RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

Big Data Challenges in Simulation-based Science

Manish Parashar*

Rutgers Discovery Informatics Institute (RDI^2)
Department of Computer Science

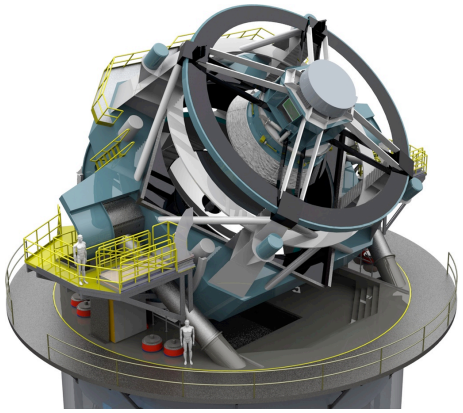
***Hoang Bui, Tong Jin, Qian Sun, Fan Zhang, ADIOS Team, and other ex-students and collaborators….**

Outline

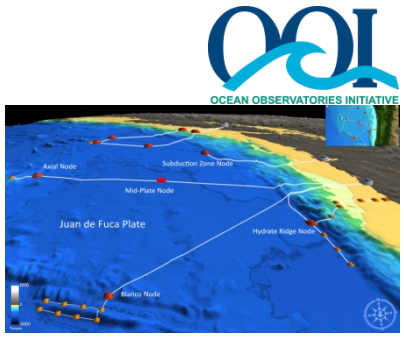
- **Data Grand Challenges**
  - Data challenges of simulation-based science
- Rethinking the simulations -> insights pipeline
- The DataSpaces Project
- Conclusion

Data-driven Discovery in Science

Nearly every field of discovery is transitioning from “data poor” to “data rich”

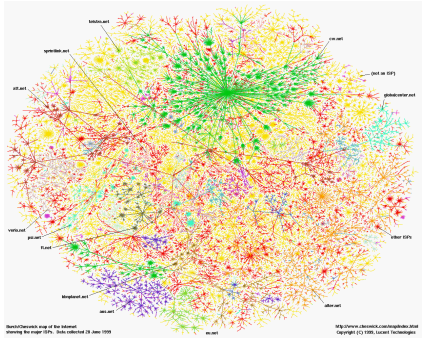


Astronomy: LSST



Oceanography: OOI

Physics: LHC



Sociology: The Web

Biology: Sequencing

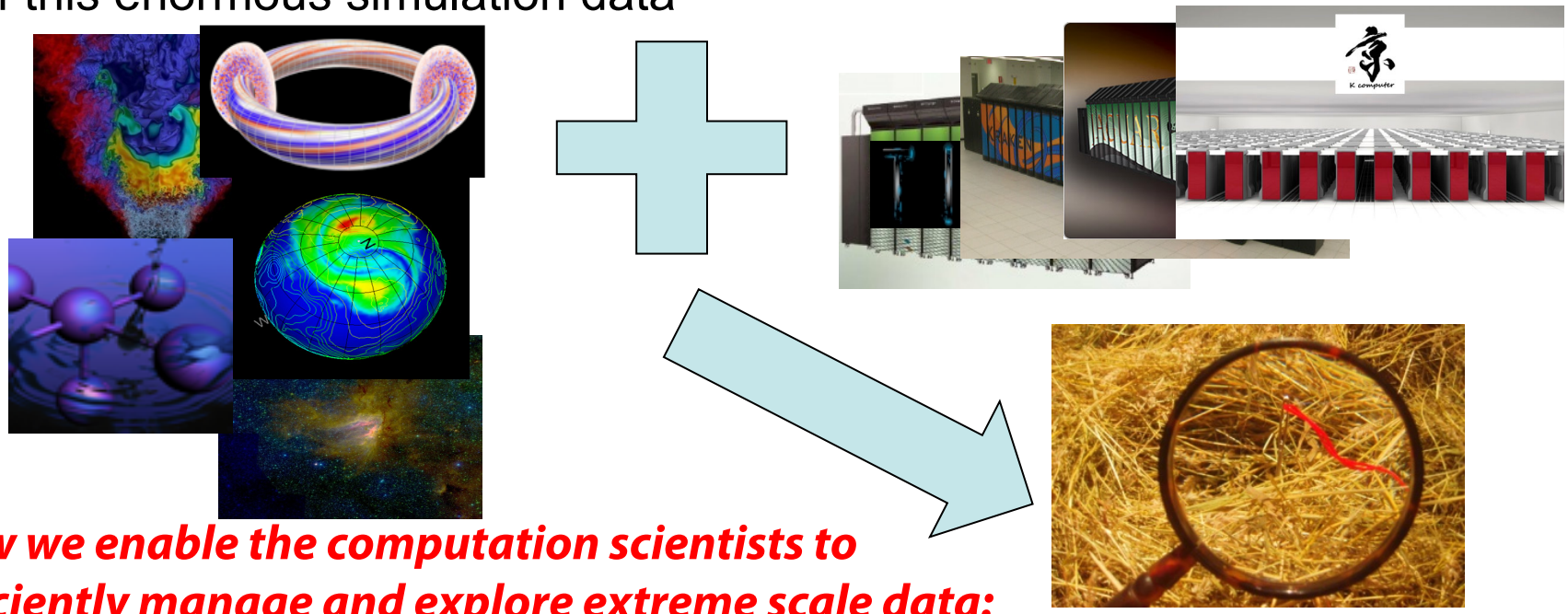
Economics: POS terminals

Neuroscience: EEG, fMRI

Credit: Ed Lazowska, Univ. of WA

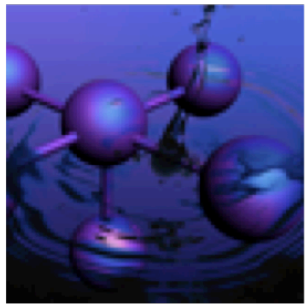
Scientific Discovery through Simulations

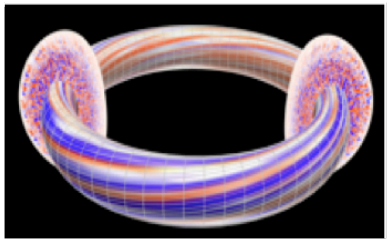
- Scientific simulations running on high-end computing systems generate huge amounts of data!
  - If a single core produces 2MB/minute on average, one of these machines could generate simulation data between ~**170TB** per hour -> ~**700PB** per day -> ~**1.4EB** per year
- Successful scientific discovery depends on a comprehensive understanding of this enormous simulation data



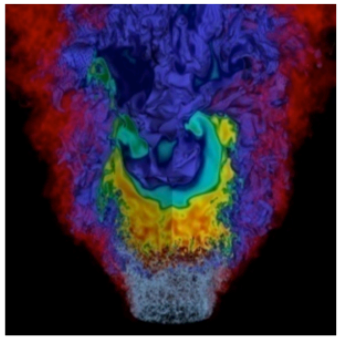
How we enable the computation scientists to efficiently manage and explore extreme scale data: "find the needles in haystack" ??

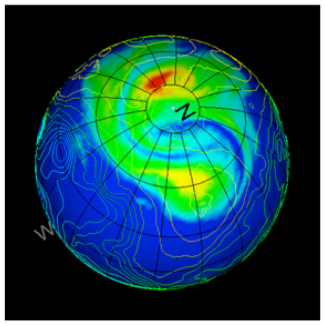
Scientific Discovery through Simulations


Molecular Simulation


Plasma Fusion


Astrophysics


Combustion


Climate Modeling

- Complex workflows integrating coupled models, data management/ processing, analytics
  - Tight / loose coupling, data driven, ensembles
- Advanced numerical methods (E.g., Adaptive Mesh Refinement)
- Integrated (online) analytics, uncertainty quantification, …

- Complex, heterogeneous components
- Large data volumes and data rates
- Data re-distribution (MxNxP), data transformations
- Dynamic data exchange patterns
- Strict performance/overhead constraints

Traditional *Simulation -> Insight* Pipelines Break Down

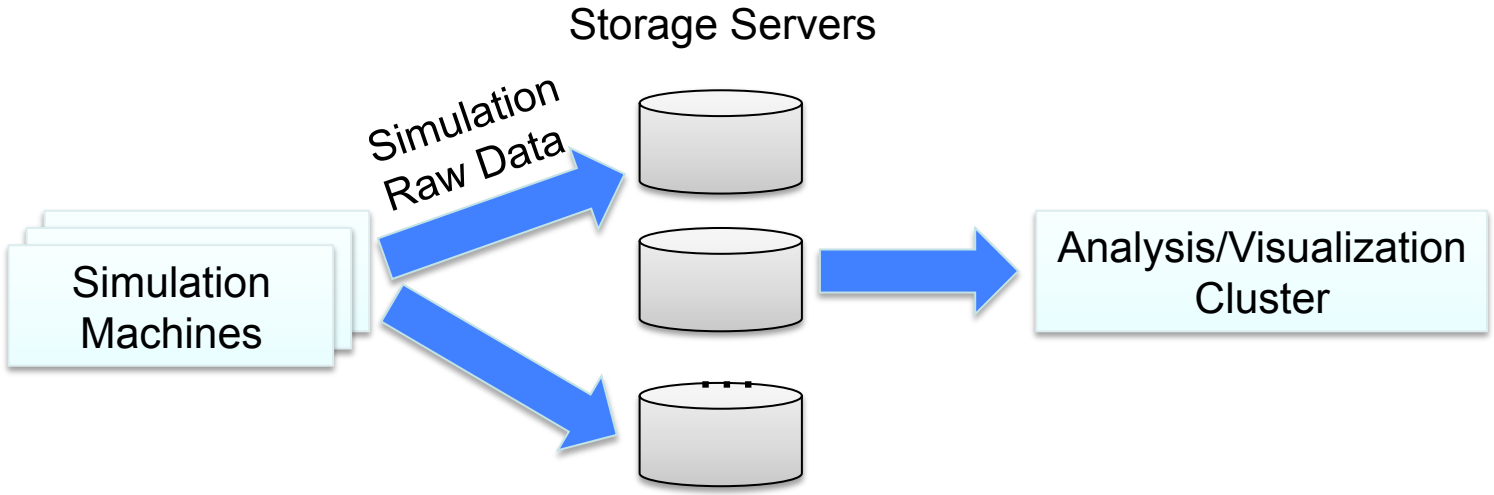


Figure. Traditional data analysis pipeline

- Traditional ***simulation -> insight*** pipeline
  - Run large-scale simulation workflows on large supercomputers
  - Dump data to parallel disk systems
  - Export data to archives
  - Move data to users' sites – usually selected subsets
  - Perform data manipulations and analysis on mid-size clusters
  - Collect experimental / observational data
  - Move to analysis sites
  - Perform comparison of experimental/observational to validate simulation data

Challenges Faced by Traditional HPC Data Pipelines

- **Data analysis challenge**
  - Can current data mining, manipulation and visualization algorithms still work effectively on extreme scale machine?
- **I/O and storage challenge**
  - Increasing performance gap: disks are outpaced by computing speed
- **Data movement challenge**
  - Lots of data movement between simulation and analysis machines, between coupled multi-physics simulation components -> longer latencies
  - Improving data locality is critical: do work where the data resides!
- **Energy challenge**
  - Future extreme systems are designed to have low-power chips – however, much greater power consumption will be due to memory and data movement!

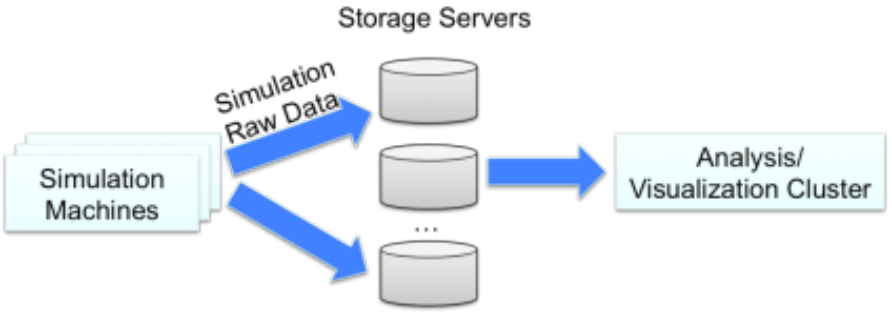
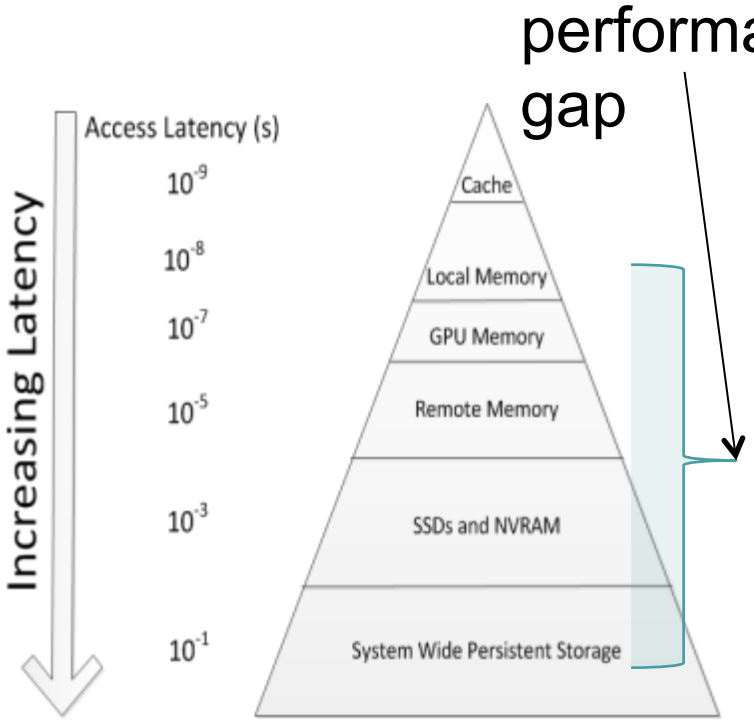


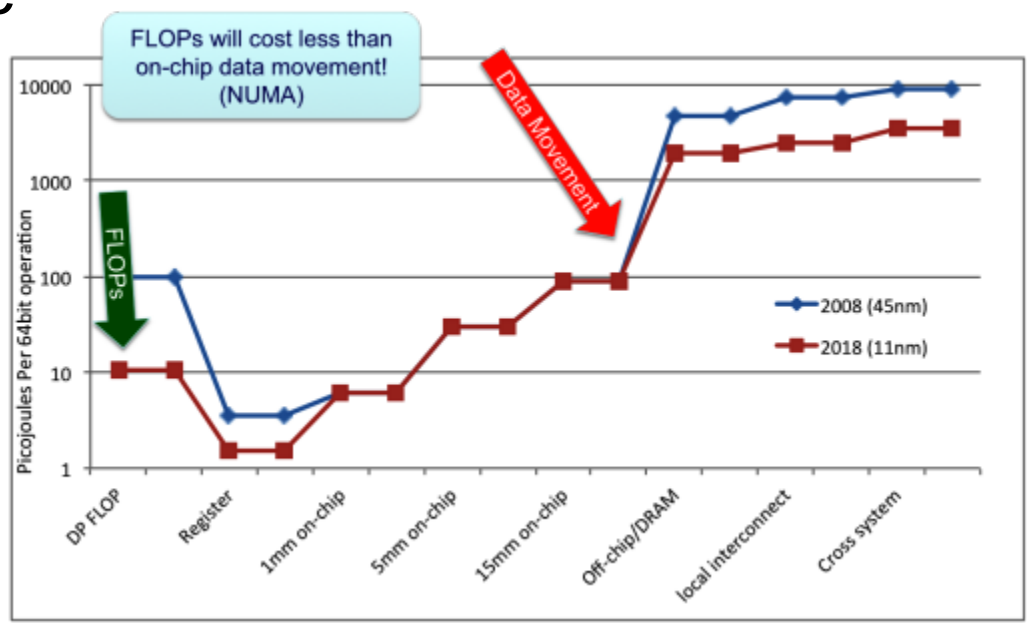
Figure. Traditional data analysis pipeline

*The **costs of data movement** are increasing and dominating!*

The Cost of Data Movement

- Moving data between node memory and persistent storage is slow!
- The energy cost of moving data is a significant concern





$$\text{Energy\_move\_data} = \frac{\text{bitrate} * \text{length}^2}{\text{cross\_section\_area\_of\_wire}}$$

From K. Yelick, "Software and Algorithms for Exascale: Ten Ways to Waste an Exascale Computer"

Rethinking the Data Management Pipeline – Hybrid Staging + In-Situ & In-Transit Execution

Issues/Challenges

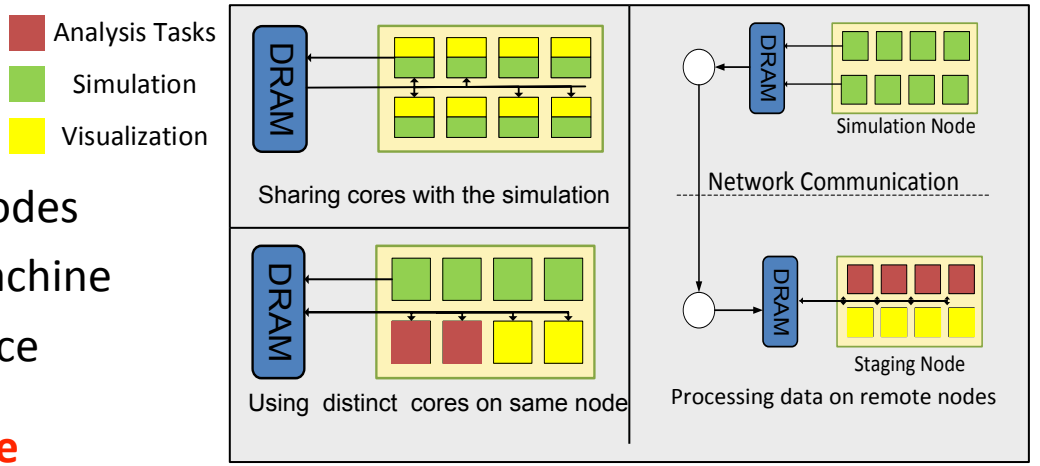
- Programming abstractions/systems
- Mapping and scheduling
- Control and data flow
- Autonomic runtime
- Link with the external workflow

Systems

- Glean, Darshan, FlexPath, …..

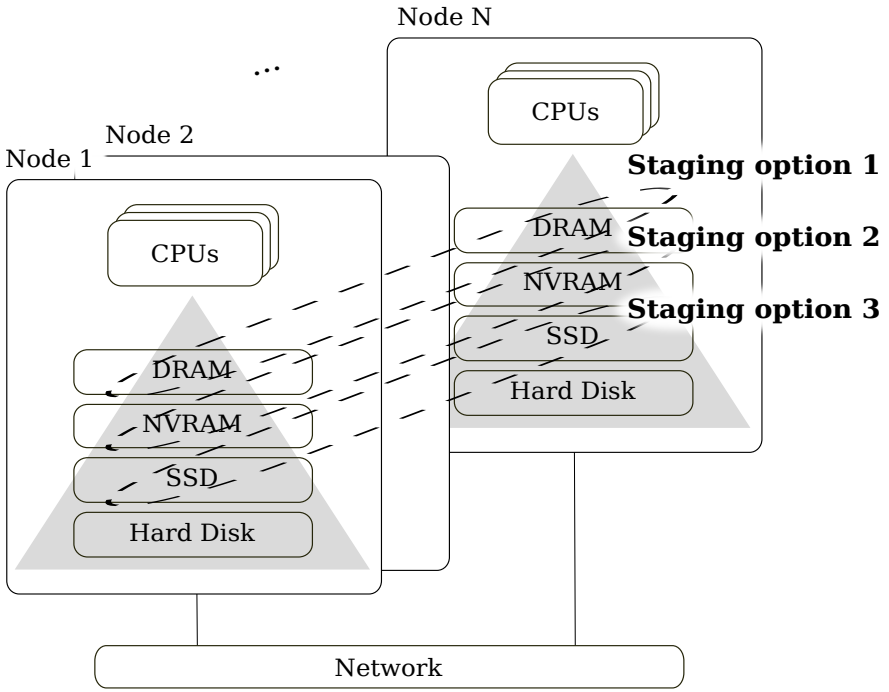
Design space of possible workflow architectures

- **Location of the compute resources**
  - Same cores as the simulation (in situ)
  - Some (dedicated) cores on the same nodes
  - Some dedicated nodes on the same machine
  - Dedicated nodes on an external resource



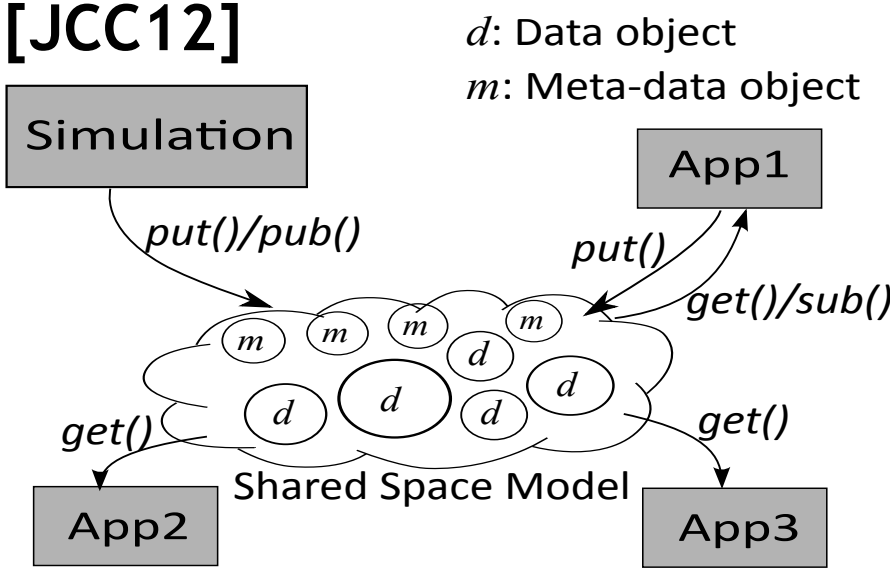
- **Data access, placement, and persistence**
  - Direct access to simulation data structures
  - Shared memory access via hand-off / copy
  - Shared memory access via non-volatile near node storage (NVRAM)
  - Data transfer to dedicated nodes or external resources

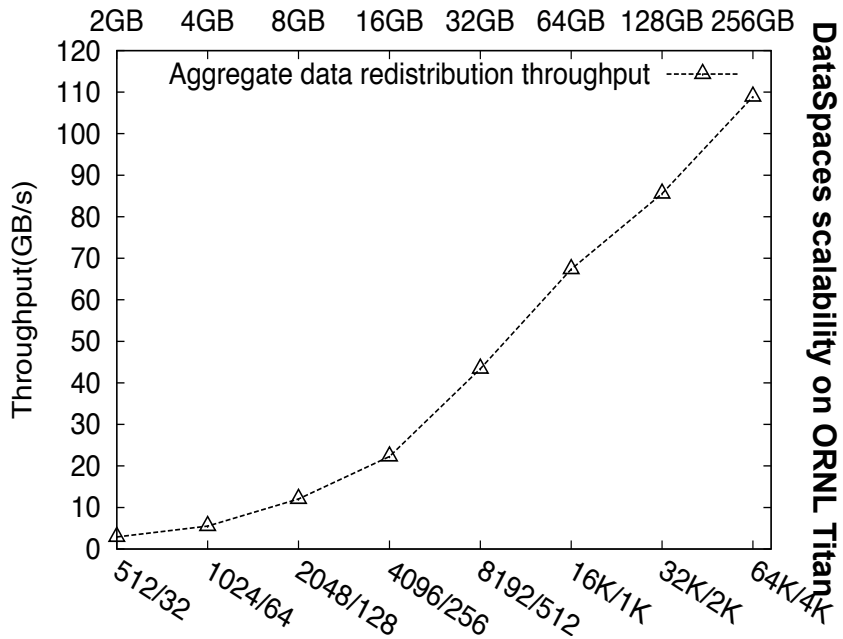
- **Synchronization and scheduling**
  - Execute synchronously with simulation every n^{th} simulation time step
  - Execute asynchronously



DataSpaces: A Scalable Shared Space Abstraction for Hybrid Data Staging [JCC12]

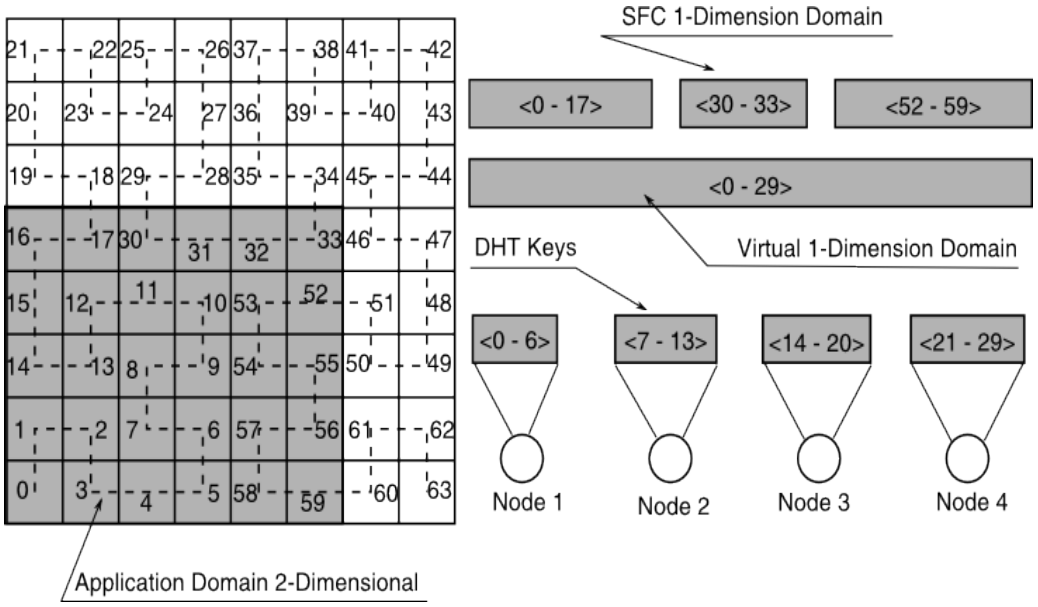
- Shared-space programming abstraction over hybrid staging
  - Simple API for coordination, interaction and messaging
  - Provides a global-view programming abstraction
  - Distributed, associative, in-deep-memory object store
  - Online data indexing, flexible querying
- Exposed as a persistent service
- Autonomic cross-layer runtime management
- High-throughput/low-latency asynchronous data transport





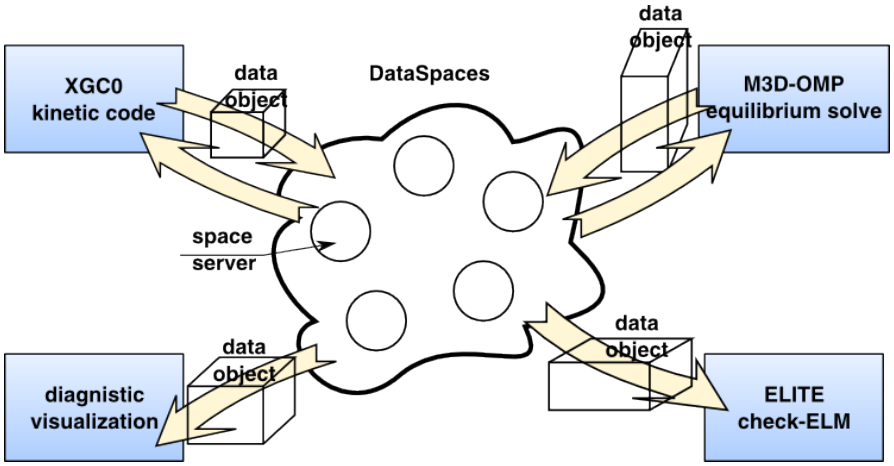
DataSpaces Abstraction: Indexing + DHT [HPDC10]

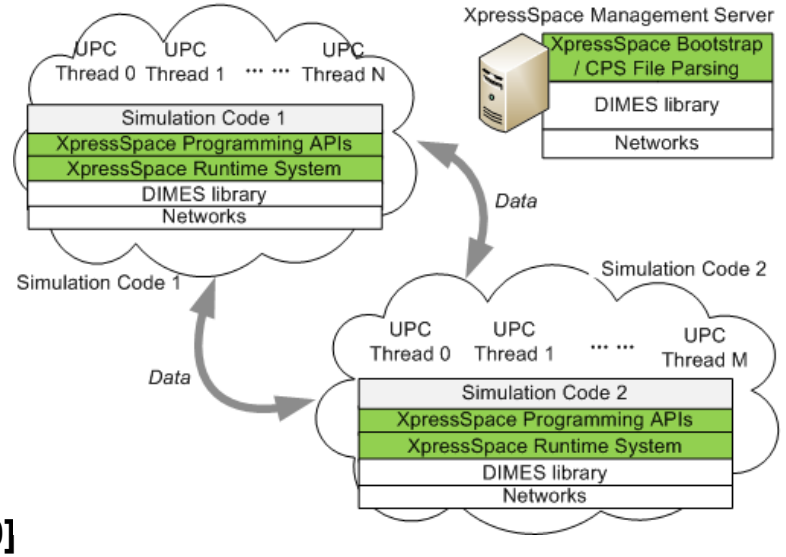
- Dynamically constructed structured overlay using hybrid staging cores
- Index constructed online using SFC mappings and applications attributes
  - E.g., application domain, data, field values of interest, etc.
- DHT used to maintain meta-data information
  - E.g., geometric descriptors for the shared data, FastBit indices, etc.
- Data objects load-balanced separately across staging cores



RUTGERS

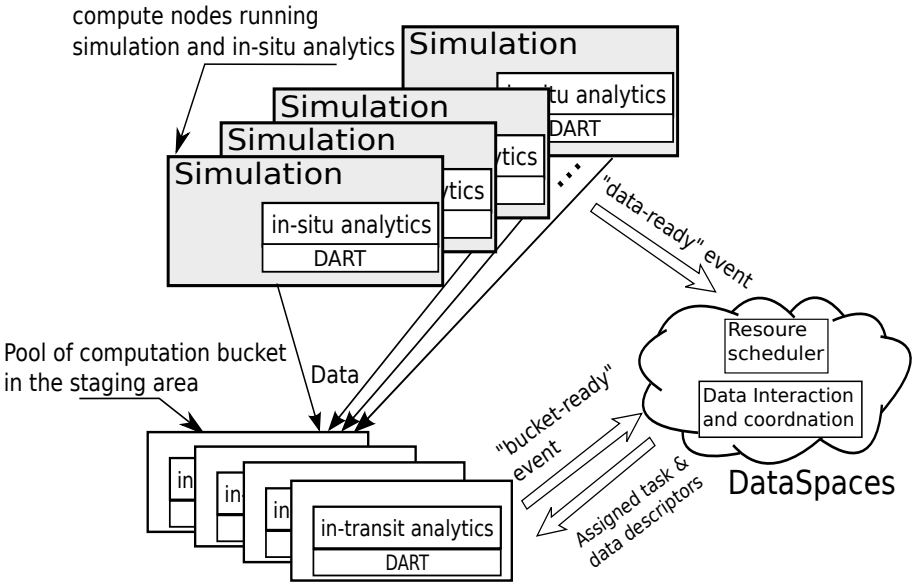
DataSpaces: Enabling Coupled Scientific Workflows at Extreme Scales

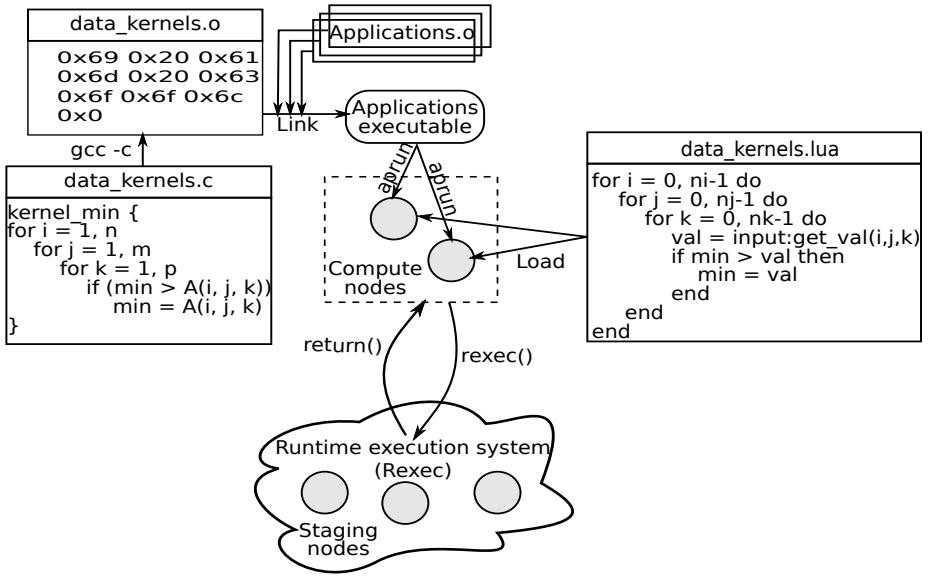




Multphysics Code Coupling at Extreme Scales [CCGrid10]

PGAS Extensions for Code Coupling [CCPE13]



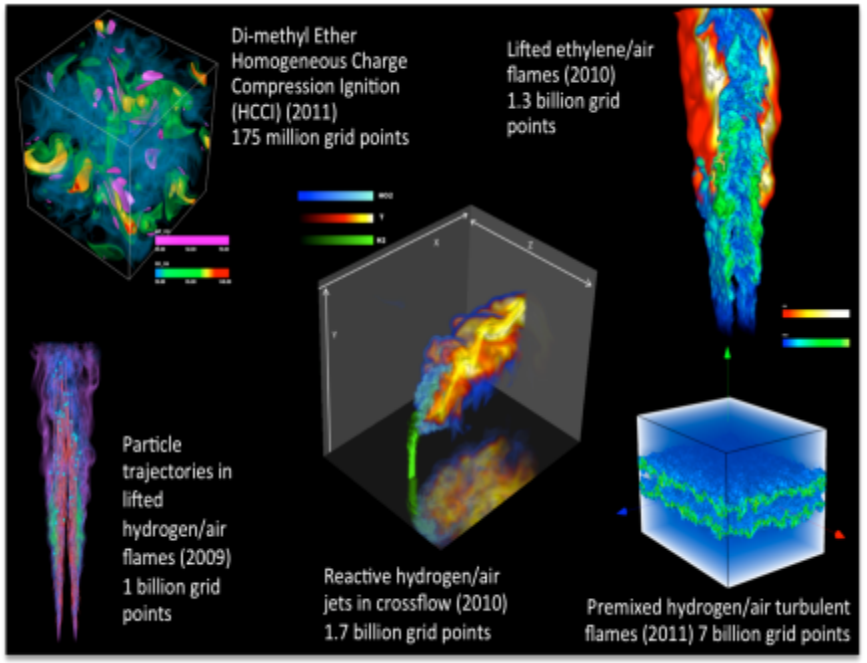


Data-centric Mappings for In-Situ Workflows [IPDPS12] **Dynamic Code Deployment In-Staging [IPDPS11]**

Integrating In-situ and In-transit Analytics [SC'12]

EXACT CENTER FOR EXASCALE SIMULATION OF COMBUSTION IN TURBULENCE

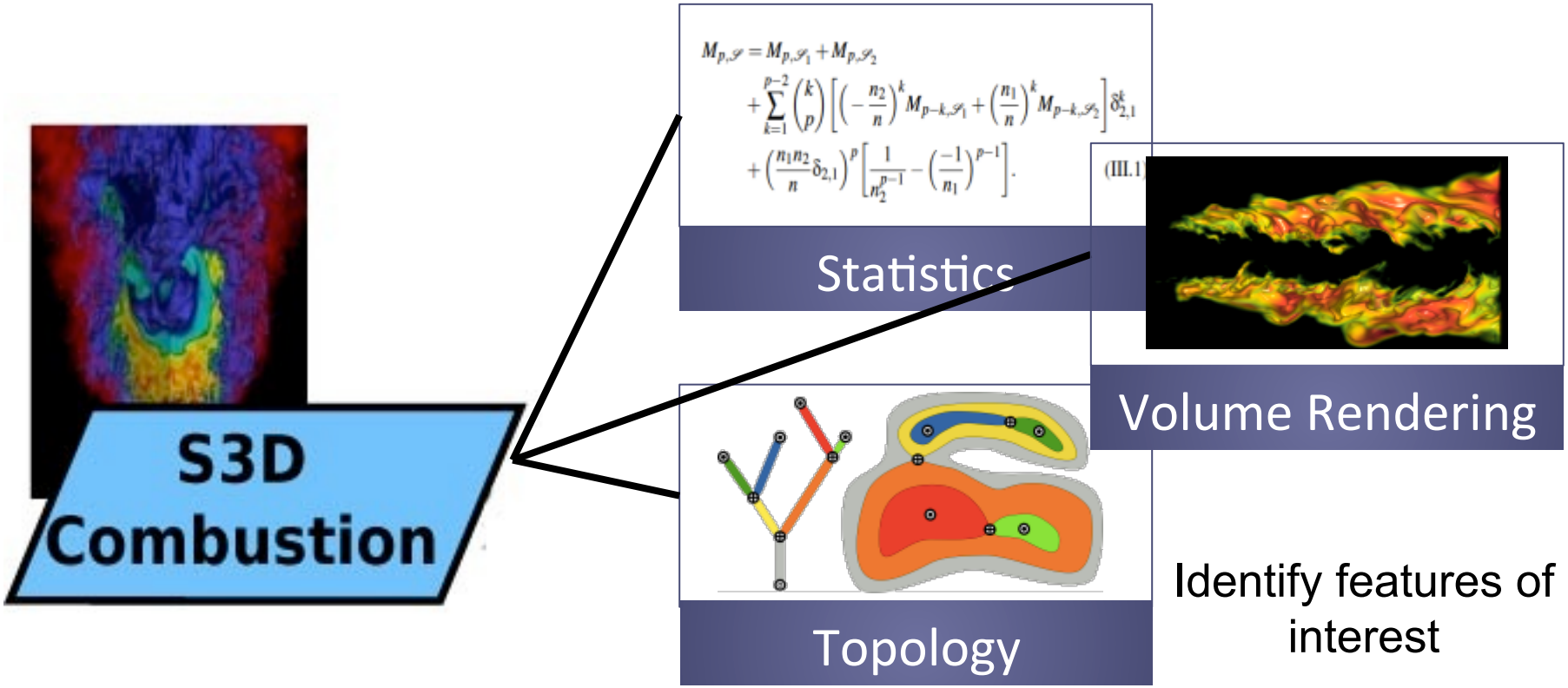
- S3D: First-principles direct numerical simulation
- Simulation resolves features on the order of 10 simulation time steps
- Currently on the order of every 400th time step can be written to disk
- Temporal fidelity is compromised when analysis is done as a post-process



Recent data sets generated by S3D, developed at the Combustion Research Facility, Sandia National Laboratories

Integrating In-situ and In-transit Analytics - Design

- Couple concurrent data analysis with simulation



*J. C. Bennett et al., *"Combining In-Situ and In-Transit Processing to Enable Extreme-Scale Scientific Analysis"*, SC'12, Salt Lake City, Utah, November, 2012.

In-situ/In-transit Data Analytics

- Online data analytics for large-scale simulation workflows
  - Combine both in-situ and in-transit placement to reduce impact on simulation, reduce network data movement, reduce total execution time
  - Adaptive runtime management: Optimize the performance of simulation-analysis workflow through adaptive data and analysis placement, data-centric task mapping

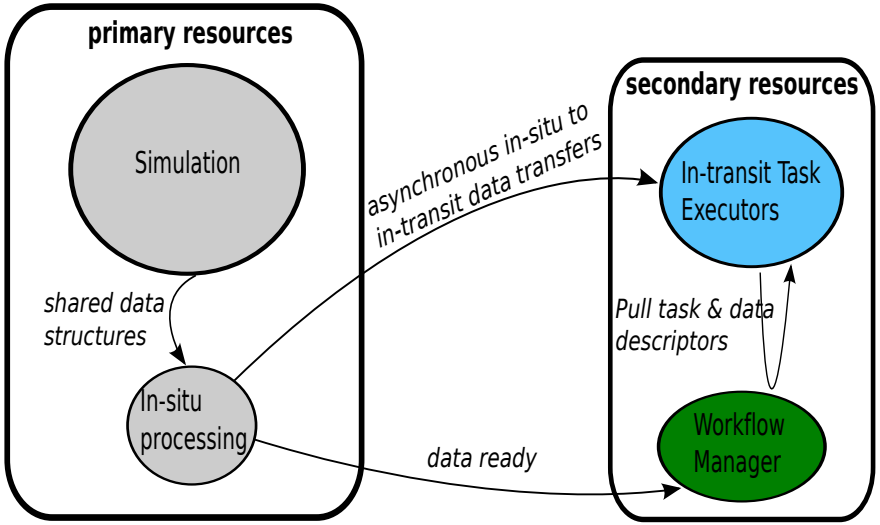


Figure. Overview of the in-situ/in-transit data analysis framework

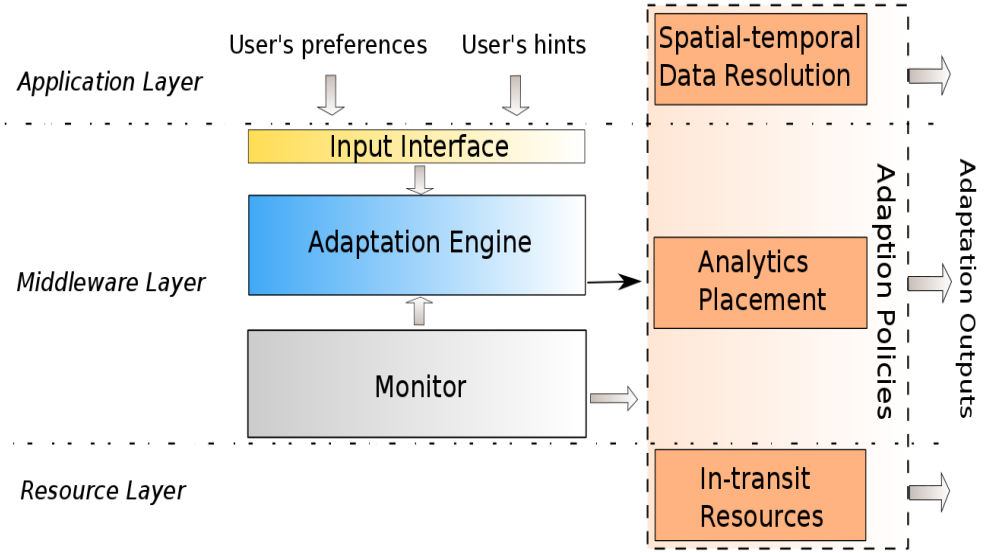
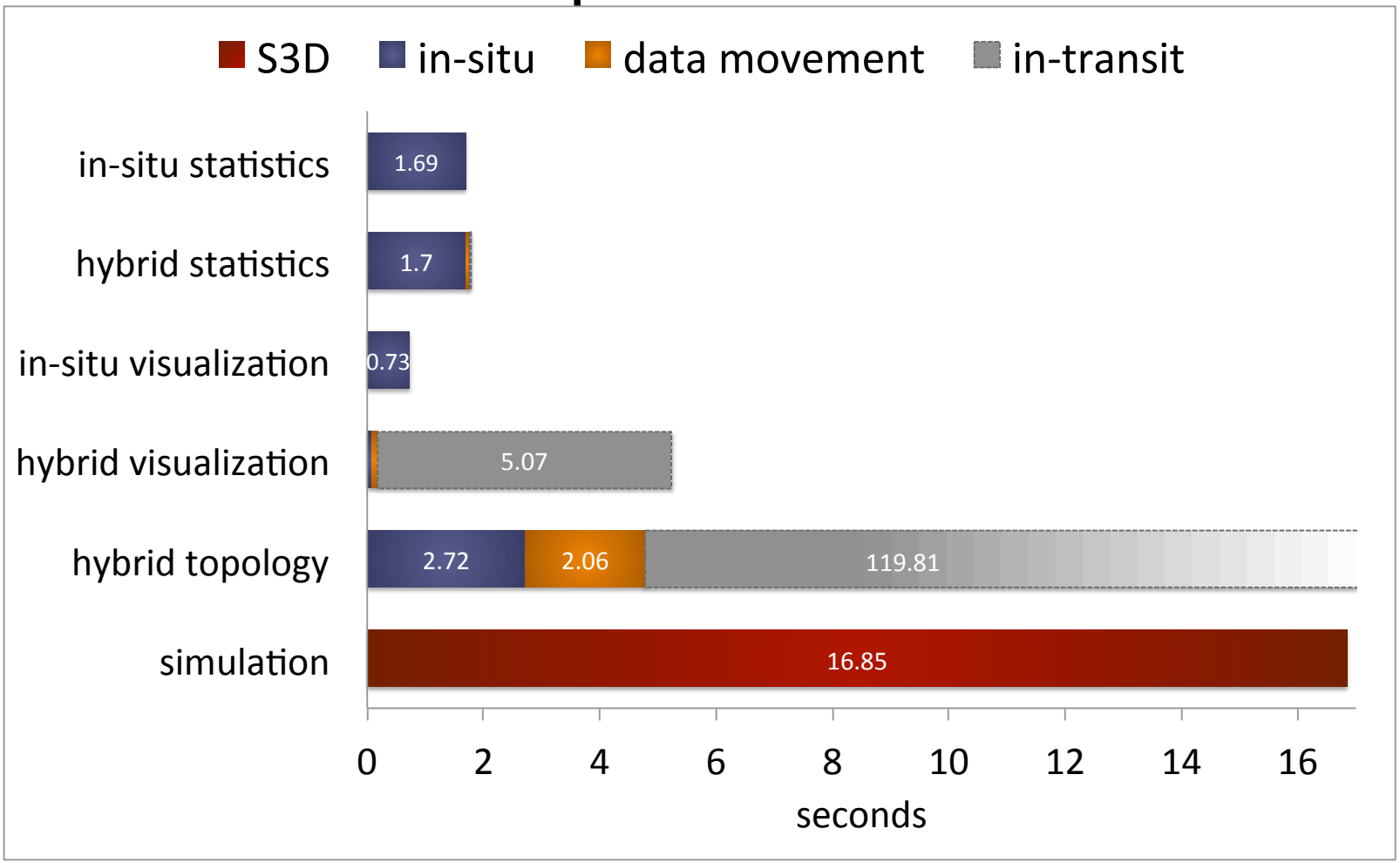
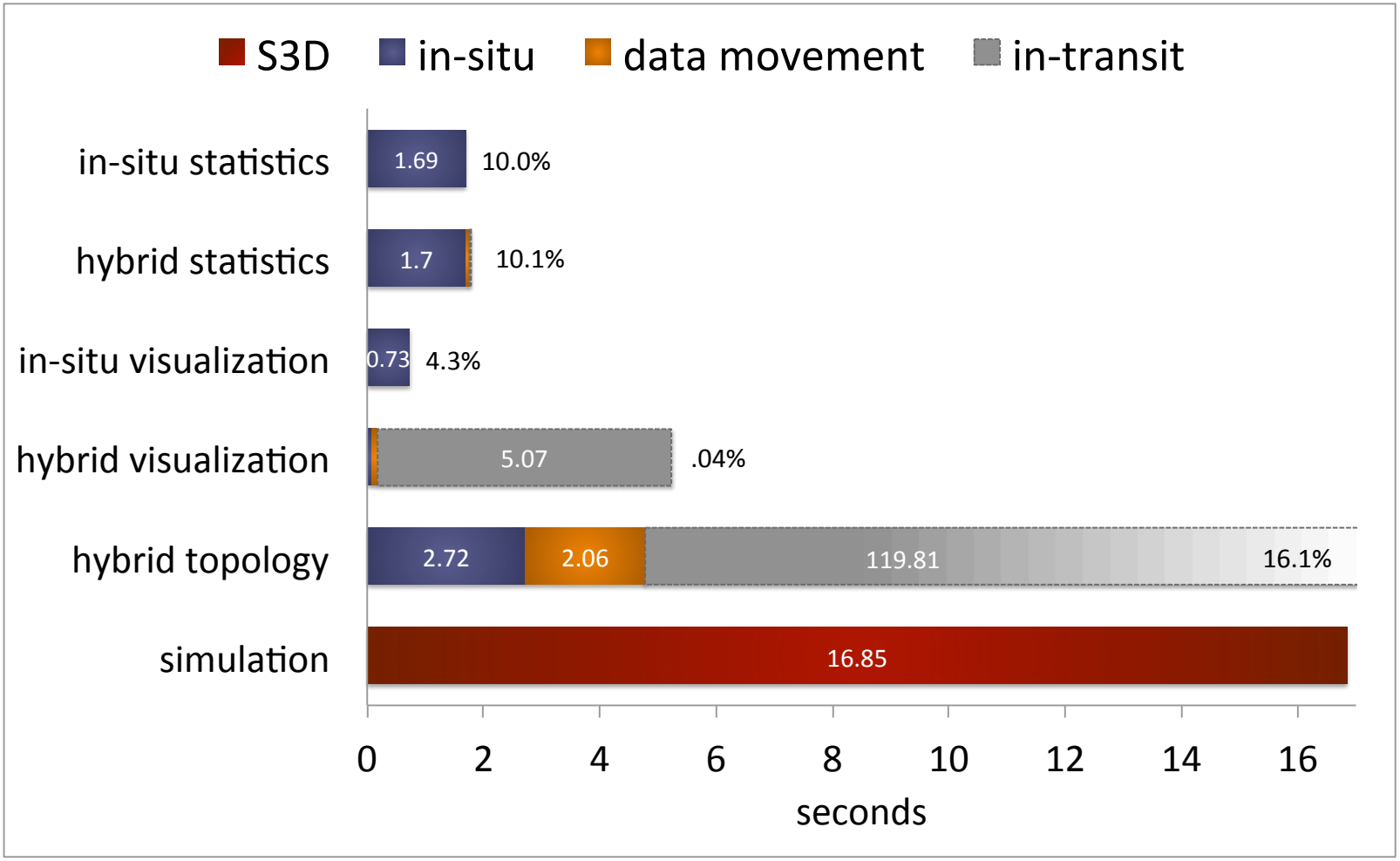


Figure. System architecture for cross-layer runtime adaptation

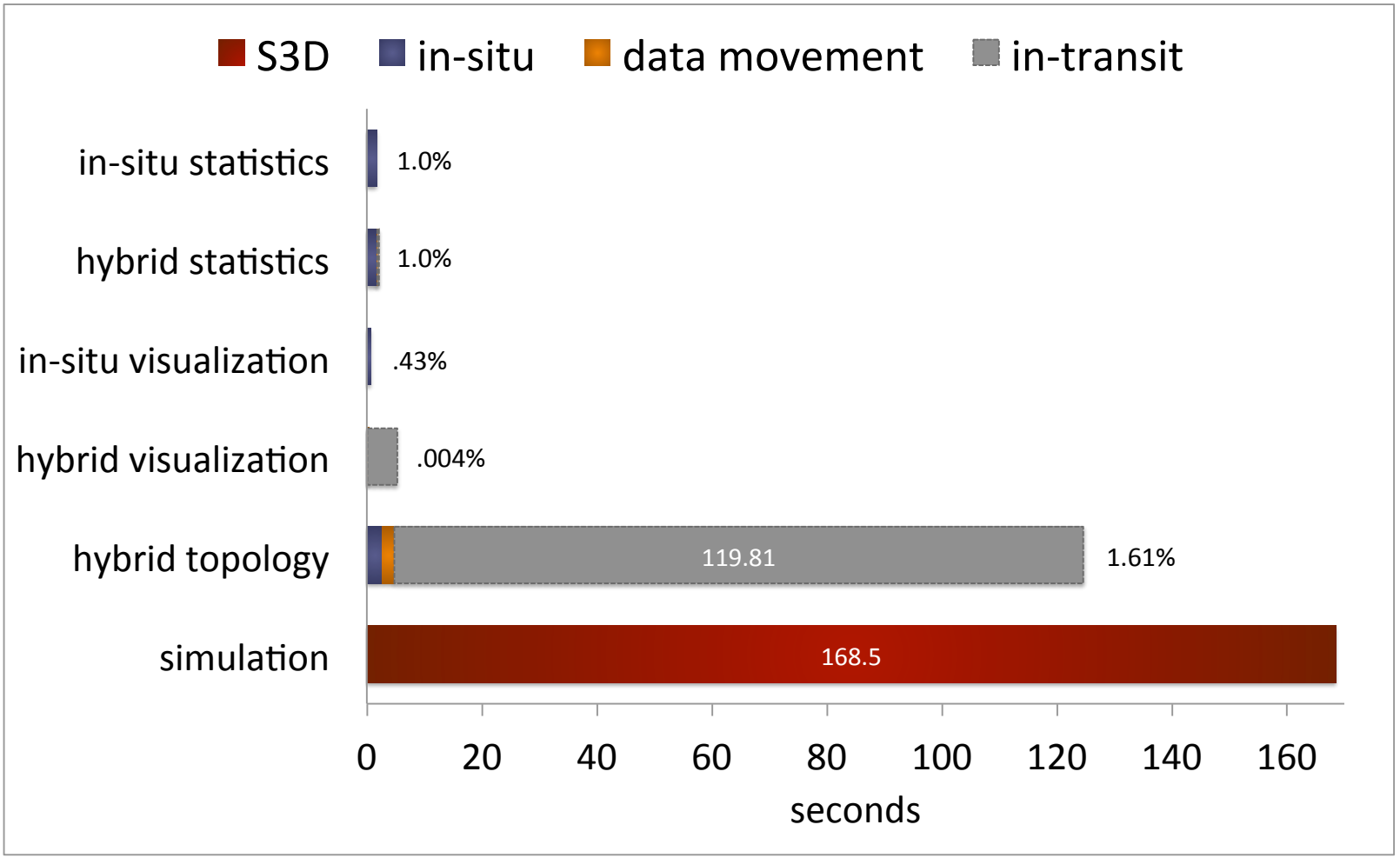
Simulation case study with S3D: Timing results for 4896 cores and analysis every simulation time step



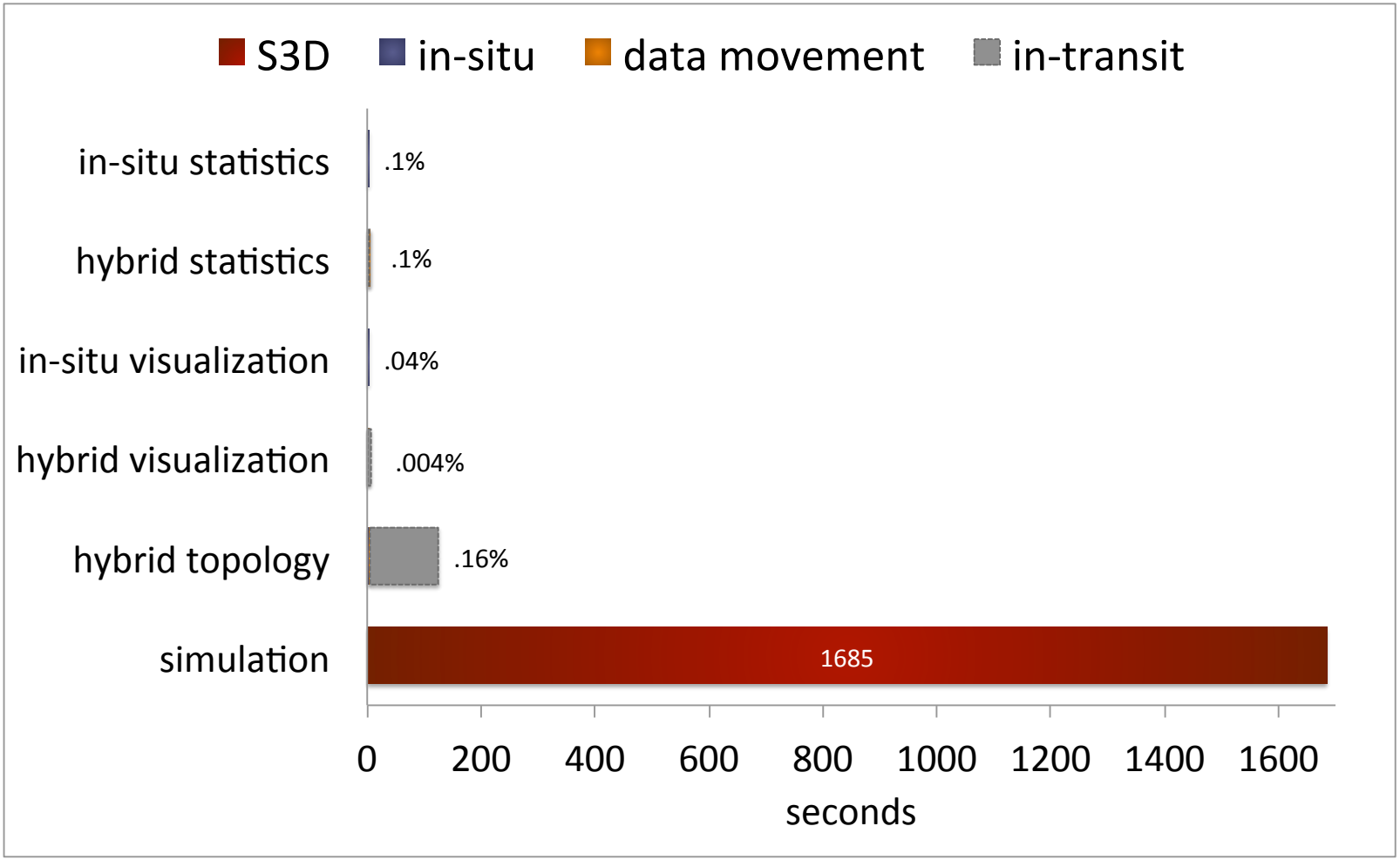
Simulation case study with S3D: Timing results for 4896 cores and analysis every simulation time step



Simulation case study with S3D: Timing results for 4896 cores and analysis every 10th simulation time step

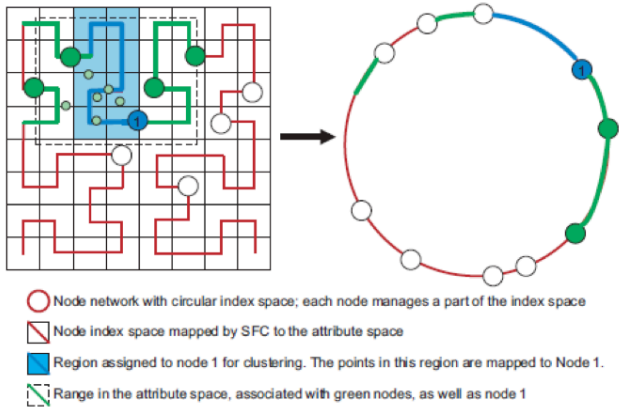


Simulation case study with S3D: Timing results for 4896 cores and analysis every 100th simulation time step

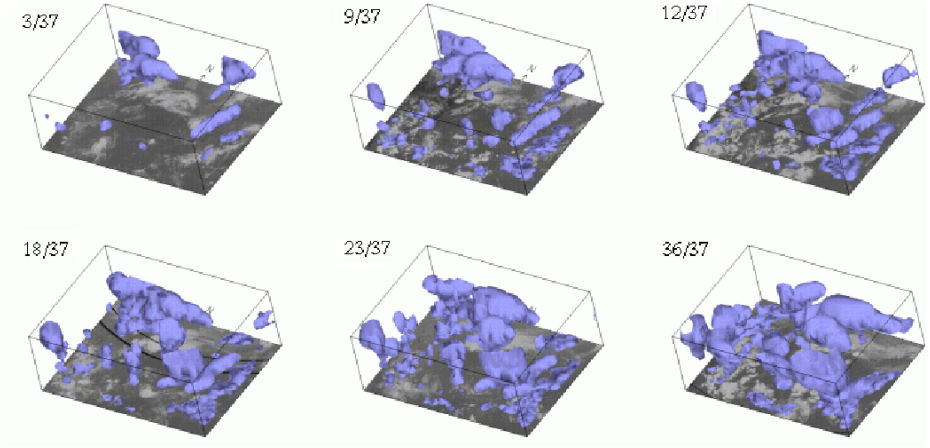


RUTGERS

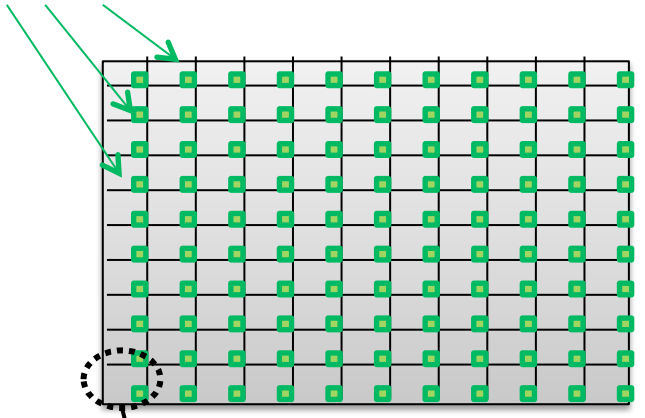
In-Situ Feature Extraction and Tracking using Decentralized Online Clustering (DISC'12)



DOC Overlay



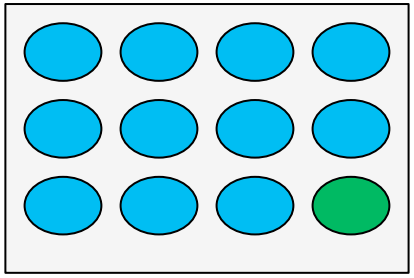
DOC workers executed in-situ on simulation machines



Simulation Compute Nodes

Processor core runs simulation

Processor core runs DOC worker



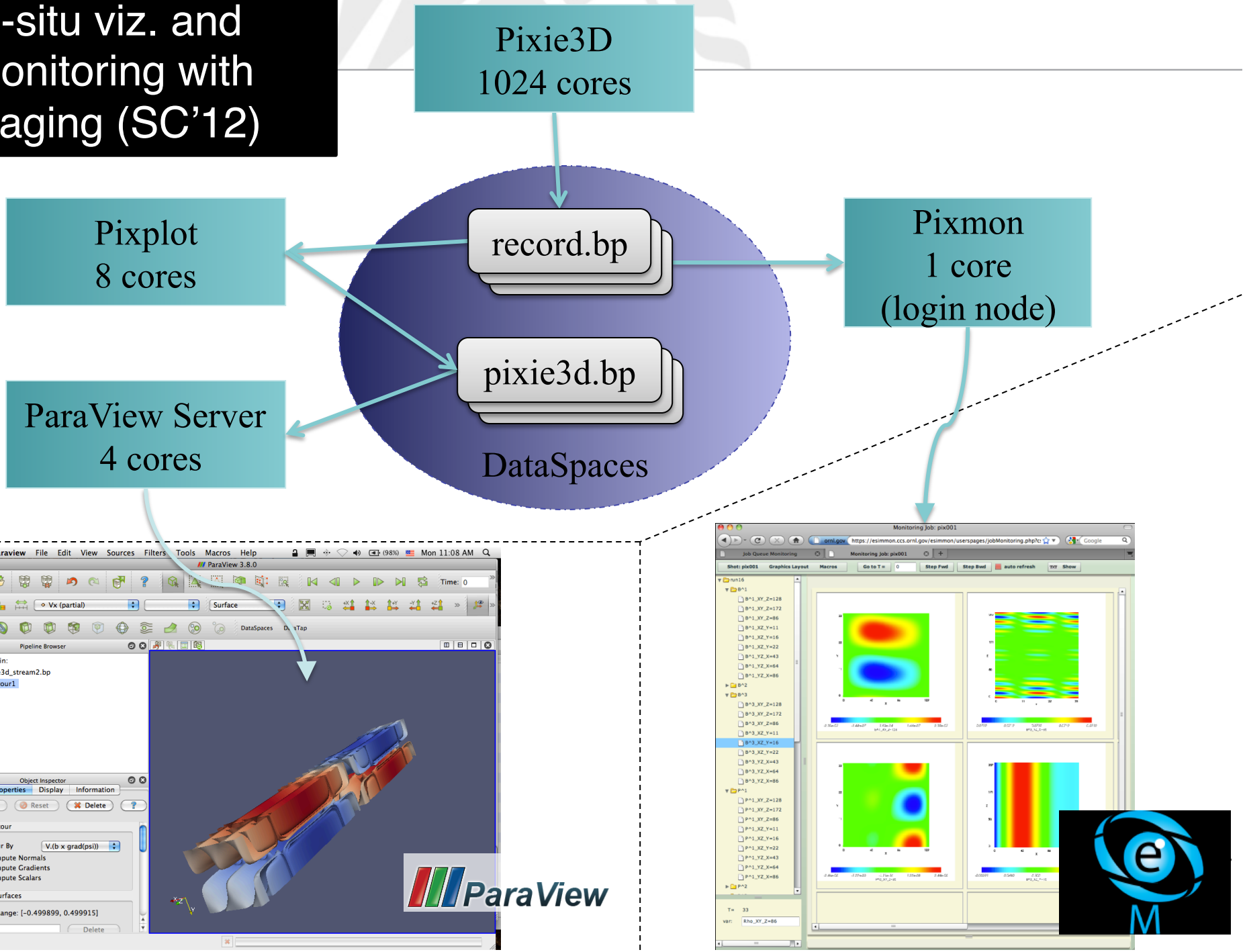
One compute node

Benefits of runtime feature extraction and tracking

(1) Scientists can follow the events of interest (or data of interest)

(2) Scientists can do real-time monitoring of the running simulations

In-situ viz. and monitoring with staging (SC'12)



RUTGERS

Scalable Online Data Indexing and Querying (BDAC'14, CCPE'15)

- Enable query-driven data analytics to extract insights from the large volume of scientific simulation data

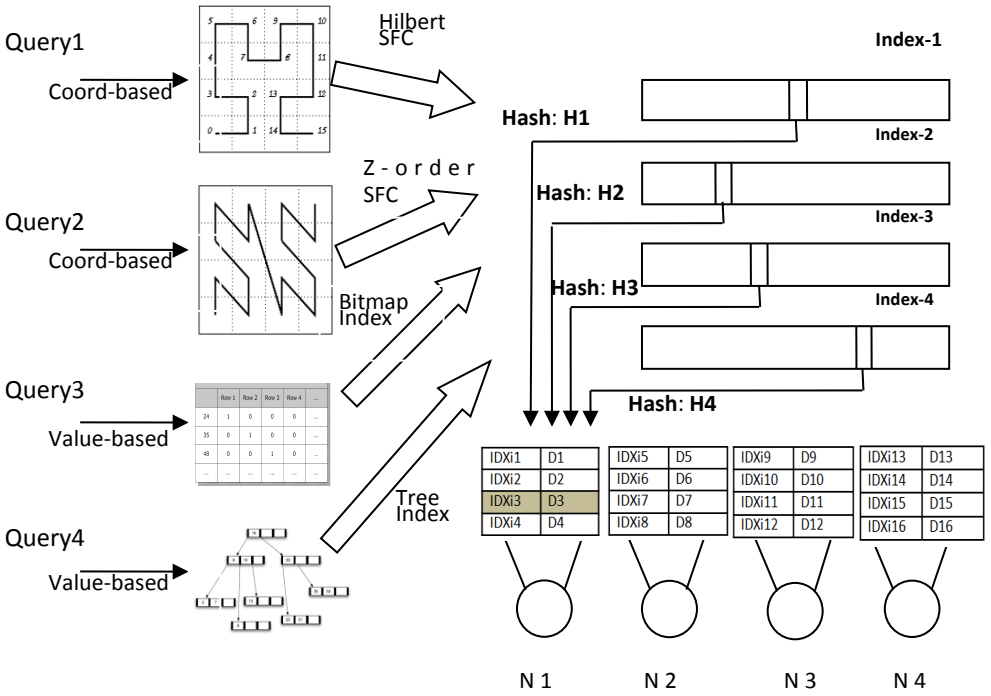


Figure. Conceptual view of the programmable online indexing & querying using multiple index

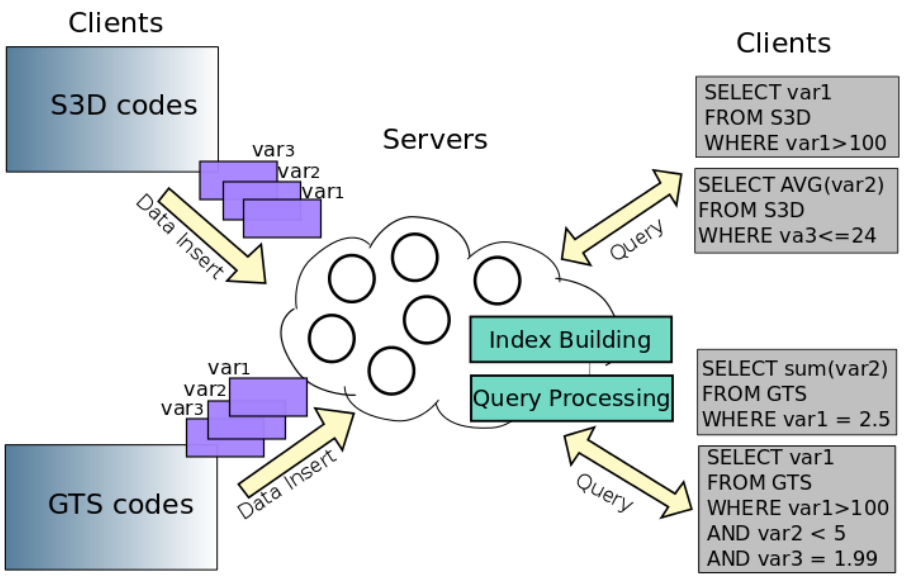


Figure. Value-based online index and query framework using FastBit

Multi-tiered Data Staging with Autonomic Data Placement Adaptation – Overview (IPDPS'15)

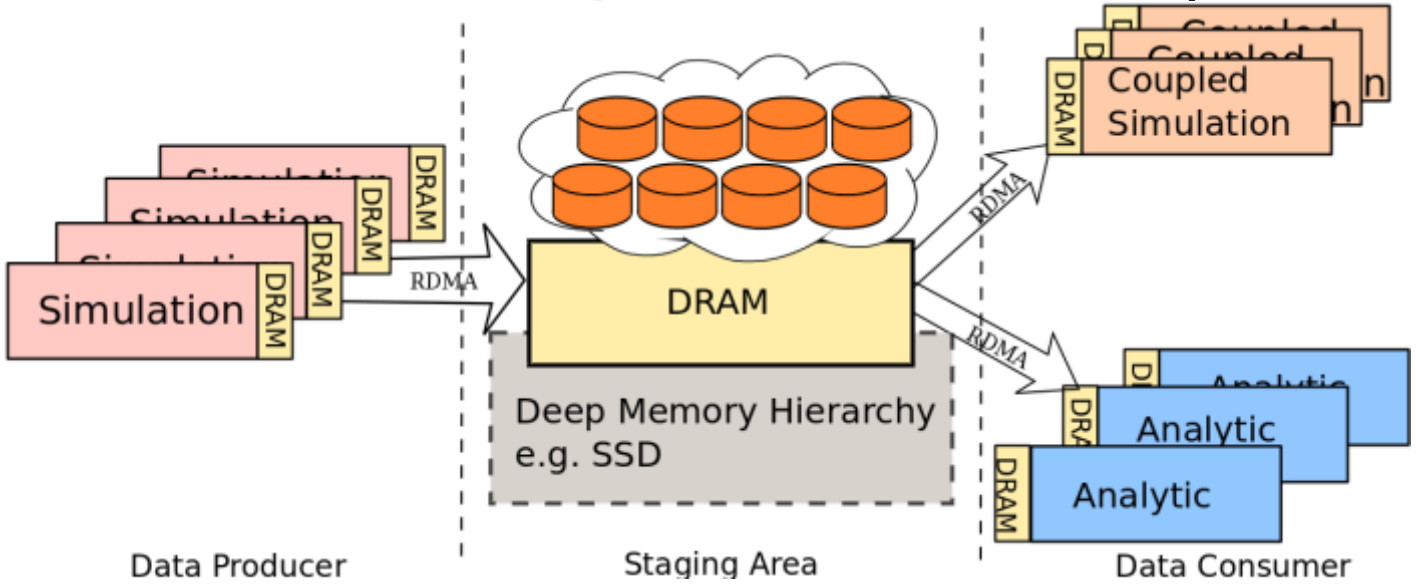
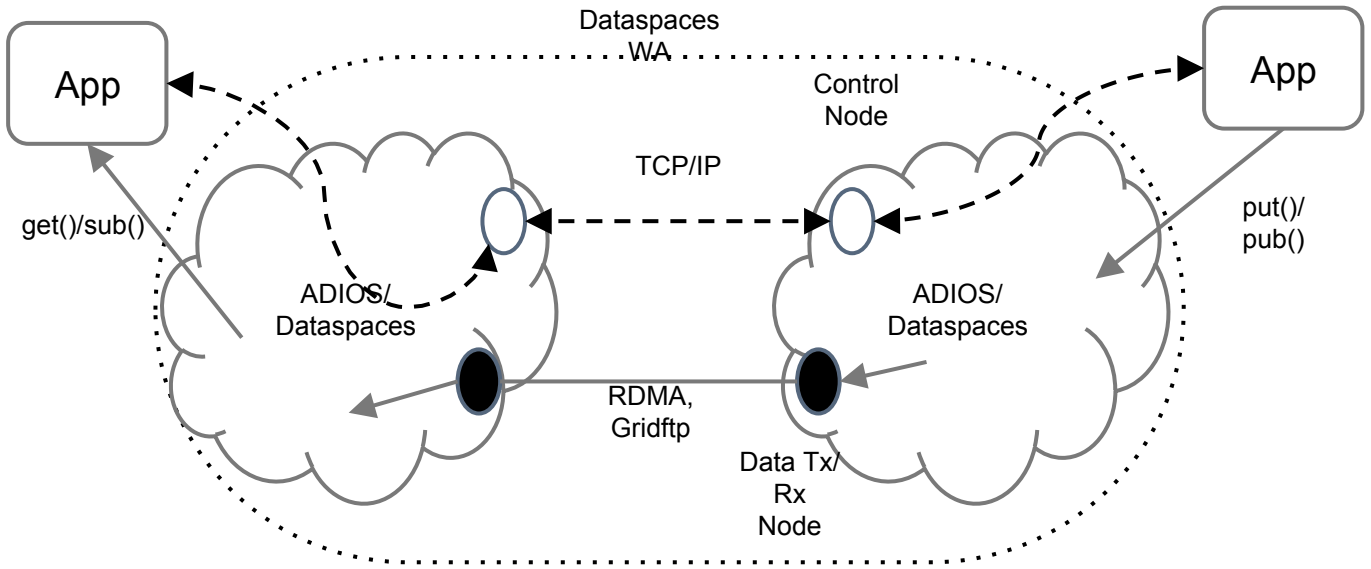


Fig. Conceptual framework for realizing multi-tiered staging for coupled data-intensive simulation workflows.

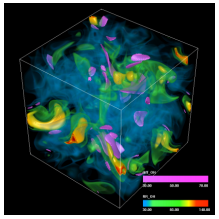
- **A multi-tiered data staging approach that uses both DRAM and SSD to support data-intensive simulation workflows with large data coupling requirements.**
- **Efficient application-aware data placement mechanism that leverages user provided data access pattern information**

Wide-area Staging (Demo at SC'14)

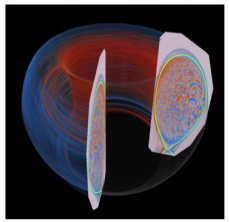
- Wide-area shared staging abstraction achieved by interconnecting multiple staging instances
  - Leverages SDN for provisioning; Workflow/Grid technologies
- Data placement optimized based on demand, availability, locality, etc.



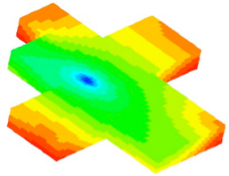
Combustion: S3D uses chemical model to simulate combustion process in order to understand the ignition characteristic of bio-fuels for automotive usage.
Publication: SC 2012.



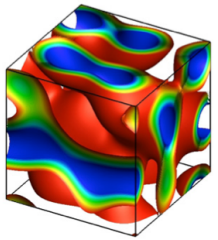
Fusion: EPSi simulation workflow provides insight into edge plasma physics in magnetic fusion devices by simulating movement of millions particles.
Publication: CCGrid 2010, Cluster 2014



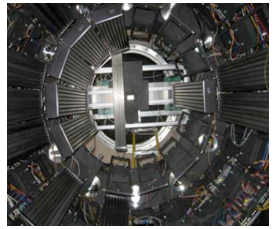
Subsurface modeling: IPARS uses multi-physics, multi-model formulations and coupled multi-block grids to support oil reservoir simulation of realistic, high-resolution reservoir studies with a million or more grids.
Publication: CCGrid 2011



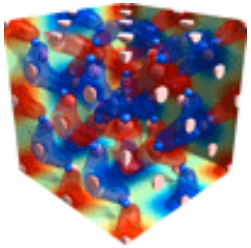
Computational mechanics: FEM coupled with AMR is often used to solve heat transfer or fluid dynamic problems. AMR only performs refinements on important region of the mesh. (work in progress)



Material Science: SNS workflow enables integration of data, analysis, and modeling tools for materials science experiments to accelerate scientific discoveries which requires beam lines data query capability. (work in progress)



Material Science: QMCPack utilizes quantum Monte Carlo (QMC) studies in heterogeneous catalysis of transition metal nanoparticles, phase transitions, properties of materials under pressure, and strongly correlated materials. (work in progress)

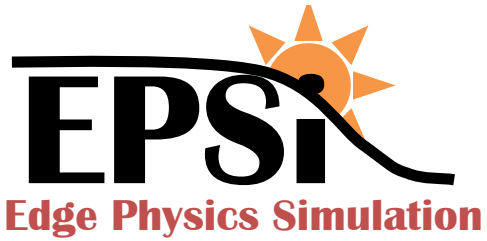


Summary & Conclusions

- Complex applications running on high-end systems generate extreme amounts of data that must be managed and analyzed to get insights
  - Data costs (performance, latency, energy) are quickly dominating
  - Traditional data management/analytics pipelines are breaking down
- Hybrid data staging, in-situ workflow execution, etc. can address this challenges
  - Users to efficiently intertwine applications, libraries, middleware for complex analytics
- Many challenges; Programming, mapping and scheduling, control and data flow, autonomic runtime management….
  - The DataSpaces project explores solutions at various levels

RUTGERS

Thank You!



Manish Parashar

Email: parashar@rutgers.edu

WWW: parashar.rutgers.edu

WWW: dataspaces.org

DataSpaces Team @ RU