project. This team science structure has been used successfully in the past, and allows scientific projects to effectively reap the benefits of state-of-the-art computing.

Workshop Background and Structure

A Cross Connects Workshop, the third in a series, was held at Lawrence Berkeley Lab, Feb 10-11, 2015 with a focus on *Improving Data Mobility & Management for International Cosmology*. The workshop brought together more than 50 members of the community in person and another 75 viewed the live stream of the event. ESnet's Greg Bell described this workshop as "one of the best yet."

The Cross Connects series brings together professionals at the intersection of scientific research, experimental facilities and cyberinfrastructure. Only by bringing these three groups together can complex, end-to-end data mobility and management challenges be discussed *and* addressed. The first two workshops in the series focused on the life sciences and climate sciences. A fourth workshop is being considered for life sciences in 2016.

Appendix

Workshop Planning Committee

- Julian Borrill, Lawrence Berkeley National Laboratory, Computational Cosmology Center
- Andrew Connolly, University of Washington, Dept. of Astronomy, LSST
- Eli Dart, Lawrence Berkeley National Laboratory, ESnet
- Susan Evett, Internet2
- Brooklin Gore, Lawrence Berkeley National Laboratory, ESnet
- Salman Habib, Argonne National Laboratory, Kavli Institute for Cosmological Physics
- John Hicks, Internet2
- Steven Myers, National Radio Astronomy Observatory
- Inder Monga, Lawrence Berkeley National Laboratory, ESnet
- Peter Nugent, Lawrence Berkeley National Laboratory, Computational Cosmology Center
- Don Petravick, University of Illinois, National Center for Supercomputing Applications, Dark Energy Survey
- Lauren Rotman, Lawrence Berkeley National Laboratory, ESnet
- Rollin Thomas, Lawrence Berkeley National Laboratory, Computational Cosmology Center