Big Data and Scientific Computing: Some Initial Thoughts

Rob Ross

Mathematics and Computer Science Division

Argonne National Laboratory

rross@mcs.anl.gov

Defining Big Data

M. Stonebreaker defines big data with “the 3 V’s”. A big data application has one of the following [1]:

- **Big Volume –** The application consumes terabytes (TB) or more data.
- **Big Velocity –** An application has much data, moving very fast.
- **Big Variety –** The application integrates data from a large variety of data sources.

Similarly, the International Data Corporation (IDC) defines big data projects [2] to:

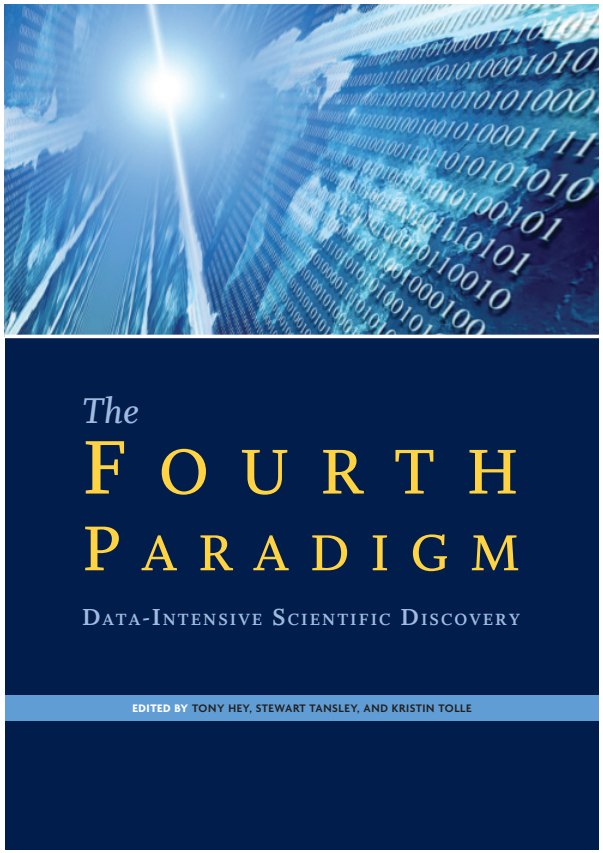
- Involve the collection of more than 100 terabytes of data, or
- High-speed, real-time streaming of data, or
- Projects with data growing by 60 percent or more a year.
- Typically involve two or more data formats.

[1] http://siliconangle.com/blog/2011/12/29/newsql-will-prevail-in-2012-says-mits-michael-stonebraker/

[2] S. Lehr’s NYT summary of IDC, “Big Data: Global Overview,” March 2012.

Data Intensive Science

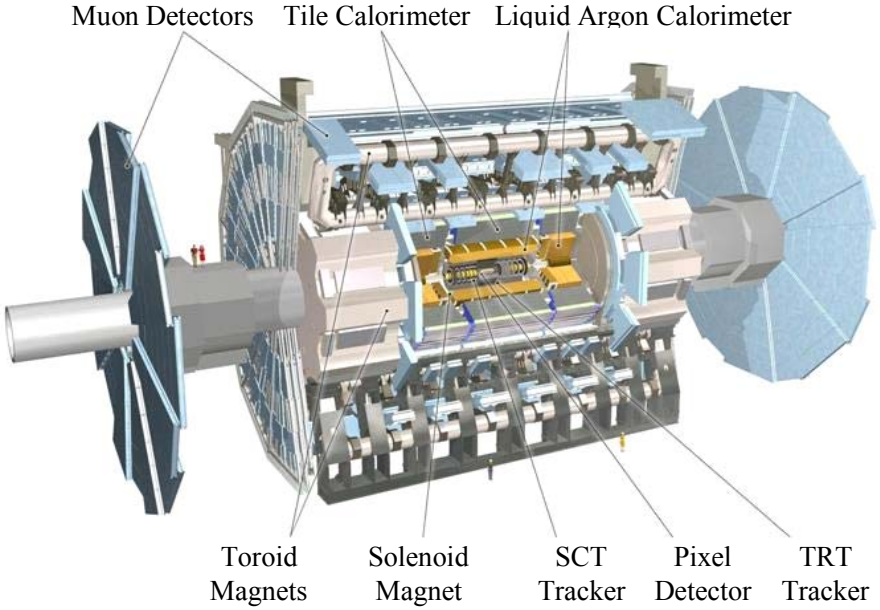
Data Intensive Science is the pursuit of scientific discoveries via the capture, curation, and analysis of big science data.



Data can come from a variety of sources:

- Experimental systems
  (e.g., ATLAS experiment at the Large Hadron Collider)
- High-throughput screening and sequencing
  (e.g., resulting in GenBank sequence database)
- Observational platforms
  (e.g., Sloan Digital Sky Survey)
- Sensor networks (e.g., environmental monitoring)
- Simulations combined with other data sources (e.g., cosmology)

ATLAS Detector at the Large Hadron Collider



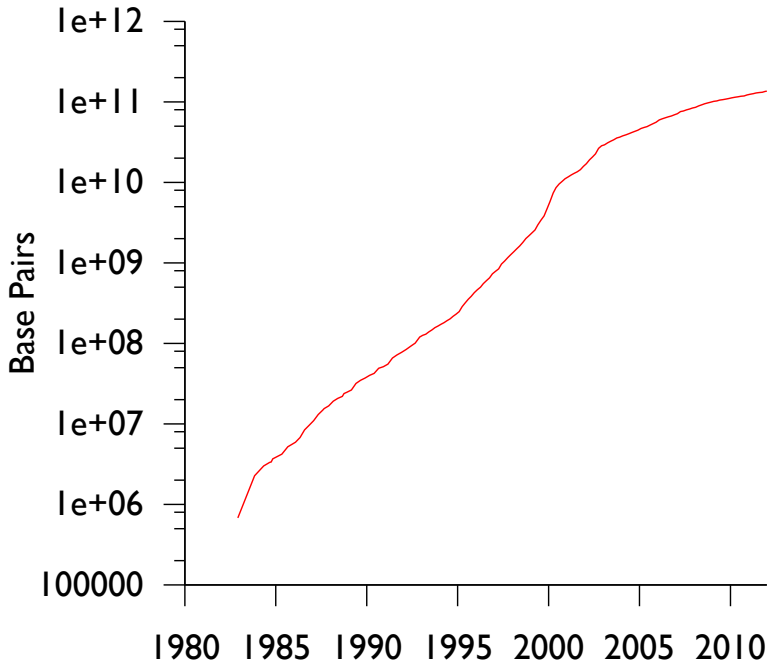
ATLAS has an "onion-layer" structure with the innermost being the inner detector (Pixels, Silicon Tracker, and Transition Radiation Tracker), followed by the calorimeters (Liquid Argon and Tile) and the muon detectors.

- Data reduction performed in the detector and in supporting cluster systems
- Multiple data types stored:
  - "Raw" events (serialization of detector readouts)
  - Summarized events (reduced version of raw events)
  - Analysis object data capturing physics objects (e.g., jets, muons)
- Complex software infrastructure (POOL/ROOT) to serialize objects, store them, and manage relationships

P. van Gemmeren et al, "The Event Data Store and I/O Framework for the ATLAS Experiment at the Large Hadron Collider", IASDS 2009.

Genetic Sequence Data Bank (GeneBank)

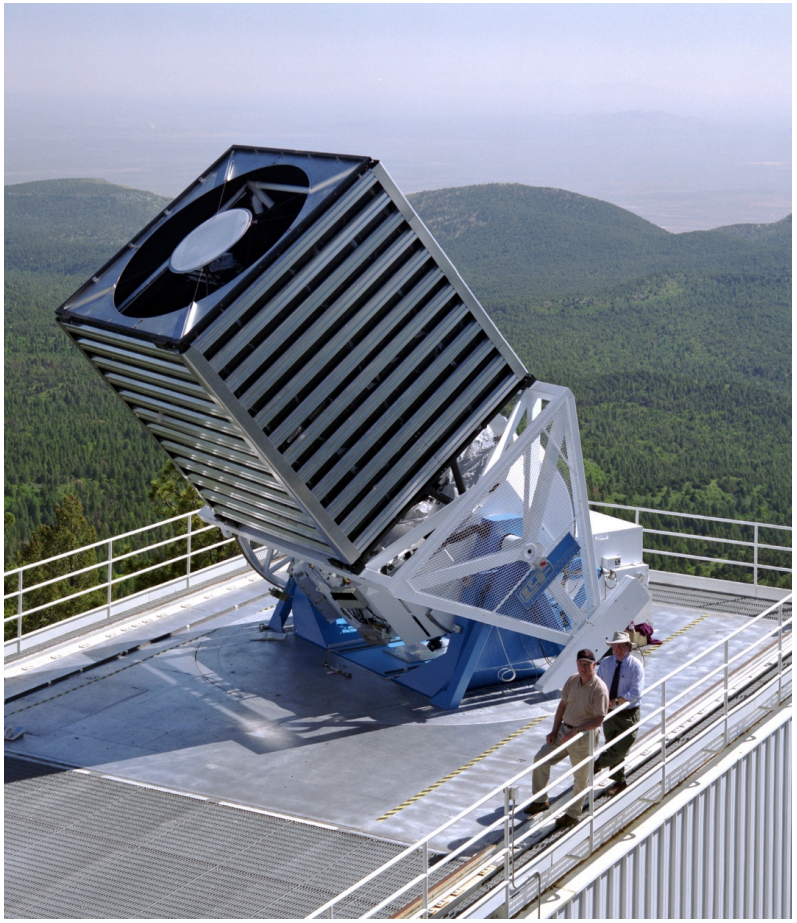
```
1        10        20        30        40        50        60        70        79
---------+---------+---------+---------+---------+---------+---------+---------
GBSMP.SEQ           Genetic Sequence Data Bank
                          April 15 1992

                  GenBank Flat File Release 74.0

                      Structural RNA Sequences

      2 loci,       236 bases, from     2 reported sequences

LOCUS       AAURRA        118 bp ss-rRNA             RNA       16-JUN-1986
DEFINITION  A.auricula-judae (mushroom) 5S ribosomal RNA.
ACCESSION   K03160
VERSION     K03160.1  GI:173593
KEYWORDS    5S ribosomal RNA; ribosomal RNA.
SOURCE      A.auricula-judae (mushroom) ribosomal RNA.
  ORGANISM  Auricularia auricula-judae
            Eukaryota; Fungi; Eumycota; Basidiomycotina; Phragmobasidiomycetes;
            Heterobasidiomycetidae; Auriculariales; Auriculariaceae.
REFERENCE   1  (bases 1 to 118)
  AUTHORS   Huysmans,E., Dams,E., Vandenberghe,A. and De Wachter,R.
  TITLE     The nucleotide sequences of the 5S rRNAs of four mushrooms and
            their use in studying the phylogenetic position of basidiomycetes
            among the eukaryotes
  JOURNAL   Nucleic Acids Res. 11, 2871-2880 (1983)
FEATURES             Location/Qualifiers
     rRNA            1..118
                     /note="5S ribosomal RNA"
BASE COUNT       27 a     34 c     34 g     23 t
ORIGIN      5' end of mature rRNA.
        1 atccacggcc ataggactct gaaagcactg catcccgtcc gatctgcaaa gttaaccaga
       61 gtaccgccca gttagtacca cggtgggga ccacgcggga atcctgggtg ctgtggtt
//
…
---------+---------+---------+---------+---------+---------+---------+---------
1        10        20        30        40        50        60        70        79
```



Nucleotide sequence database maintained by National Center for Biotechnology Information.

1790 formatted text files, ~500 Gbytes, 100+ billion base pairs, significant curation demands, growing exponentially.

Sloan Digital Sky Survey



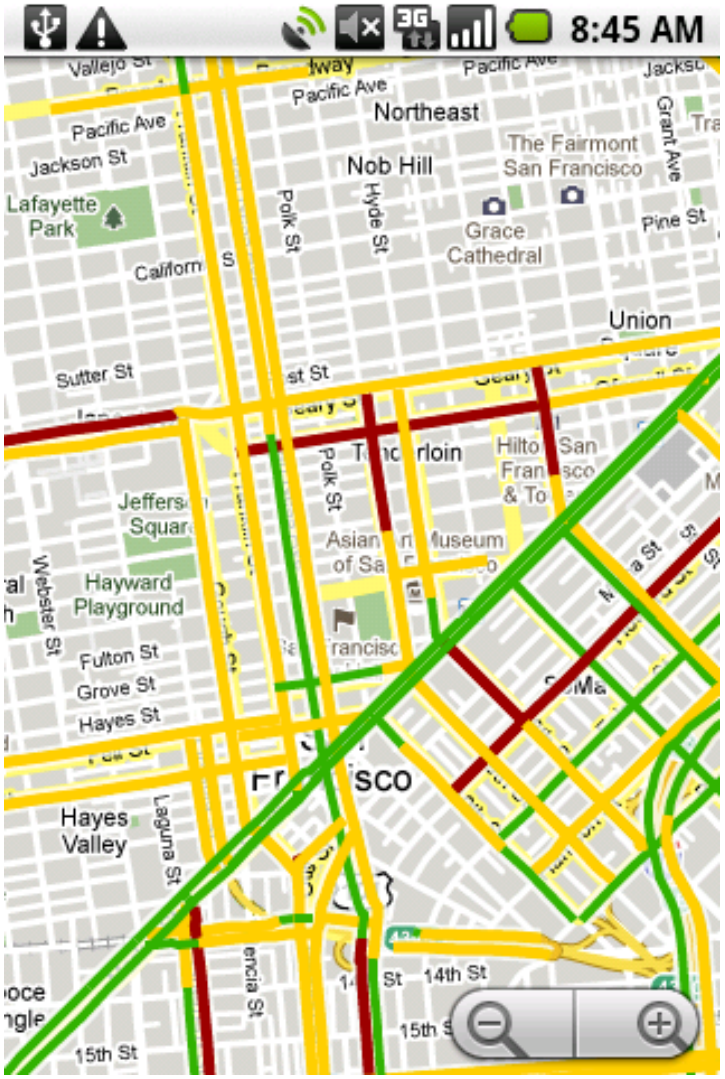
- The instrument:
  - 2.5m wide-angle optical telescope
  - 120 Mpixels
  - Integrated spectrograph
  - Online since 2000
- Original survey: 230 million objects, 930,000 galaxies (primary targets), 120,000 quasars, and 225,000 stars
- Four main datasets totaling ~40TBytes:
  - Photometric catalog (500 attributes for each element, reference to bitmap images)
  - Spectroscopic catalog (admission and absorption lines, reference 1D spectra)
  - Bitmap images in multiple color bands
  - Spectra
- Data managed/accessed in multiple ways, SQL DB, Objectivity DB, flat files

Szalay et al, "The SDSS SkyServer – Public Access to the Sloan Digital Sky Survey Data," ACM SIGMOD, 2002.

Sloan Digital Sky Survey web site, http://www.sdss.org/.

Fermilab Visual Media Services, "Sloan Digital Sky Survey telescope," http://www.sdss.org/gallery/gal_photos.html, 1998.

Crowdsourced Traffic Congestion Monitoring



- Mobile phones as a sensor network
- Enabled in Google Maps for mobile
- Augments existing sensors:
  - Point detectors (e.g., inductance loops, radar, video)
  - Beacon-based probes (e.g., electronic toll passes)
- Especially helpful in providing data on arterial roads

Goal: Keep scientists off the road and in their laboratories.

"The bright side of sitting in traffic: Crowdsourcing road congestion data," http://googleblog.blogspot.fr/2009/08/bright-side-of-sitting-in-traffic.html

Combining Simulation with Observations

Large Synoptic Survey Telescope

- Capabilities
  - 3D maps of mass distribution
  - Time-lapse imaging of faint, moving objects
  - Census of solar system down to 100m objects
- Wide field survey telescope
  - Comes online in 2019
  - 3.2 Gpixel, exposure every 20 secs
  - 1.28 PBytes per year
- Over 100 PBytes of data after processing

http://www.lsst.org

Hybrid/Hardware Accelerated Cosmology Code Framework

- Building understanding of structure formation of universe
- Simulation critical to understanding latter, nonlinear, half of history
- Code ported to multiple leadership computing platforms, running at full scale on Intrepid BG/P system, Mira BG/Q system

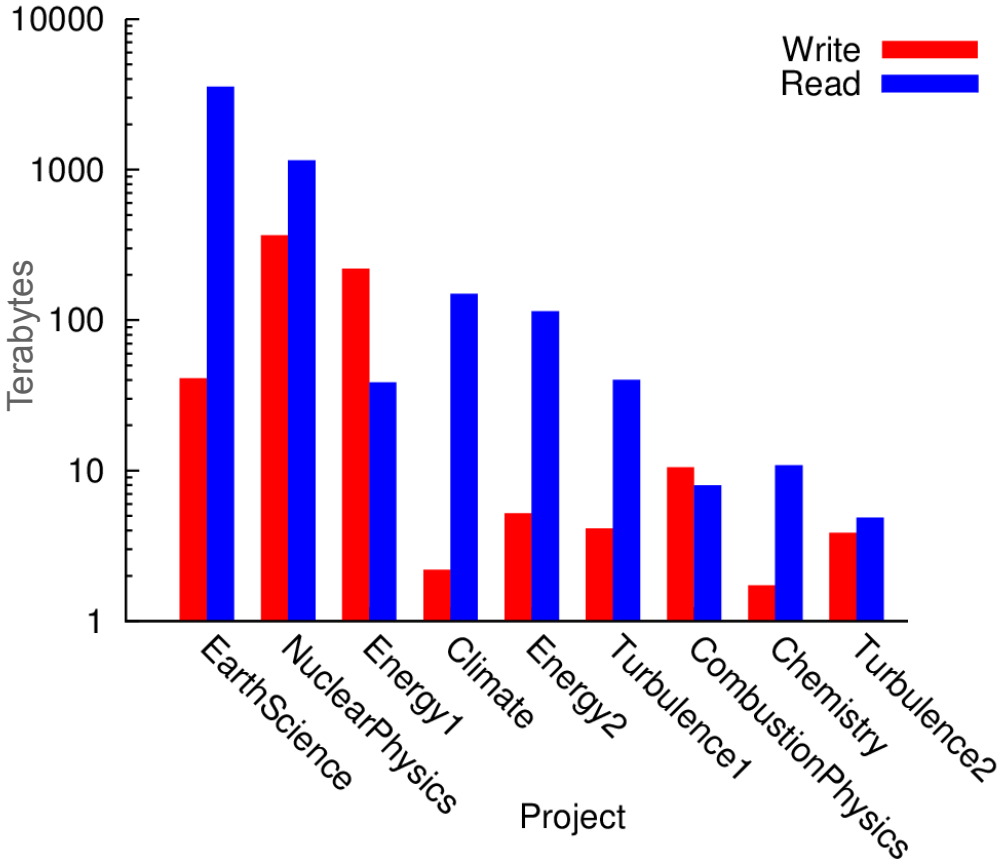
Goal: Combine results of simulations with the observations, using statistical methods, to infer the dynamical laws governing the evolution of the universe.

What about Computational Science?

Computational Science has Big Volume

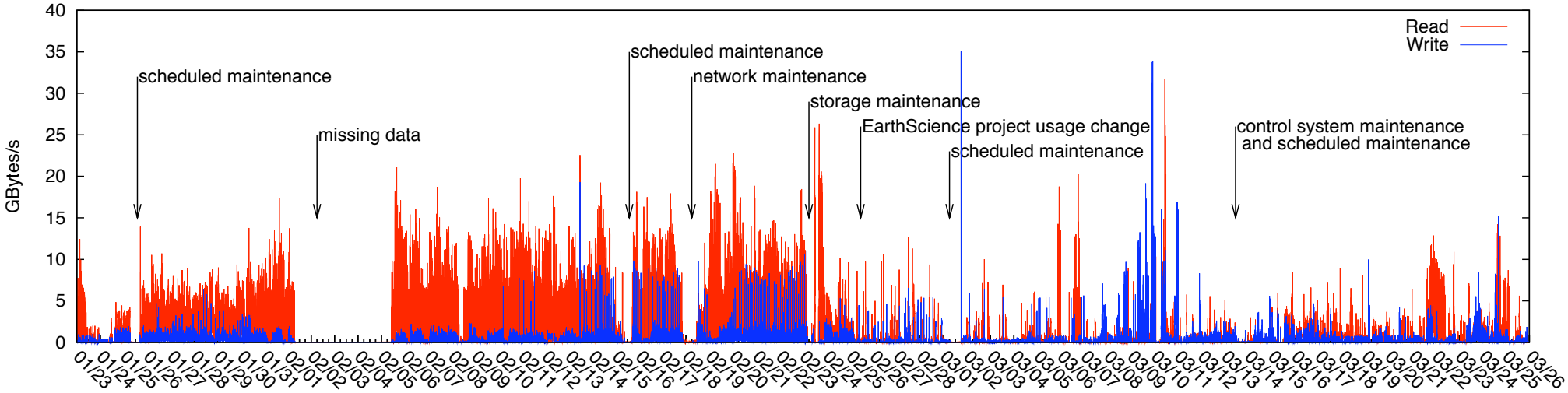
Data requirements for select 2011 INCITE applications at ALCF

PI	Project	On-line Data (TBytes)	Off-line Data (TBytes)
Khokhlov	Combustion in Gaseous Mixtures	1	17
Baker	Protein Structure	1	2
Hinkel	Laser-Plasma Interactions	60	60
Lamb	Type Ia Supernovae	75	300
Vary	Nuclear Structure and Reactions	6	15
Fischer	Fast Neutron Reactors	100	100
Mackenzie	Lattice QCD	300	70
Vashishta	Fracture Behavior in Materials	12	72
Moser	Engineering Design of Fluid Systems	3	200
Lele	Multi-material Mixing	215	100
Kurien	Turbulent Flows	10	20
Jordan	Earthquake Wave Propagation	1000	1000
Tang	Fustion Reactor Design	50	100



Amount of data accessed by top I/O users during two month window on ALCF BG/P [Carns 2011].

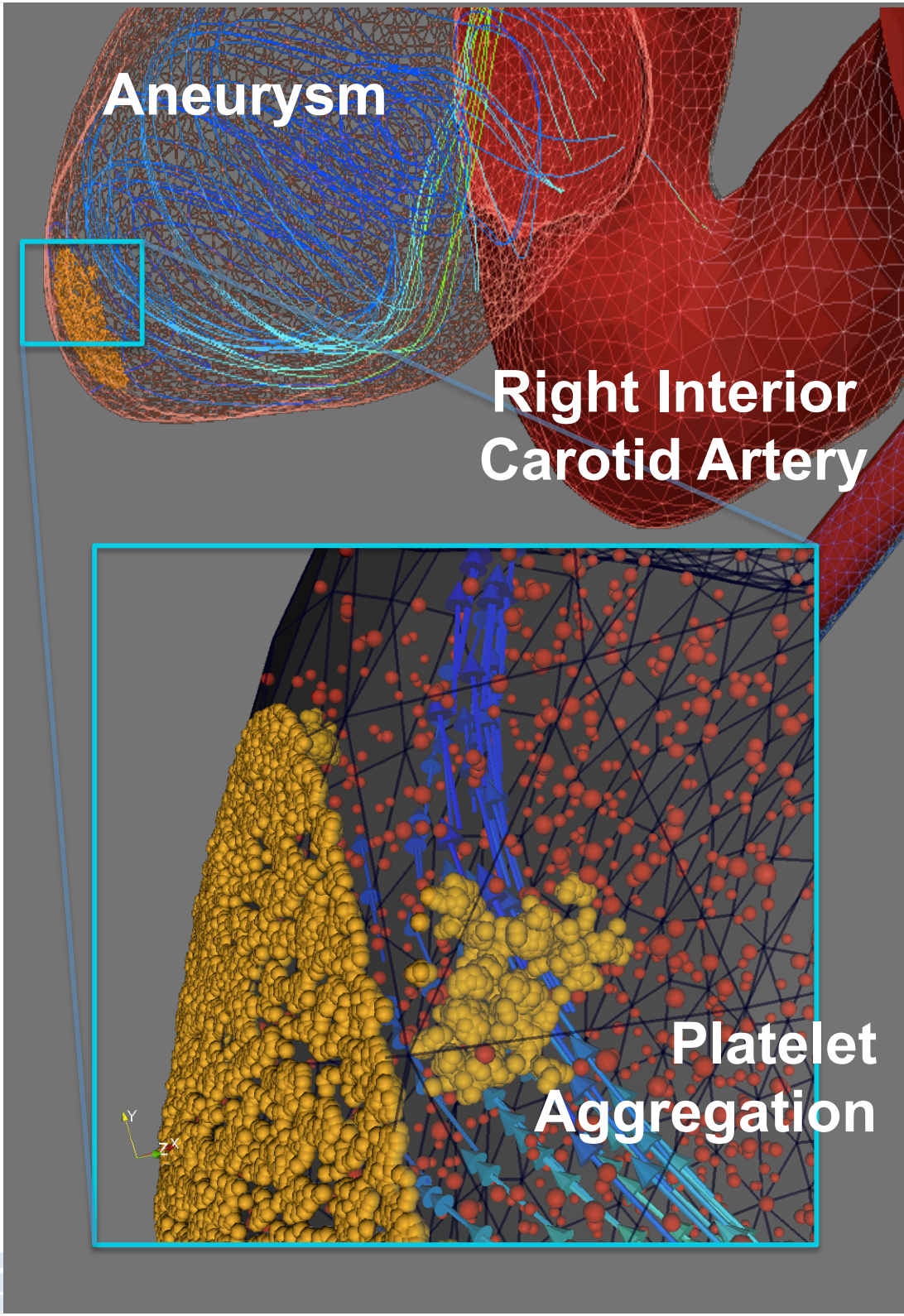
Computational Science has Big Velocity



- Shown here: aggregate I/O throughput on Argonne BG/P storage servers at one minute intervals. Peaks in the 10s of Gbytes/sec.
- Blue Waters peak I/O rates of 1TByte/sec!

Computational Science has Big(ish) Variety

- Complexity as an artifact of science problems and codes:
  - Coupled multi-scale simulations generate multi-component dataset.
  - Atomistic data representations for plasma, red blood cells, and platelets from MD simulation.
  - Field data for ensemble average solution generated by spectral element method hydrodynamics code [Grinberg 2011, Insley 2011]



Much Computational Science is Data Intensive Science

- The leadership computing system is an instrument
- Recall: Data intensive science is the pursuit of scientific discoveries via the capture, curation, and analysis of big science data
  - **Capture** increasingly includes in situ data reduction, the simulation analog of detector triggers
  - **Curation** includes storing provenance necessary to repeat simulations
  - **Analysis** often includes reorganizing data into formats more amenable to processing, generation of derived datasets

An Example

Understanding behavior of a laser pulse propagating through a hydrogen plasma

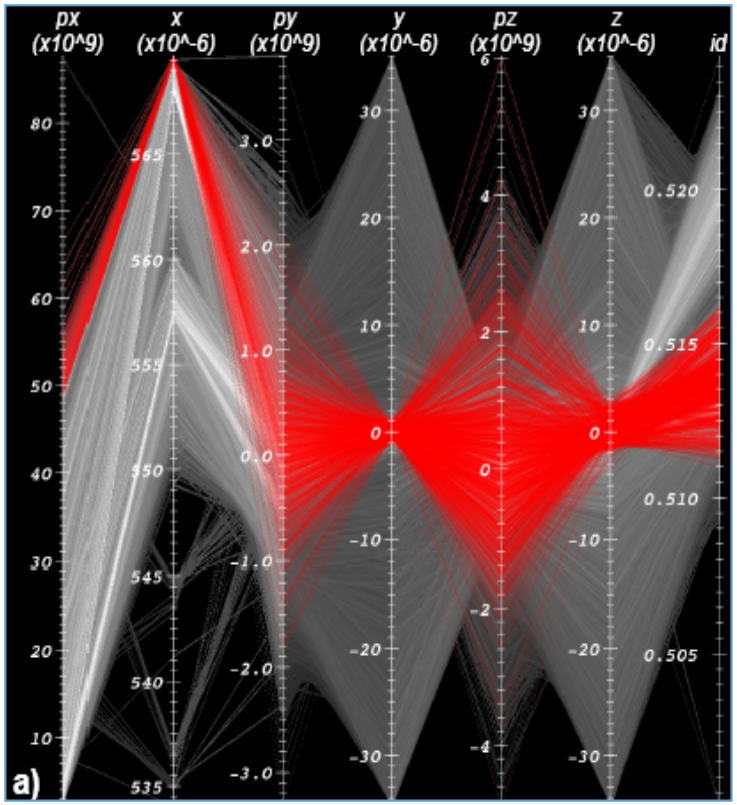
- VORPAL code used to simulate laser wakefield particle accelerator
  - 3D simulation
  - 30 timesteps
  - 90 million particles per timestep, ~5 Gbytes of data per timestep (circa 2008, a “big” run might have 100+ billion particles/timestep now)
- Questions:
  - Which particles become accelerated? How are they accelerated?
  - How did the beam form? How did it evolve?
- Data management, analysis, and visualization:
  - **Data model support** – HDF5, H5Part to store data with appropriate metadata
  - **Indexing** – FastBit to enable quick identification of particles of interest, associate particles between timesteps
  - **Visualization** – Parallel coordinates view to help user select particles, VisIt as deployment vehicle

Rubel et al. High performance multivariate visual data exploration for extremely large data. SC08. November, 2008.

Beam Selection

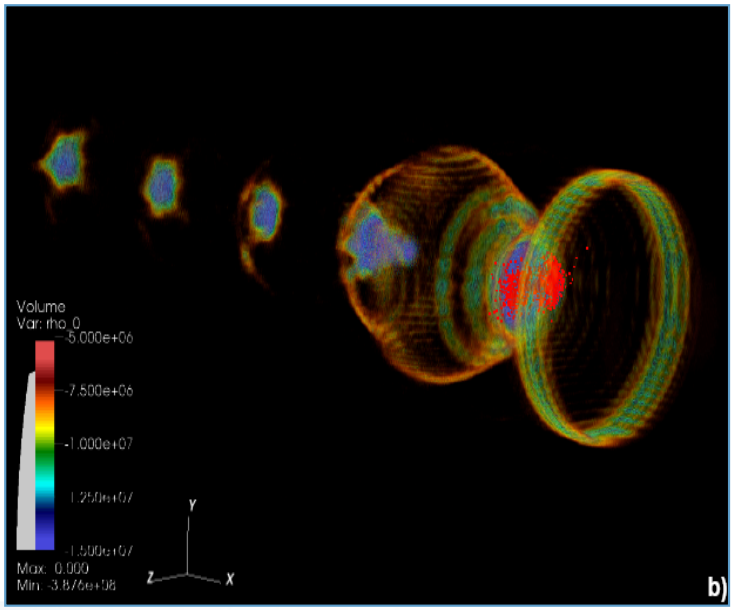
Parallel coordinates view of $t = 12$

- Grey particles represent initial selection ($px > 2*10^9$)
- Red particles represent "focus particles" in first wake period following pulse ($px > 4.856*10^{10}$) && ($x > 5.649*10^{-4}$)



Volume rendering of plasma density with focus particles included in red ($t = 12$)

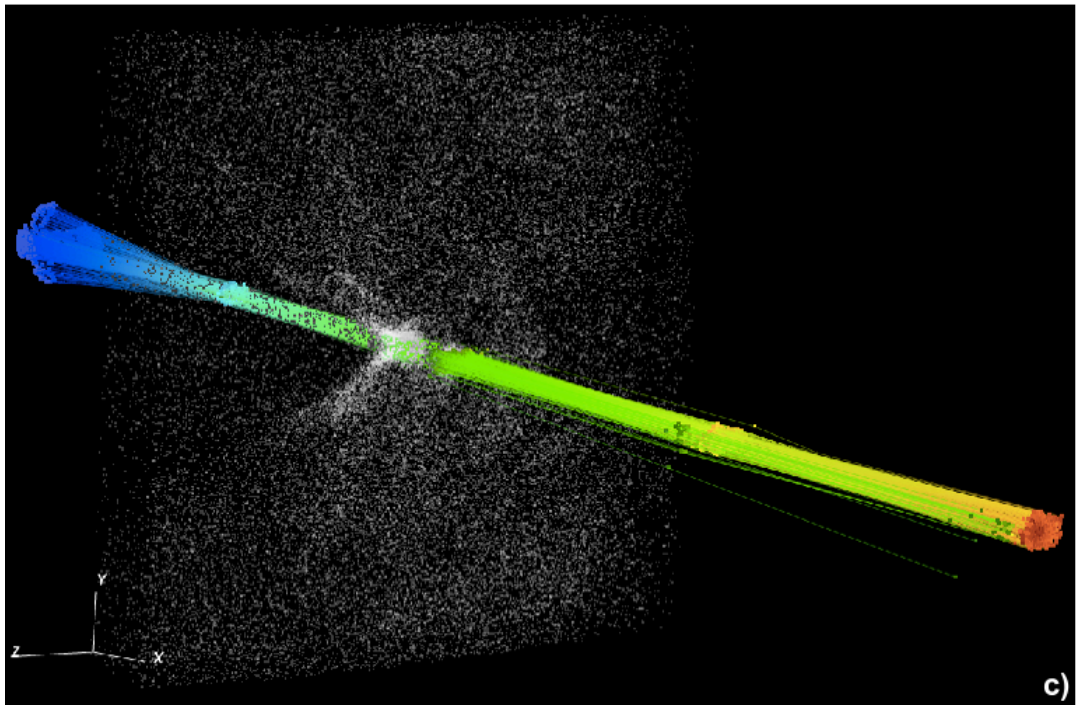
- Helps locate beam within wake



Tracing Particles over Time

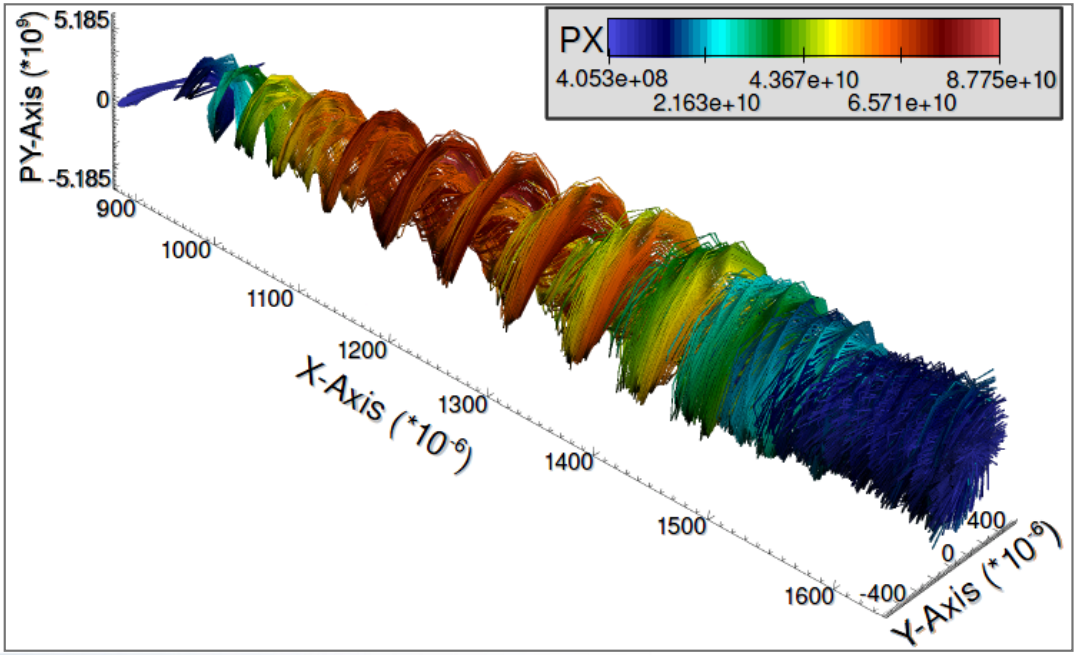
Tracing particles back to *t* = 9 and forward to *t* = 14 allows scientist to see acceleration over time period:

- Heatmap shows particles constantly accelerated over time (increase in *px*, left to right).
- Grey particles show initial selection (for reference).



More recent work shows:

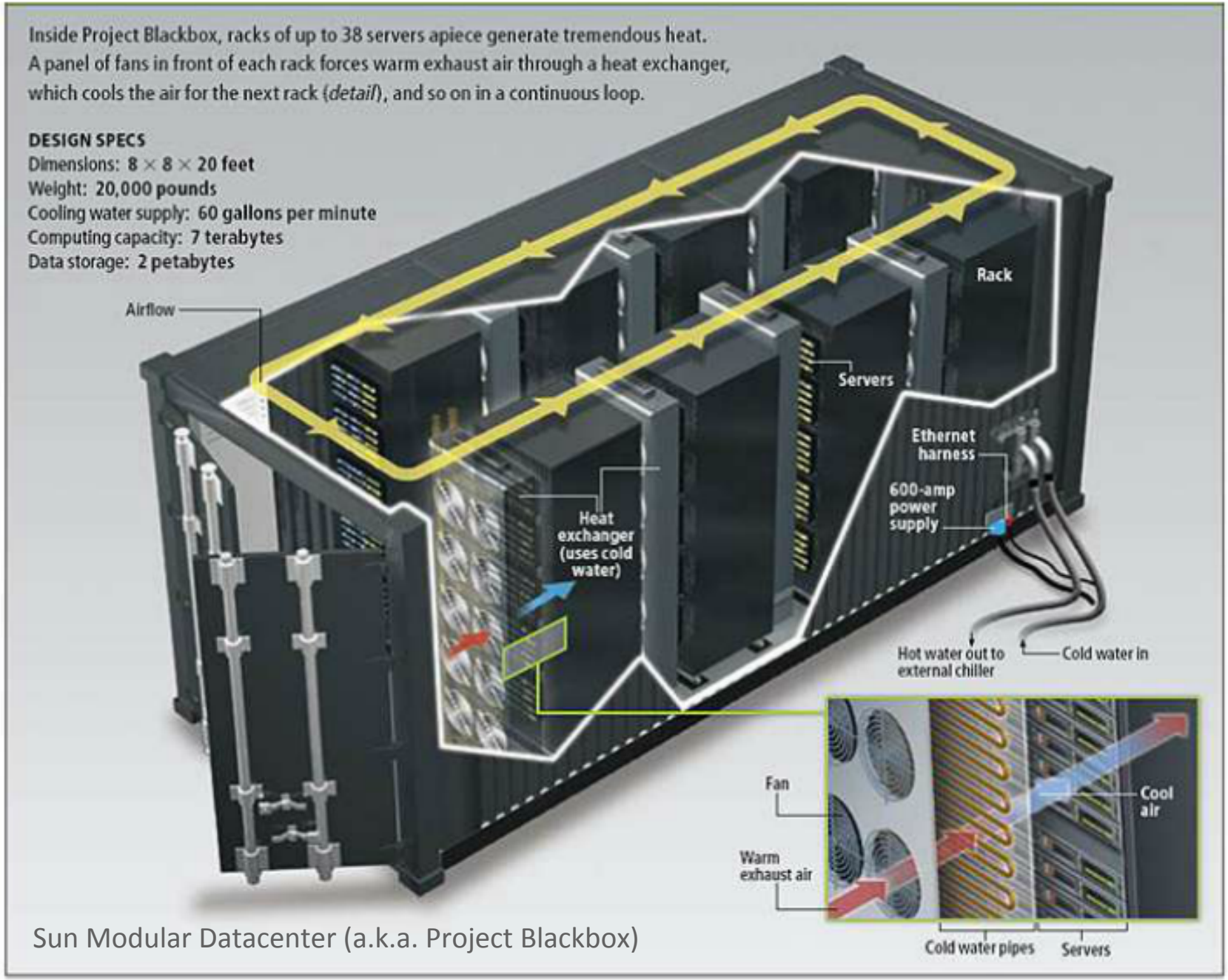
- Particles start out slow (blue, left), undergo acceleration (reds), then slow again as the plasma wave outruns them (blue, right).
- Spiral structure shows particles oscillating transversely in the focusing field (new science).



How might big data influence leading computing facilities and systems?

Extreme-Scale Computational Science Systems

Extreme-Scale Data Intensive Systems



Sun Modular Datacenter (a.k.a. Project Blackbox)

Shifting Emphasis at Compute Facilities?

- As data becomes increasingly difficult/costly to move, emphasis shifts further towards moving analysis to the data.
- In the context of experimental facilities, this means co-locating data intensive computing systems.
- Leading data intensive computing systems will become magnets for important datasets.

- Where sites host both data intensive science projects and computational science systems, opportunities exist for leveraging data intensive computing infrastructure in the computational science context as well.

Overhauling the Traditional HPC Support Systems

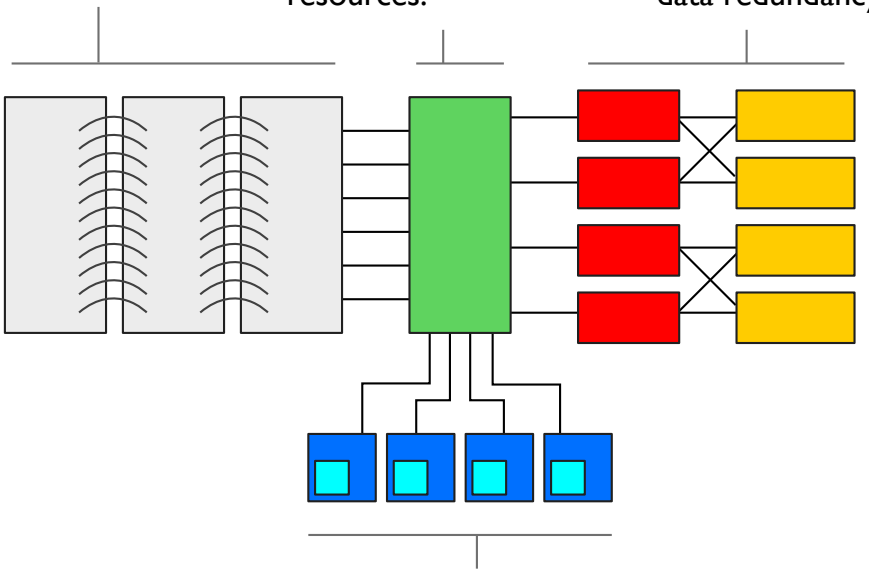
Data intensive computing systems hold the promise of replacing traditional enterprise storage and visualization clusters while supporting a wide variety of new science endeavors.

Current Leadership Computing Architecture

Leadership computing system executes simulation codes in batch mode.

Commodity network attaches leadership computing system to storage and analysis resources.

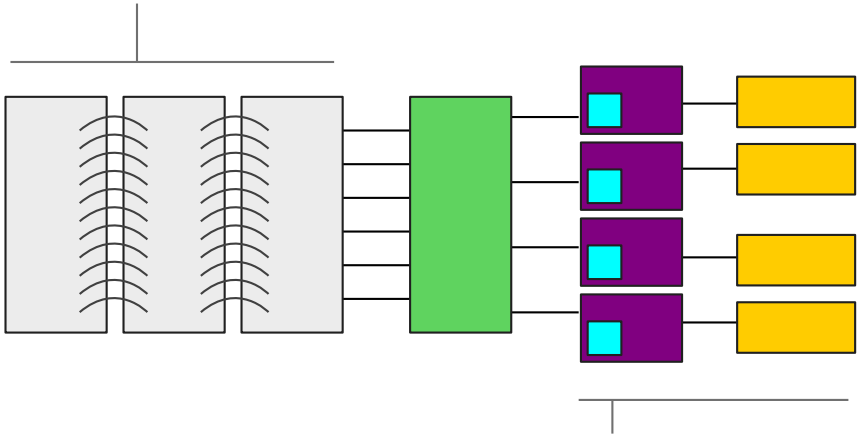
Storage nodes run parallel file system server software. Attached to enterprise storage for data redundancy.



Visualization nodes perform analysis calculations. Usually multi-core nodes with GPU resources.

Leadership Computing System with Data Intensive Back-end

Scientific codes execute unchanged, writing to storage as before.

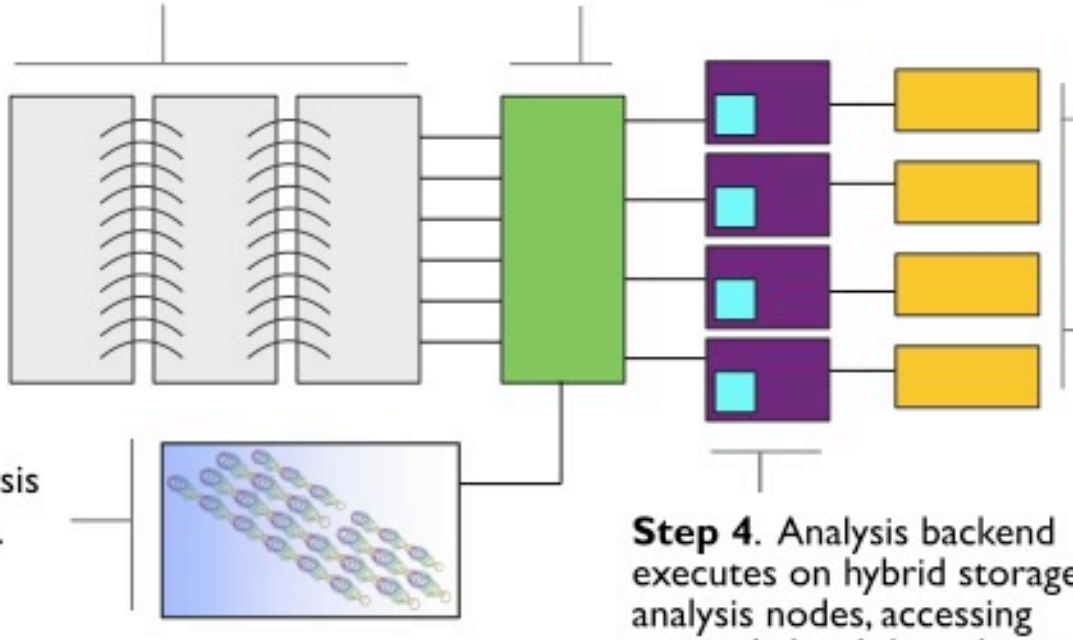


Data intensive back-end system places analysis operations close to data. Commodity storage reduces overall system cost, but requires more sophisticated storage software.

Laser Pulse Propagation with Data Intensive Computing Support

Step 1. VORPAL simulation executes unchanged on compute nodes.

Step 2. Data written through H5Part/HDF5 directly to objects. Semantic structure stored alongside.



Step 3. Fastbit indices calculated using active storage functions, leveraging semantic information stored alongside data.

Step 5. Access patterns are observed and system reorganizes data, storing alternative layouts.

User drives analysis via GUI frontend.

Step 4. Analysis backend executes on hybrid storage/ analysis nodes, accessing primarily local data elements.

Applying Ourselves

Data Intensive Science Areas at Argonne

- Materials science
- Atmospheric science
- Cosmology
- Computational biology
- Urban planning
- High energy physics
- Facility monitoring via sensor networks

Research Areas in Data Intensive Scientific Computing

- Capture
  - Efficient methods for persisting data
  - Impedance matching between data sources and data intensive computing system
  - Methods of leveraging heterogeneous storage
  - Targeted data reduction (i.e., determine and retain what the scientist needs)
- Curation
  - Automation of provenance collection
  - Accelerating development of ontologies and schemas for science data
  - Storage (especially with respect to long-term resilience)
- Analysis
  - Programming models for scientific data analysis
  - Runtime, scheduling, operating system support for large scale DISC systems
  - New algorithms for analyzing large, complex scientific datasets
    - Statistical, graph, …
  - Tools for data movement and sharing
- Reusable tools to support multiple domains

Thanks!