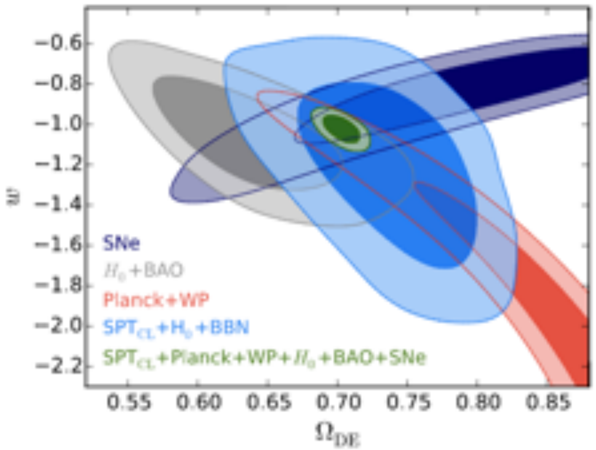
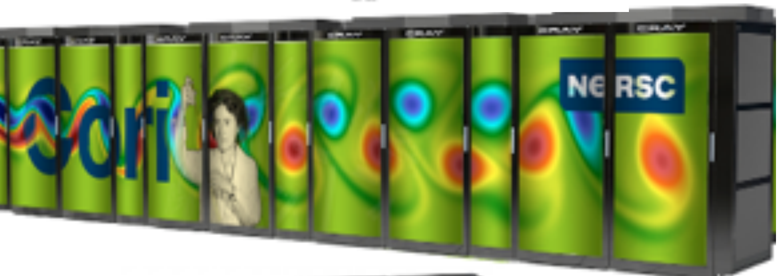
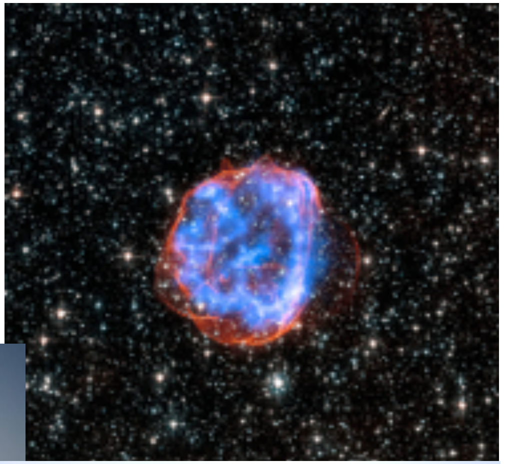
Streaming Data in Cosmology

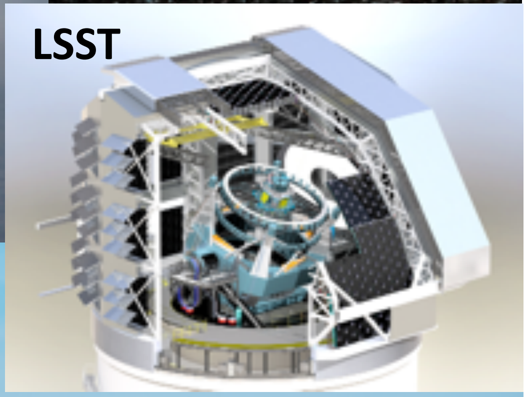
Salman Habib
Argonne National Laboratory

Stream 2016, March 22, 2016

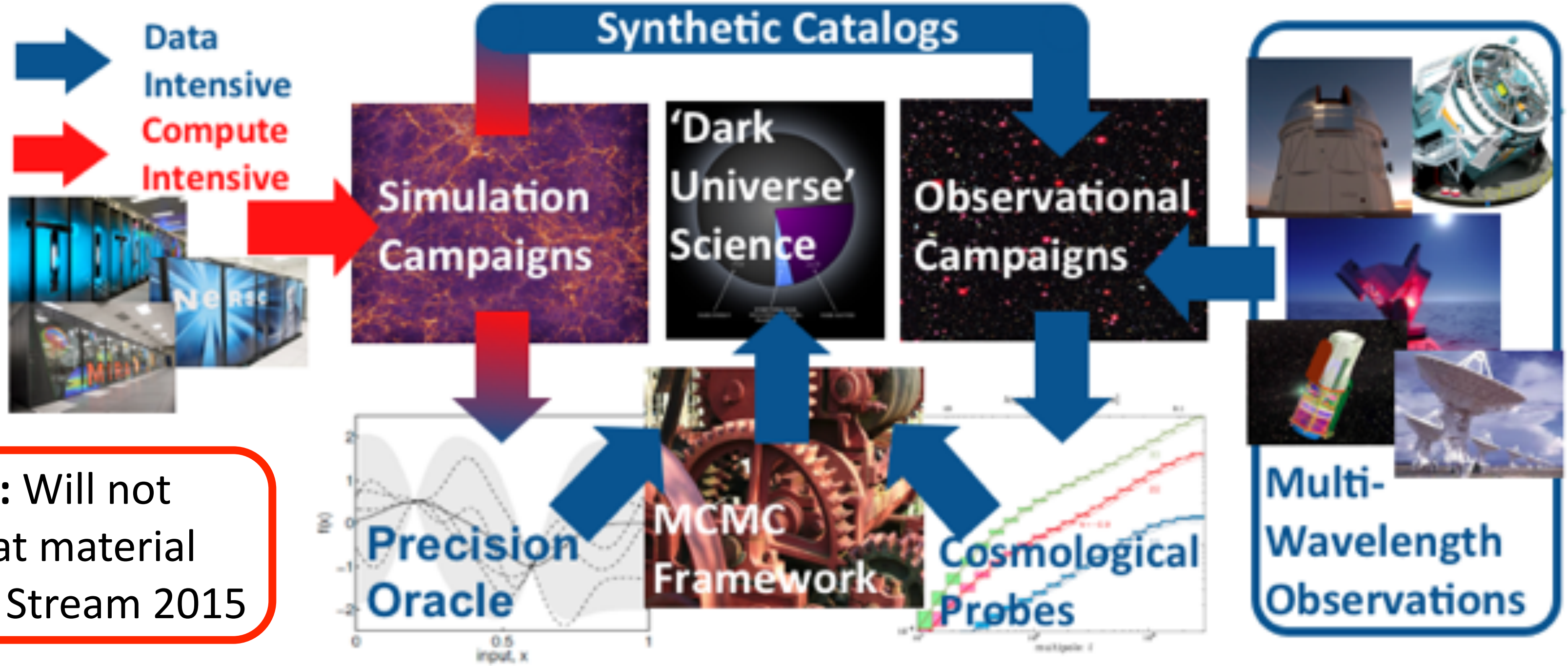








Data Flows in Cosmology: The Big Picture

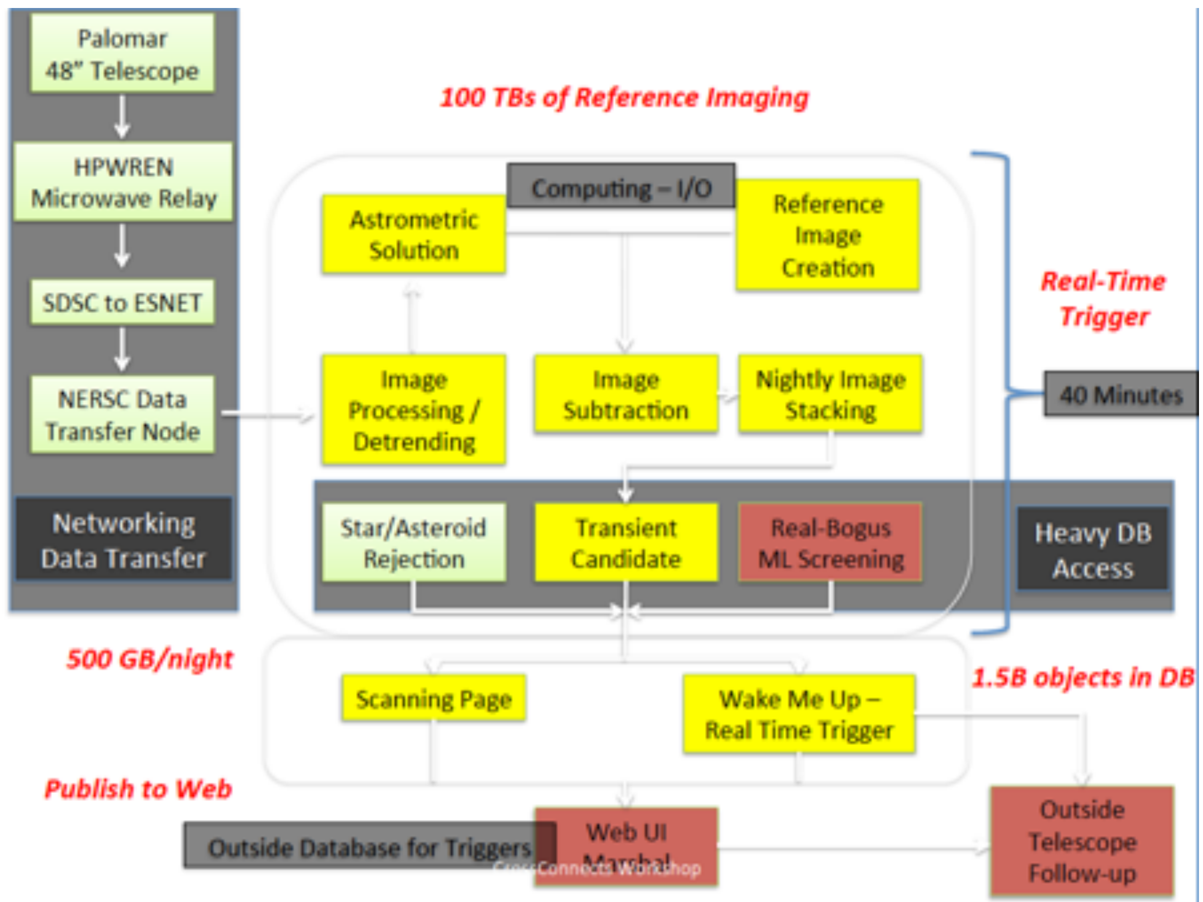


Note: Will not repeat material from Stream 2015

- **Data/Computation in Cosmology:** Data flows and associated analytics play an essential role in cosmology (combination of streaming and offline analyses)
- **Streaming Data:**
  - **Observations:** CMB experiments (ACT, SPT, —), optical transients (Sn surveys, GW follow-ups, —), radio surveys
  - **Simulations:** Large datastreams (in situ and co-scheduled data transformation)
  - **Analytics:** Transient classification pipelines, imaging pipelines

Data Flow Example

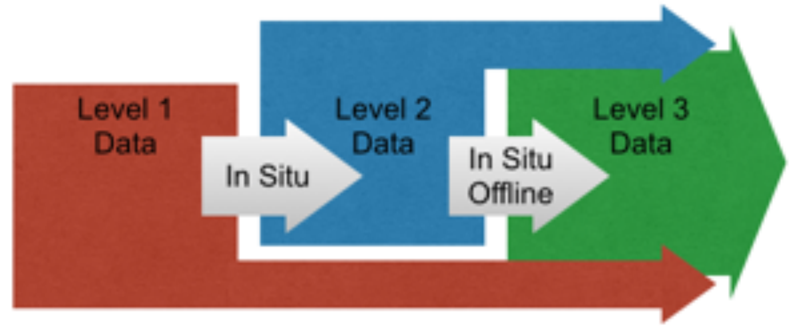
- **Transient Surveys:** Optical searches for transients (e.g., DES, LSST, PTF) can have cadences in the range of fractions of minutes to minutes, current data rates are about 500 GB/ night — LSST can go up to 20TB/night, about 10K alerts/night
- **Machine Learning:** Major opportunity for machine learning for filtering and classification of transient sources (potentially one in a million interesting events) demonstrated at NERSC with PTF

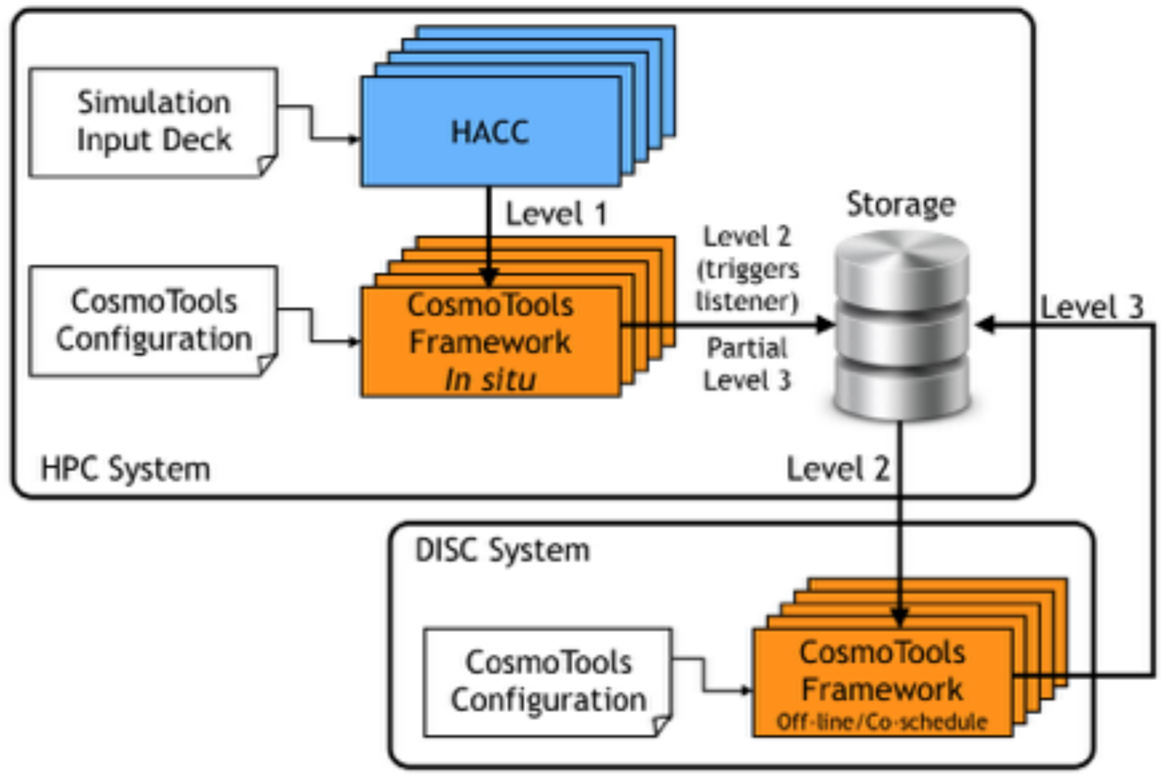


Palomar Transient Factory (courtesy Peter Nugent)

In Situ Analysis and Co-Scheduling

- **Analysis Dataflows:** Analysis data flows are complex and any future strategy must combine elements of in situ and offline approaches (Flops vs. IO/ storage imbalance)
- **CosmoTools Test:** Test of coordinated offline analysis ("co-scheduling")
- **Portability:** Analysis routines implemented using PISTON (part of VTK-m, built on NVIDIA's Thrust library)
- **Example Case (Titan):** Large halo analysis (strong scaling bottleneck) offloaded to alternative resource using a listener script that looks for appropriate output files

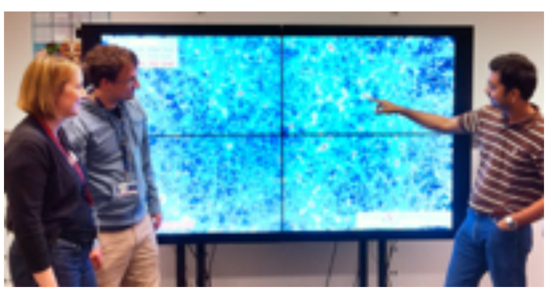


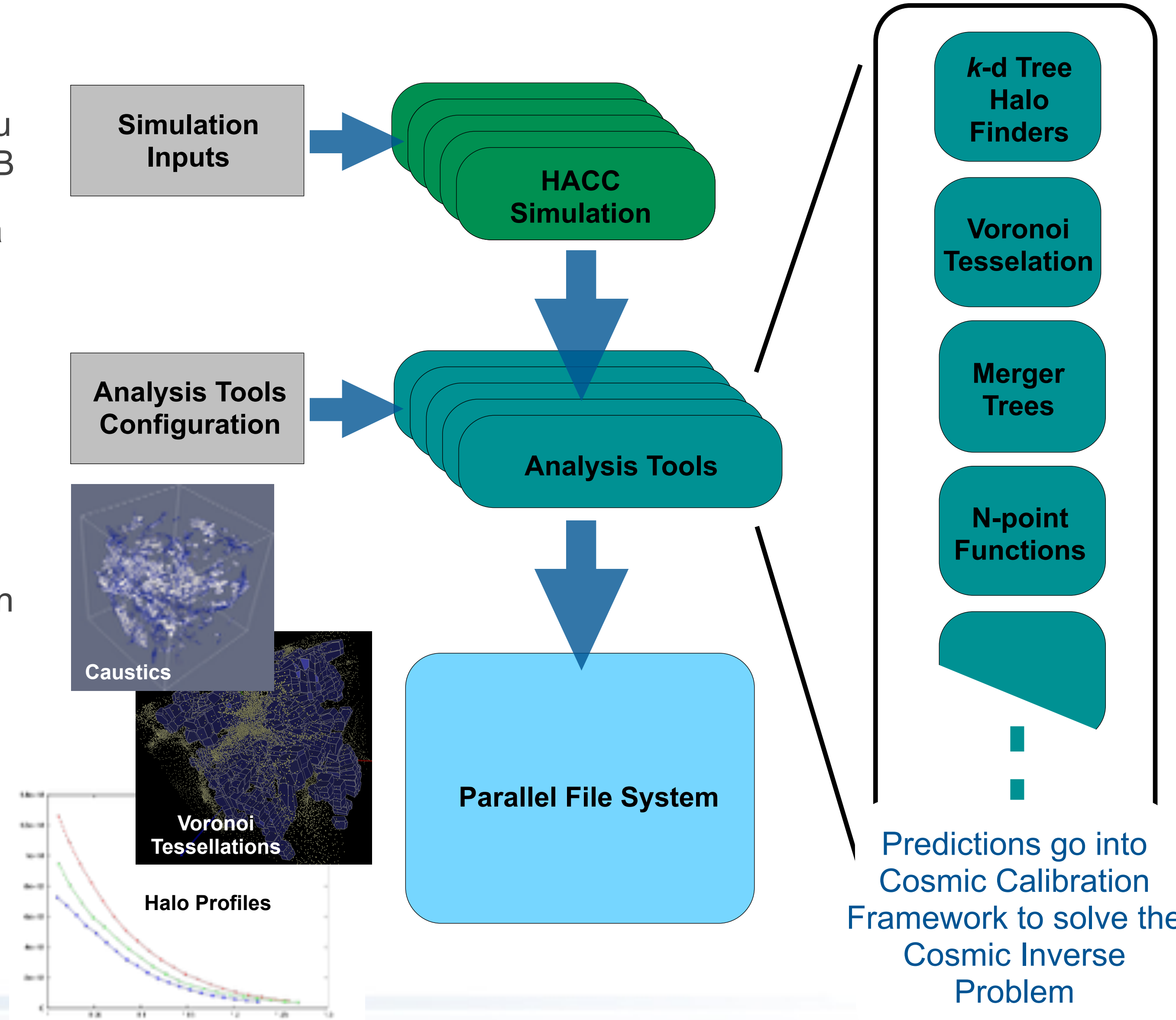


Sewell et al. 2015, SC15 Technical Paper

In Situ Analysis Example

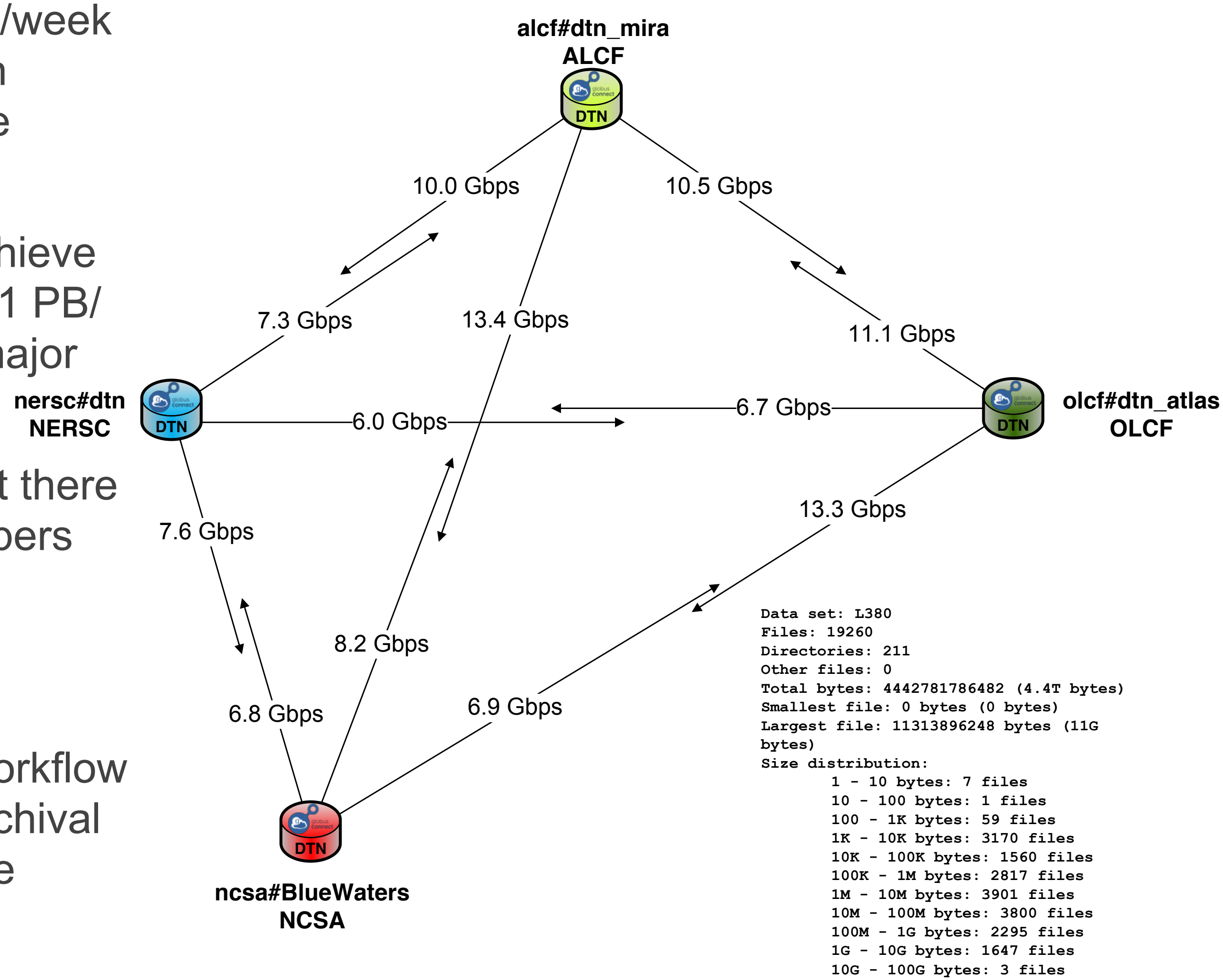
- **Data Reduction:** A trillion particle simulation with 100 analysis steps has a storage requirement of ~4 PB -- in situ analysis reduces it to ~200 TB
- **I/O Chokepoints:** Large data analyses difficult because I/O time > analysis time, plus scheduling overhead
- **Fast Algorithms:** Analysis time is only a fraction of a full simulation timestep
- **Ease of Workflow:** Large analyses difficult to manage in post-processing





Offline Data Flow: Large-Scale Data Movement

- **Offline Data Flows:** Cosmological simulation data flows already require ~PB/week capability, next-generation streaming data will require similar bandwidth
- **ESnet Project:** Aim to achieve a production capability of 1 PB/ week (FS to FS) across major compute sites
- **Status:** Very close but not there yet (600+ TB/week); numbers from a simulation dataset "package" (4 TB)
- **Future:** Automate entire process within the data workflow including retrieval from archival storage (HPSS); add more compute/data hubs



Petascale DTN project, courtesy Eli Dart

Extreme-Scale Analytics Systems (EASy) Project (ASCR/HEP)

- **New Approaches to Large-Scale Data Analytics:** Combine aspects of High Performance Computing, Data-Intensive Computing, and High Throughput Computing to develop new pathways for large-scale scientific analyses enabled through Science Portals
- **EASy Elements (Initial focus on cosmological simulations and surveys):**
  - **Surveys:** DESI, LSST, SPT, —
  - **Software Stack:** Run complex software stacks on demand (containers and virtual machines)
  - **Resilience:** Handle job stream failures and restarts
  - **Resource Flexibility:** Run complex workflows with dynamic resource requirements
  - **Wide-Area Data Awareness:** Seamlessly move computing to data and vice versa; access to remote databases and data consistency
  - **Automated Workloads:** Run automated production workflows
  - **End-to-End Simulation-Based Analyses:** Run analysis workflows on simulations and data using a combination of in situ and offline/co-scheduling approaches

EASy Project: Infrastructure Components

Component	Description	Notes
Observational Data	Data from Dark Energy Survey (DES), Sloan Digital Sky Survey (SDSS), South Pole Telescope (SPT), and upcoming surveys (DESI, LSST, WFIRST, —)	Make selected data subsets available given storage limits; make analysis software available to analyze the datasets
Simulation Data	Simulations for optical surveys (raw data, object catalogs, synthetic catalogs, predictions for observables); simulations for CMB observations	Very large amounts of simulation data need to be made available; hierarchical data views; data compression methods
Data Storage	Multi-layered storage on NVRAM, spinning disk, and disk-fronted tape (technologies include RAM disk, HPSS, parallel file systems)	Current storage availability for the project is ~PB on spinning disk; larger resources available within HPSS; RAM disk testbeds
Data Transfer	Data transfer synced with computational infrastructure and resources; data transfer as integral component of data-intensive workflows	Use of Globus transfer as an agreed mechanism; current separate project with ESnet to have a production capability at 1PB/week
Computational Infrastructure	Wide range of computational resources include high performance computing, high throughput computing, and data-intensive computing platforms	How to bring together a number of distinct resources to solve analysis tasks in a layered fashion? What is the optimal mix?
Computational Resources	Resources at NERSC include Edison and Cori Phase 1; at Argonne, Cooley, Jupiter/Magellan, Theta (future)	Melding HPC and cluster resources; testbeds for using HPC resources for data-intensive tasks and elastic computing paradigms
Containers and Virtualization	Running large-scale workflows with complex software stacks; allowing for interactive as well as batch modes for running jobs; use of web portals	Data management and analysis workflows, especially workflows that combine simulation and observational datastreams
Algorithmic Advances	New data-intensive algorithms with improved scaling properties, including approximate algorithms with error bounds; new statistical methods	As data volumes increase rapidly, new algorithms are needed to produce results in finite time, especially for interactive appliications

Future Challenges

- **Data Filtering and Classification:** The major challenges for machine learning approaches are high levels of throughput and lack of training datasets — these approaches are the only ones that are likely to succeed, however
- **Data Access:** View of streaming as "one-shot" is actually a statement of a technology limitation; to overcome this will require cheap and fast storage with databases (or equivalent) overlays
- **Software Management:** Current data pipelines can be very complex (although not very computationally intensive) with many software interdependencies — work using VMs and containers shows substantial promise
- **Resource Management:** Cloud resources have attractive features, such as on-demand allocation — can enterprise-level science requirements for high-throughput data analytics be met by the cloud?