SPOT Suite: HPC for Light Sources

Craig E. Tull, PhD
Computing Research Division

ALS Users’ Meeting
Advanced Tomographic Reconstruction Algorithms
October 8, 2013 @ LBNL, Berkeley, CA
Data Intensive Science

- Data Intensive Science is becoming the norm for experiment, observation, and simulation.
- President Obama March 29, 2012 “Big Data Research and Development Initiative.”
  By improving our ability to extract knowledge and insights from large and complex collections of digital data.

The Office of Basic Energy Sciences (BES) BES Scientific User Facilities have supported a number of efforts aimed at assisting users with data management and analysis of big data, which can be as big as terabytes (10^{12} bytes) of data per day from a single experiment. For example, the Accelerating Data Acquisition, Reduction and Analysis (ADARA) project addresses the data workflow needs of the Spallation Neutron Source (SNS) data system to provide real-time analysis for experimental control, and the Coherent X-ray Imaging Data Bank has been created to maximize data availability and more efficient use of synchrotron light sources.

In October 2011, the Data and Communications in Basic Energy Sciences workshop sponsored by BES and ASCR identified needs in experimental data that could impact the progress of scientific discovery.
Data is Reshaping HPC

- The observational dataset for the Large Synoptic Survey Telescope will be ~100 PB
- The Daya Bay project will require simulations which will use over 128 PB of aggregate memory

- By 2017 ATLAS/CMS will have generated 190 PB
- Light Source Data Projections:
  - 2009: 65 TB/yr
  - 2011: 312 TB/yr
  - 2013: 1.9 PB /yr
  - EB in 2021?
  - FELs expected to generate data at a Terabit per second
NERSC’s Data Directions

• Field cutting edge HPC Storage and Data Technologies
• Bring BigData solutions to science partners who are ready
  – Bioinformatics, bioimaging and health (JGI, NGBI LDRD, UCSF/NIH)
  – Photon science, Synchrotrons (ALS, SLAC, NGLS)
  – Materials Genomics, Manufacturing (MatProj, BES, JCESR, CA)
• Science Gateways
  – Bring advanced computing and data with the world’s scientists
  – Database-driven workflows and storage
  – Scalable structured and unstructured object stores
  – Software solutions to traverse massive data for search or analysis
  – Sophisticated web-based gateways to interact with and leverage data
  – Comprehensive scientific data curation beyond simple archiving
Photo-system II, Nick Sauter (LBL): A single experiment in 2013 generated 150TB over four days. Data was relayed to NERSC via Esnet, analysis required 135K CPU hours of computing.

LCLS-II will require > 100M CPU hours per experiment. Higher resolution and advanced image analysis could grow computational complexity. Some algorithms scale an M*NlogN for M images of N pixels.
Diversity is our Challenge

**Reciprocal Space** (Scattering)
- HipGISAXS/HipMC parallel Scattering

**Real Space** (Tomography)
- Arec3d, QuantCT, CrunchFlow

**Spectroscopy** (MicroXas)
- ShirleyXAS (MSD)
- BerkeleyGW (NERSC)

**Hybrid** (COSMIC)
- Ptychographic reconstruction
ESnet: DOE’s Leadership and Production Network

DOE Science Network:
- 72% annual traffic growth
- International collaborations
- Bandwidth reservations and monitoring

Advanced Networking Initiative (ANI):
- 100 Gbps network
- Contract to Internet2 signed in July
- Demo planned in late 2011
- Separate network research testbed serving 17 projects

ESnet+ANI, DOE will be the world leader in networking for science.
NERSC Facility Leads DOE in Scientific Computing Productivity

NERSC computing for science
- 4000 users, 500 projects
- 1500 publications per year

Systems designed for science
- 1.3PF Petaflop Cray system, Hopper
  - 2nd Fastest computer in US and one of 2 Petaflop systems in Office of Science
  - 26 PB HPSS tape storage (150M files); 6 PB GPFS disk system
  - GPU (Dirac), capacity (Carver), and specialized systems (PDSF, JGI, Plank)
SPOT Suite – ALS Data Analysis LDRD

• “Towards an end-to-end solution for light source data” – Craig E. Tull (CRD), Jack Deslippe (NERSC), Alex Hexemer (ALS), David Prendergast (MSD), Brian Tierney (ESNet)
  – FY2013 focus is on real-time analysis of beam-time data.
• Researching workflows and metadata for 5 ALS Beamlines: Tomography, GISAXS, XAS Microdiffraction, Ptychography
• Integrating Facility resources from ALS, ESNet, and NERSC to extend ALS Facility’s capability.
• FY2014 to integrate supercomputer simulations with automated data management and analysis.
How the user sees end-to-end.

Framework

User Initiated Simulation
User Initiated Analysis
Real-time Analysis
Real-time Analysis Pipeline
Data Pipeline
HPC Storage and Compute

* LDRD R&D

User

Control & Feedback

Experiment

Real-time Analysis

Compare
Experiment

User Initiated Simulation

Control & Feedback

Real-time Analysis Pipeline

Data Pipeline

HPC Storage and Compute
Real-time analysis will have high-impact.

- 3-D beam-time feedback requires stopping data taking for > 1 hour for data processing.
- Many users analyze a small fraction.

Geologic CO2 Sequestration
Jonathan Ajo-Franklin
Time-resolved, in-situ Tomography

- Turbines are more efficient at high temperature
- Full 3D quantitative images inform and constrain computer models

SPOT Suite Components:

• Beamline DTN: ESNet's DTN Reference Architecture
• Data Suitcasing: Beamline specific HDF5 conversion code
• SPADE Data Transfer: generic
• MetaData DataBase: MongoDB at NERSC
• Data Portal: browsing, monitoring, analysis, simulation, AAA
• NERSC*: Carver, Hopper, Edison, NGFS, HPSS, SDG
• HPC-optimized Photon Science Codes:
  – Fiji, GridRec, ImgRec, XMAS, ShirleyXAS, HipGISAXS…
• Rabbit-MQ Job submission system: (in development)
• Workflow System: (SDAV PostDoc hire)
• WAN Data Access & Sharing: (SDAV PostDoc hire)
FY13 Significant Accomplishments

• 4 Tomography codes ported to NERSC
  – Parallelized run-time speed-ups of x8 for average scan. Multiple reconstructions. HDF5 input & output (no parallel I/O yet).
• Integrated Queriable/Browsable Interface
  – Raw & Derived Data, MetaData, Provenance
• Prototype interface for XMAS (Micro-diffraction)
• 1-BL insights: Workflow/throughput optimizations, metadata/provenance, data parallelism (transverse vs. longitudinal), ensemble results, human steering, processing locality, transfer modes (file vs stream)

March 21, 2013 - First successful round-trip test of workflow-triggered "real-Wme" data analysis with replayed data.

On Mar 20, 2013, at 09:34, Dula Parkinson <dparkinson@gmail.com> wrote:
That is awesome! It makes me want to cry, it's so beautiful. OK, maybe not cry, but to see how far this has come is really, really great.

- Jonathan Ajo-Franklin Beamwme (04/18/2013 - 04/20/2013)

"At present, we are hobbled during dynamic experiments by the lack of real-Wme feedback, due primarily to the delay in processing newly acquired data. This LDRD is showing real progress in eliminating this constraint; initial tests during our dynamic runs have demonstrated that the pipeline, with small modifications, will soon replace our current processing approach and dramatically improve the speed at which we can evaluate experimental data."

From: Dula Parkinson <dparkinson@gmail.com>
Subject: Re: spot.nersc.gov
Date: August 1, 2013 19:35:56 PDT
OK, I just checked out some volume renderings of data collected this morning--from my laptop, on vacation at Lake Tahoe, with a super-slow internet connection. I really never would have thought that was possible a year ago, and here it is a reality.
SPOT Suite – ALS Data Analysis LDRD

- Collaboration between ALS, MSD and CRD, NERSC, ESNet.
- Real time tomography codes ported to NERSC and in production usage. (Aug 1)
  - AUG: 34 TB, 264 CPU-days
- Integrated Queriable/Browsable Interface
  - Raw Data, Derived Data, MetaData, Provenance, Access Control & Sharing
- Development for Micro-diffraction, SAX/WAXS/GISAXS and Ptychography
- [http://spot.nersc.gov/](http://spot.nersc.gov/)
Simulated Spectrum (Shirley XAS)
Conclusion & Future Plans

• CRD brings data-intensive/large-scale science experience and advanced computational science expertise.
• ALS represents each data "theme", provides algorithm expertise, and allows at-scale tests using real data.
• CRD, MSD, NERSC supply HPC-ready simulation codes.
• NERSC, ESNet provide raw capacity and experience to solve problems at scale. (eg. DTN Reference Architecture)
• Year 2 Goals:
  – Year 1 system to 4 other Beamlines (new data & workflows)
    • WAXS/SAXS, uDiffraction, Ptycography
  – Test impact of Y1 system on 3 experiments in real beamtimes
  – Optimize Tomography performance (I/O, MPP, streaming) and HTC jobs at NERSC (eg. Rabbit-MQ based job submission)
  – Develop human steering & input strategies & capabilities
  – Integrate PostDoc from SDAV Data Pilot Project
  – Explore partnerships – LCLS, SSRL, APS, NSLS
Participants

- CRD brings data-intensive/large-scale science experience and advanced computational science expertise.
- ALS represents each data "theme", provides algorithm expertise, and allows at-scale tests using real data.
- CRD, MSD, NERSC supply HPC-ready simulation codes.
- NERSC, ESNet provide raw capacity and experience at scale
- **CRD**: C.E.Tull (25%), A.Essiari (100%), SDAV PostDoc (100%), D.Gunter, X.S.Li, Simon Patton (10%), L.Ramakrishnan
- **NERSC**: R.S.Canon, J.Deslippe (35%), D.Skinner, Tyler & Dylan (interns@25%)
- **ALS**: E.Chan, A.Hexemer, S.Marchesini, M.Marcus, D.Parkinson, N.Tamura, D.Shapiro, C.Wang
  - J.Blair (postbac@100%), R.De Costa, (postbac@40%), M.Wang (student@50%)
  - J.Lee (CSGF fellow@20%)
- **ESNet**: E.Dart, I.Monga, E.Pouyoul, B.Tierney
- **MSD**: D.Prendergast
- **ALS Participants**: J.Blair (ALS), R.De Costa, (ALS), M.Wang (ALS)
- **Science Customers**: N.Balsara (MSD/EETD, USB), J.Ajo-Franklin (ESD), R.Richie (MSD, UCB), T.Willey (LLNL)

* FY2013 off-LDRD efforts
* FY2014 off-LDRD efforts
THANK YOU
XIE-XIE 谢谢
Shokran شكرًا