



[EXPLORE THE NATIONAL LABS](#)

[EXPLORE BY TOPIC](#)

[COLLABORATIVE PROJECTS](#)

[BOOTH SCHEDULE](#)

[HOME](#)



Data Science
at Scale Team
Los Alamos
National Laboratory

MarFS -
A Scalable Near-Posix
Name Space over
Cloud Objects

Runtime and
Storage System
Co-design

[BACK TO
MAP](#)

Trinity Center of
Excellence &
Early Science Projects



MarFS - A Scalable Near-Posix Name Space over Cloud Objects

What is MarFS? Near-POSIX global scalable name space over many POSIX and non POSIX data repositories (Scalable object systems - CDMI, S3, RestFul API, etc.)

- It scales name space by sewing together multiple POSIX file systems both as parts of the tree and as parts of a single directory allowing scaling across the tree and within a single directory
- It is small amount of code (C/C++/Scripts)
 - A small Linux Fuse
 - A pretty small parallel batch copy/sync/compare/ utility
 - A set of other small parallel batch utilities for management
 - A moderate sized library both FUSE and the batch utilities call
- Data movement scales just like many scalable object systems
- Metadata scales like NxM POSIX name spaces both across the tree and within a single directory
- It is friendly to object systems by
 - Spreading very large files across many objects
 - Packing many small files into one large data object
- How about a Scalable Near-POSIX Name Space over Object Erasure ?

Best of both worlds

- Objects Systems
 - ❖ Provide massive scaling and efficient erasure techniques
 - ❖ Friendly to applications, not to people. People need a name space.
 - ❖ Huge Economic appeal (erasure enables use of inexpensive storage)
- POSIX name space is powerful but has issues scaling

The challenges

- Mismatch of POSIX an Object metadata, security, read/write semantics, efficient object/file sizes.
- No update in place with Objects
- How do we scale POSIX name space to trillions of files/directories

SC15 Presentation and talk schedule

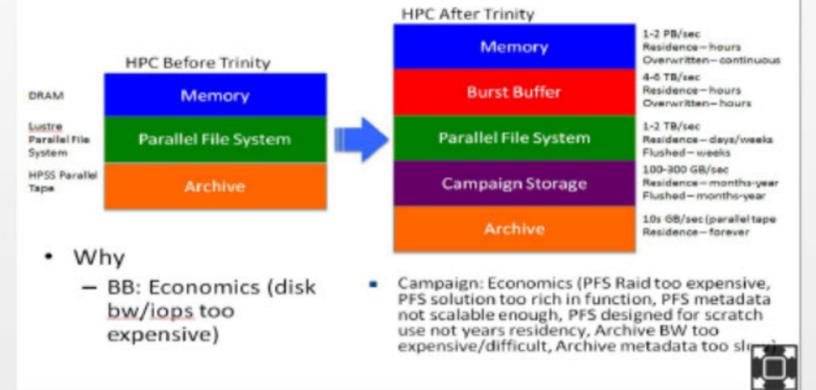
SC15 BOF, Two Tiers Scalable Storage: Building POSIX-Like Namespaces with Object Stores

Wed Nov. 18th, 2015, 5:30-7:00PM, Hilton Salon A

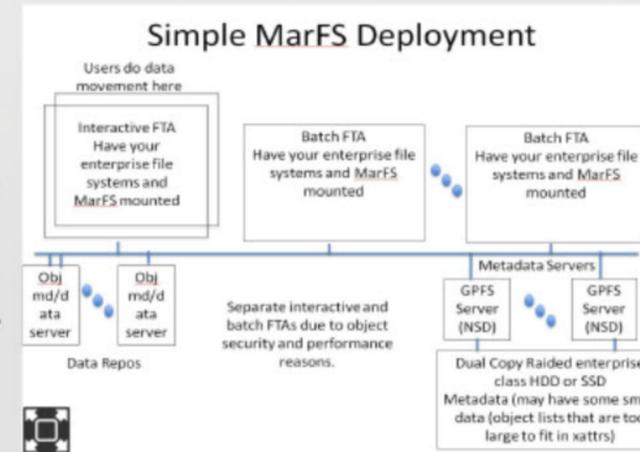
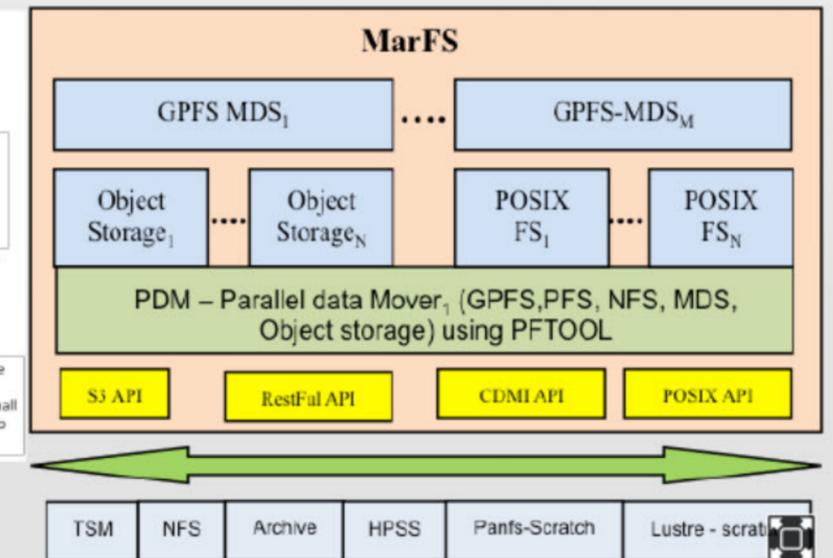
Brighttalk – webinars, Nov.10th, 2015, Speaker: Gary Grider, MarFS: A Scalable Near-POSIX Name Space over Cloud Objects

Why do we need MarFS ?

What are all these storage layers?
Why do we need all these storage layers?



We need capacity tier – Campaign Storage



BACK TO MAP

FUNDING & CREDITS

DOE transforms HPC



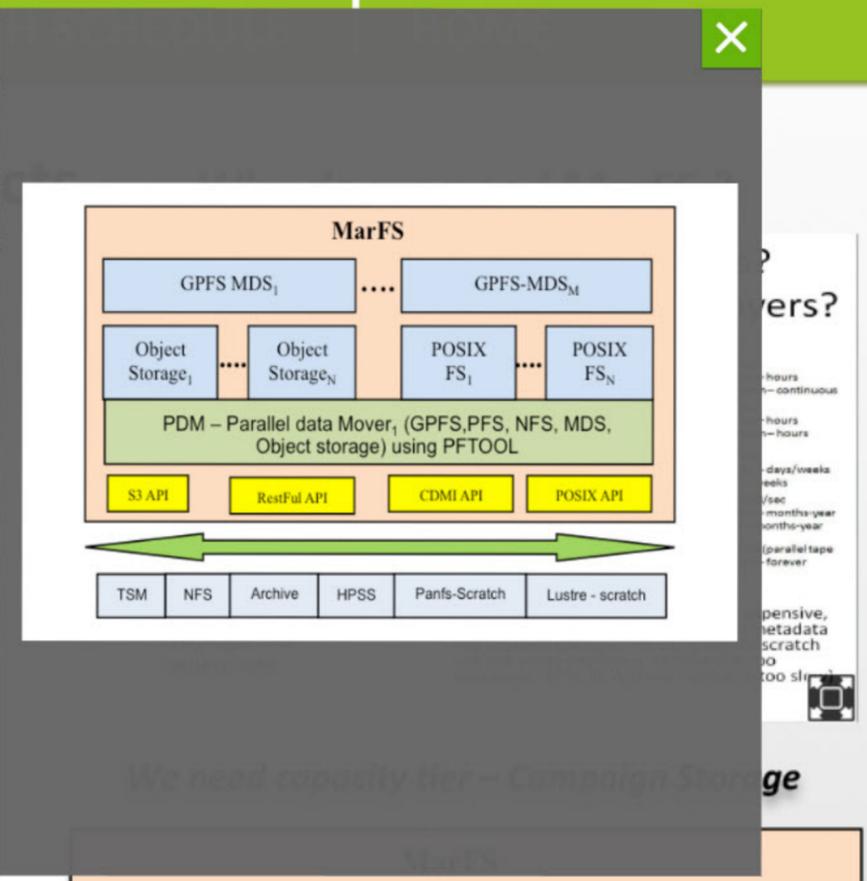
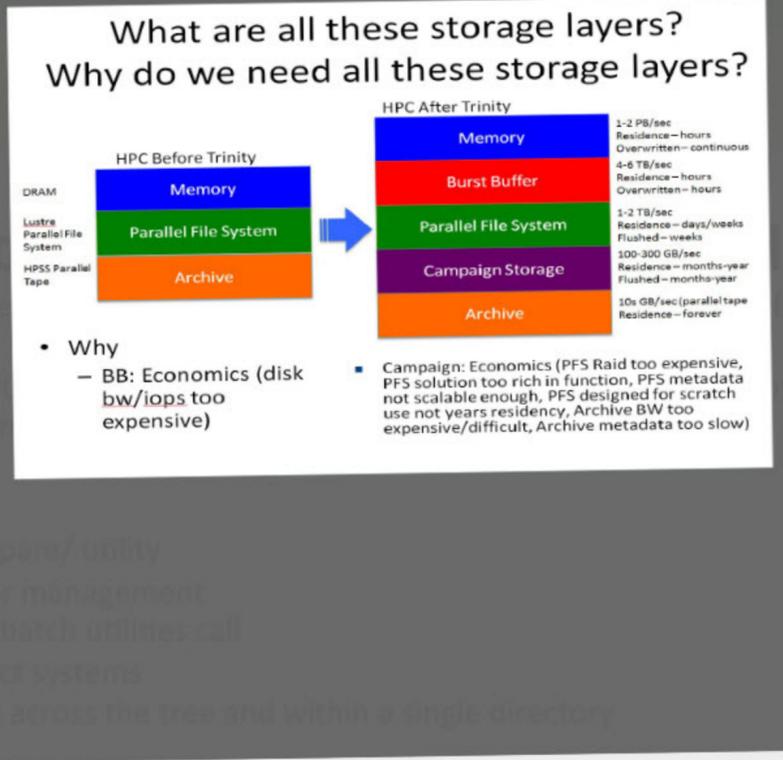
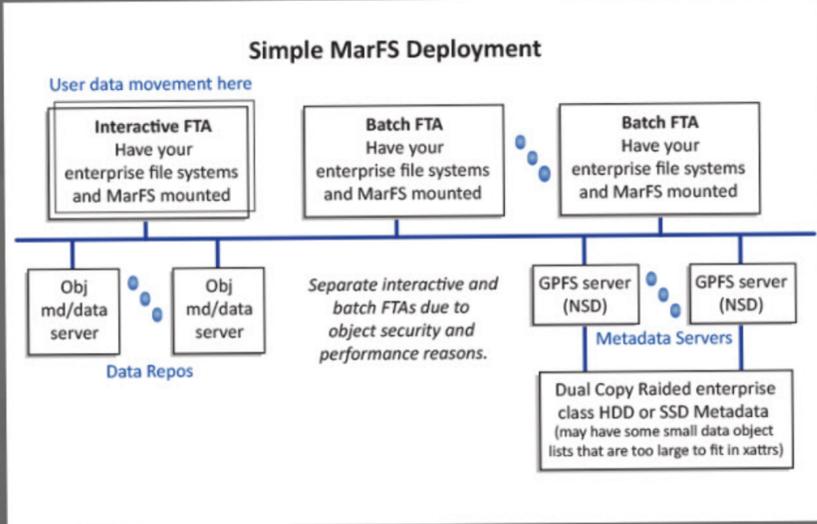
EXPLORE

TOPIC

BOOT

Obj

a single



Near-P

able name spa

together multiple

the tree and with

(Scripts)

h copy/sync/com

l batch utilities fo

both FUSE and the

any scalable obje

name spaces bot

cross many objects

to one large data object

Name Space over Object Erasure ?

and efficient erasure techniques

Priority to applications, not to people

❖ Huge Economic appeal (erasure enable)

POSIX name space is powerful but ha

The challenges

➤ Mismatch of POSIX an Object metadata, efficient object/file sizes.

➤ No update in place with Objects

➤ How do we scale POSIX name space to tr

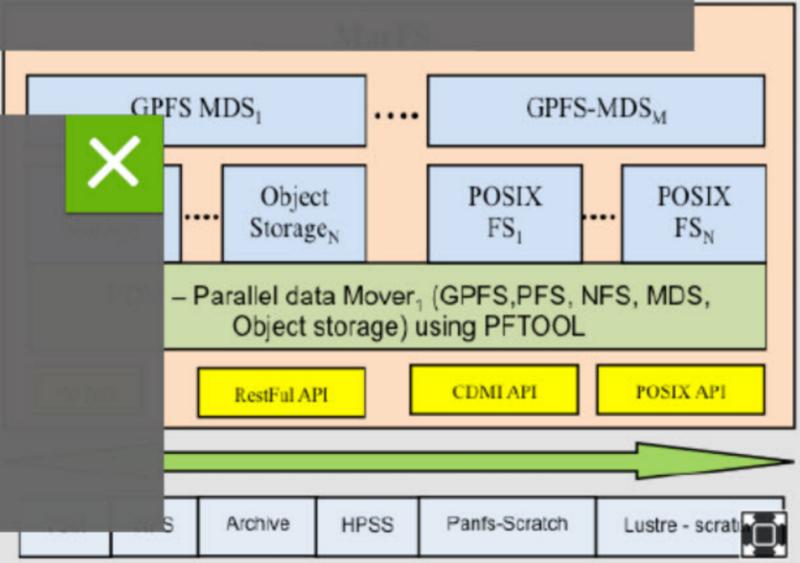
SC15 Presentation and talk schedule

SC15 BOF, Two Tiers Scalable Storage: Building P

Wed Nov. 18th, 2015, 5:30-7:00PM, Hilton Salon A

Brighttalk - webinars, Nov. 10th, 2015, Speaker: Gary Grider, MarFS: A Scalable Near-POSIX Name Space over Cloud Objects

Funding: DOE



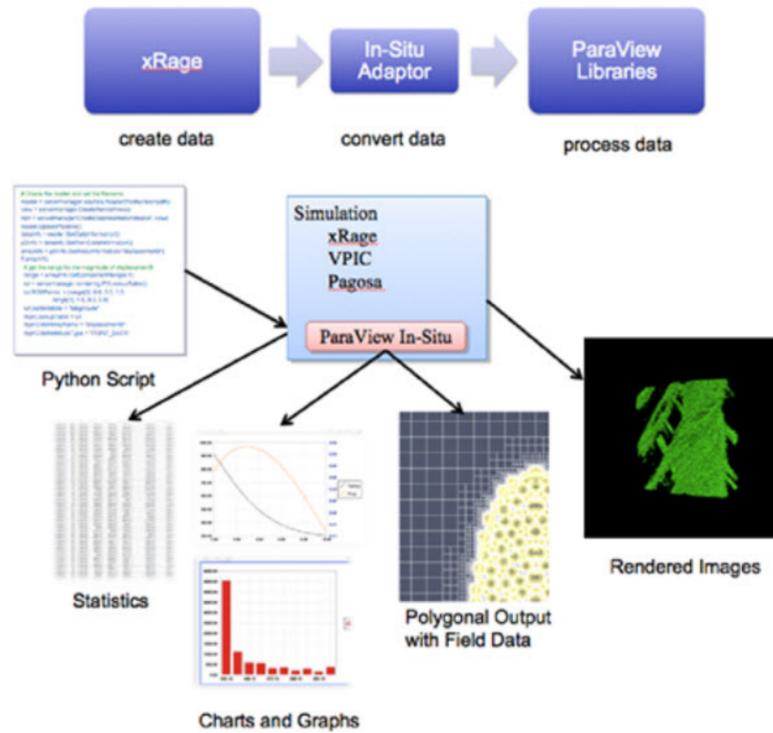
We need capacity for campaign storage

FUNDING & CREDITS

BACK TO MAP

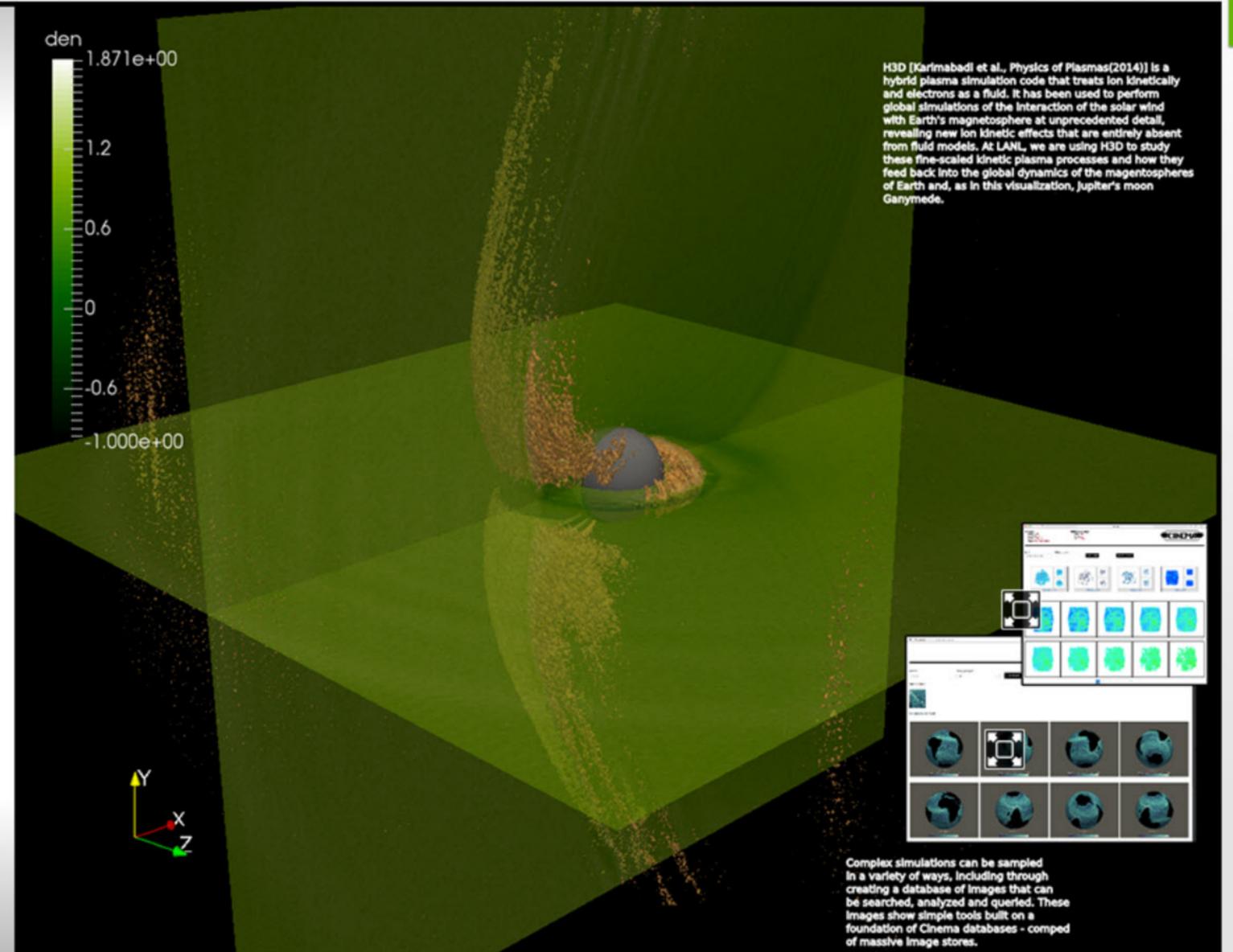


Los Alamos National Laboratory is exploring methods to sample, compress, and visualize the results from large scale simulations in-situ (while the sim. is running).



This effort combines research in analytics, workflows, compression and scheduling to provide a flexible set of options for scientists to achieve a wide range of workflows that can be tuned to specific solutions.

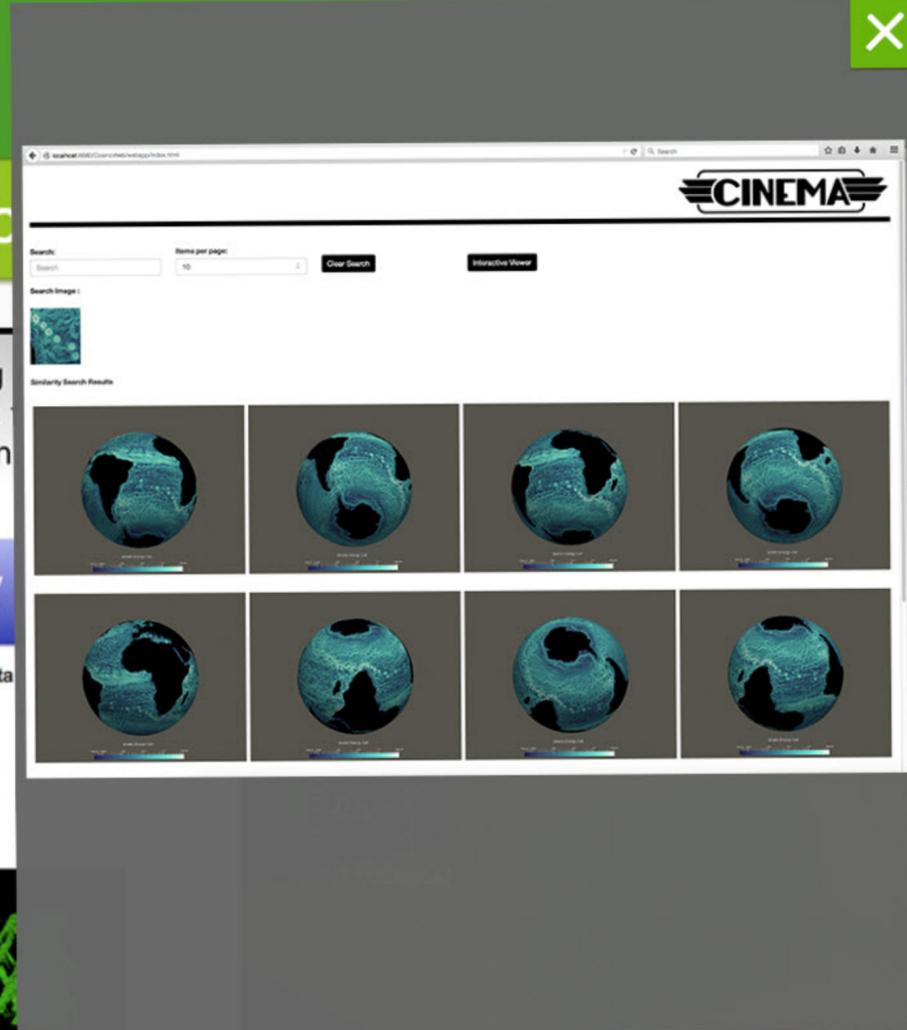
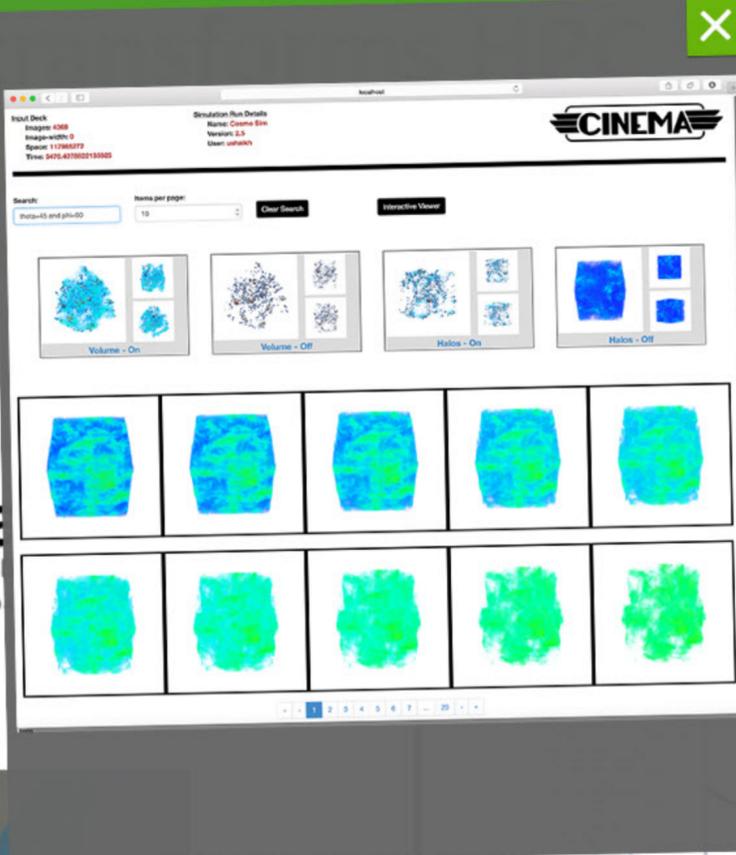
In-situ Analysis and Vis



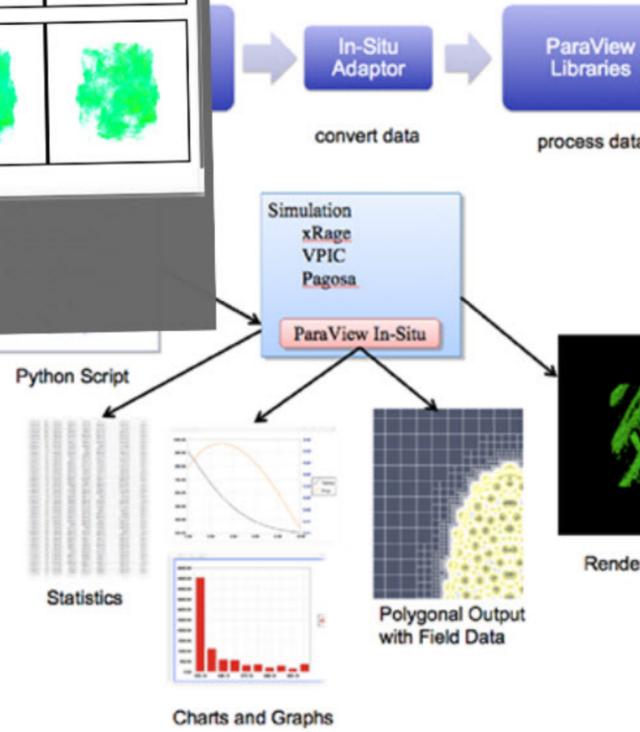
Data Science at Scale Team
Los Alamos National Laboratory



BACK TO MAP



onal Laboratory is exploring
ss, and visualize the results
ations in-situ (while the sim



H3D [Karimabadi et al., Physics of Plasmas(2014)] is a hybrid plasma simulation code that treats ion kinetically and electrons as a fluid. It has been used to perform global simulations of the interaction of the solar wind with Earth's magnetosphere at unprecedented detail, revealing new ion kinetic effects that are entirely absent from fluid models. At LANL, we are using H3D to study these fine-scaled kinetic plasma processes and how they feed back into the global dynamics of the magnetospheres of Earth and, as in this visualization, Jupiter's moon Ganymede.

Complex simulations can be sampled in a variety of ways, including through creating a database of images that can be searched, analyzed and queried. These images show simple tools built on a foundation of Cinema databases - comped of massive image stores.

This effort combines research in analytics, workflows, compression and scheduling to provide a flexible set of options for scientists to achieve a wide range of workflows that can be tuned to specific solutions.

In-situ Analysis and Vis

Data Science at Scale Team
Los Alamos National Laboratory

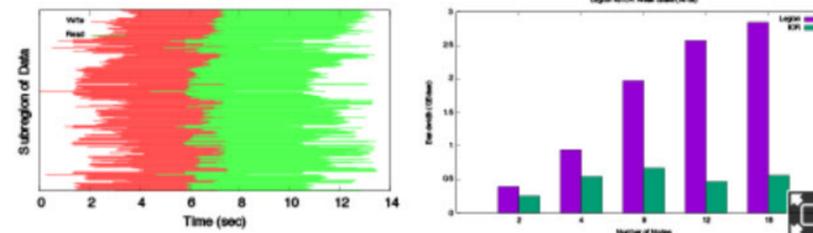


BACK TO MAP

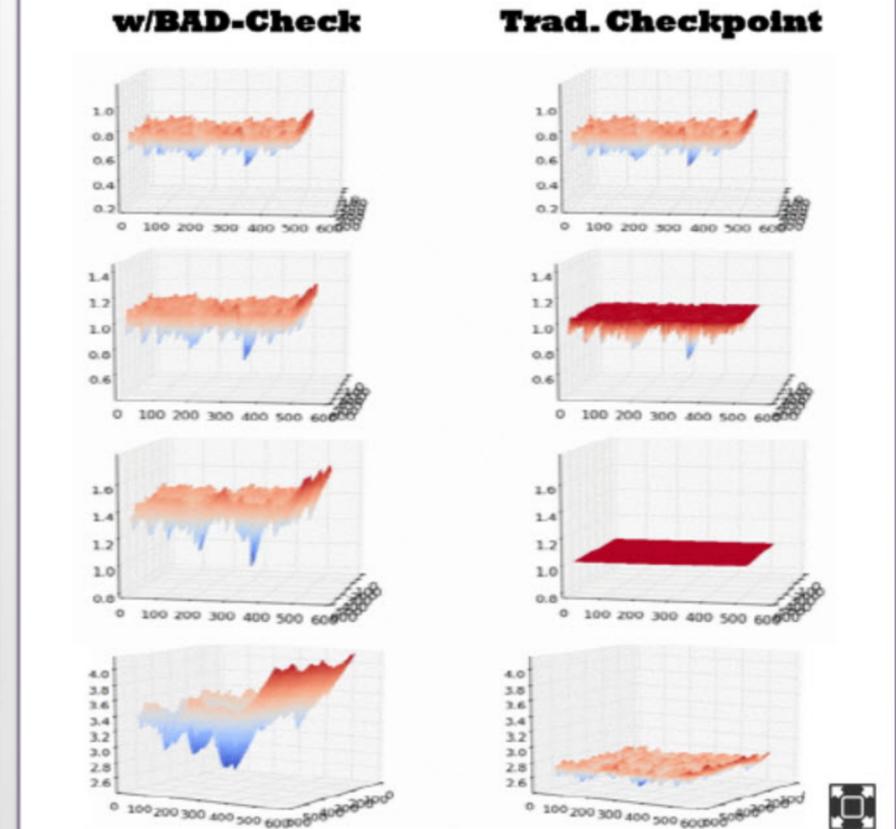
Runtime and Storage System Co-design

PIs: Brad Settlemyer and Galen Shipman

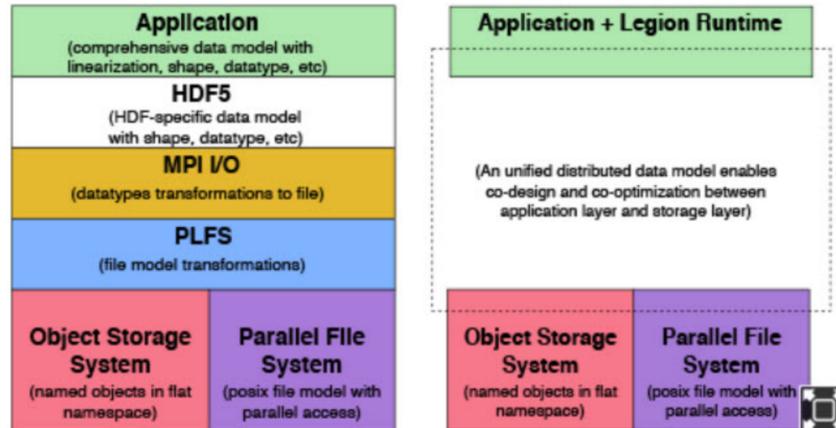
Abstract - In order to accelerate simulation science storage system workflows, research efforts at LANL span the spectrum from analytical modeling to systems software prototypes. In this poster we describe a portion of LANL's on-going storage research efforts.



In the figure above left application file region accesses are shown on the y-axis and time on the x-axis, illustrating the runtime's ability to asynchronously schedule uncoordinated I/O while providing a consistent view of the data. The performance of this approach is compared to that of traditional bulk-synchronous coordinated I/O in the above left.



Visualization demonstrating BAD-Check's improved efficiency for HIGRAD/FIRETEC.

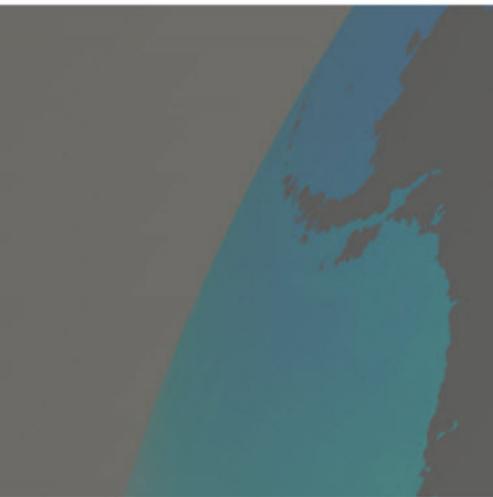


In a contemporary I/O stack, shown on the left, each layer uses a distinct data model. Our proposed architecture, shown on the right, uses a unified data model with run-time support in Legion to enable system-wide co-design and co-optimization strategies.

BAD-Check, an asynchronous, distributed transaction protocol enables asynchronous checkpoint creation



for parallel scientific simulations featuring few global communications and frequent computational hotspots with migration.



BACK TO MAP



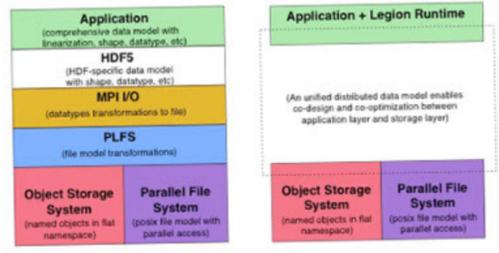
FUNDING & CREDITS

DOE transforms HPC

EXPLORE THE



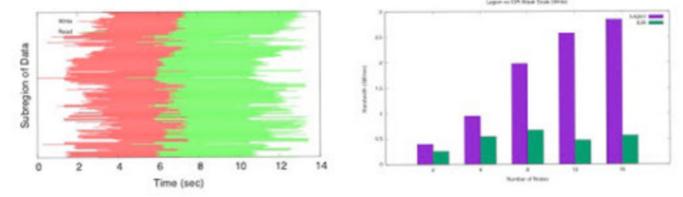
Improving scalability of I/O through runtime and storage system co-design



In a contemporary I/O stack each layer uses a distinct data model (left). Our proposed architecture uses a unified data model and run-time to enable system-wide co-design and co-optimization strategies (right).

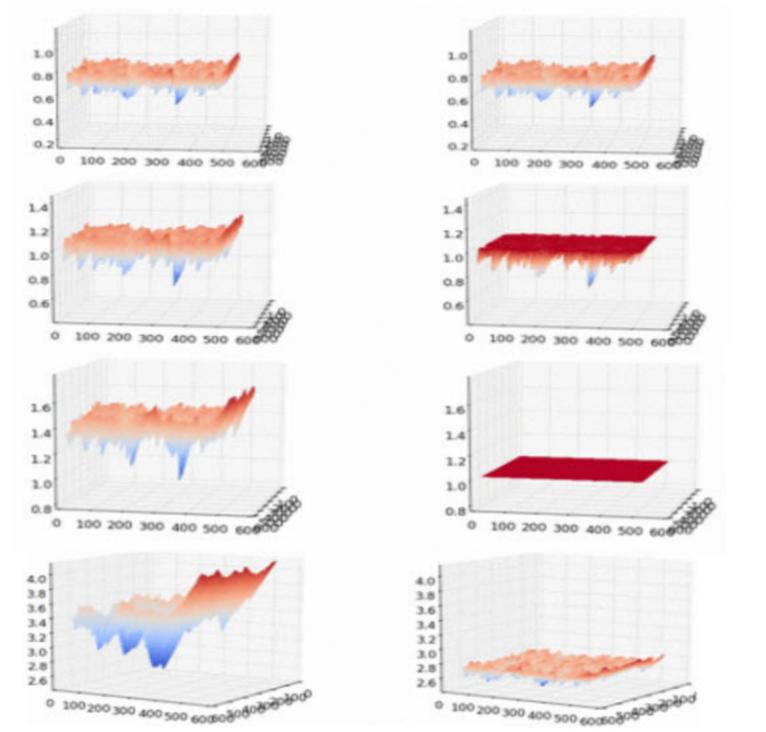
Initial results

Runtimes, such as Legion are capable of introspecting global data dependencies and utilize available computation during I/O phases while maintaining a *consistent cut* of the globally distributed data structure



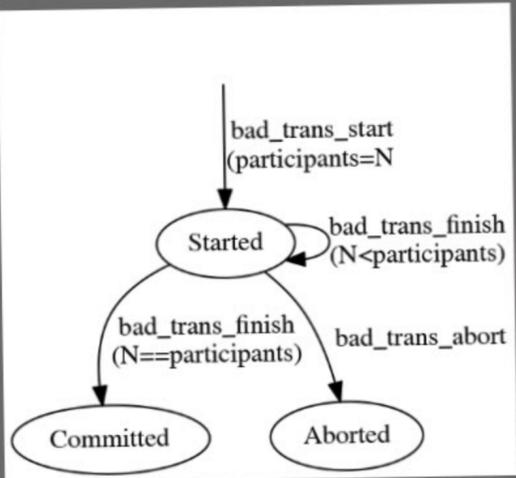
Write and Read phases demonstrating independent I/O scheduling. Legion I/O over Lustre compared with N-1 IOR over HDF5 over MPI-I/O over Posix on Lustre

w/BAD-Check Trad. Checkpoint



Visualization demonstrating BAD-Check's improved efficiency for HIGRAD/FIRETEC.

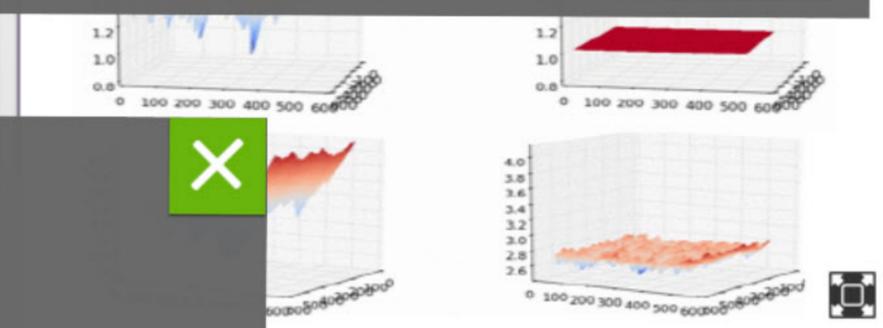
Application Application + Legion



while providing a consistent view of the data. The performance of this approach is compared to that of traditional bulk-synchronous

Funding: DOE

for parallel scientific simulations featuring few global communications and frequent computational hotspots with migration.



Visualization demonstrating BAD-Check's improved efficiency for HIGRAD/FIRETEC.

BACK TO MAP



FUNDING & CREDITS



EXPLORE THE NATIONAL LABS

EXPLORE BY TOPIC

COLLABORATIVE PROJECTS

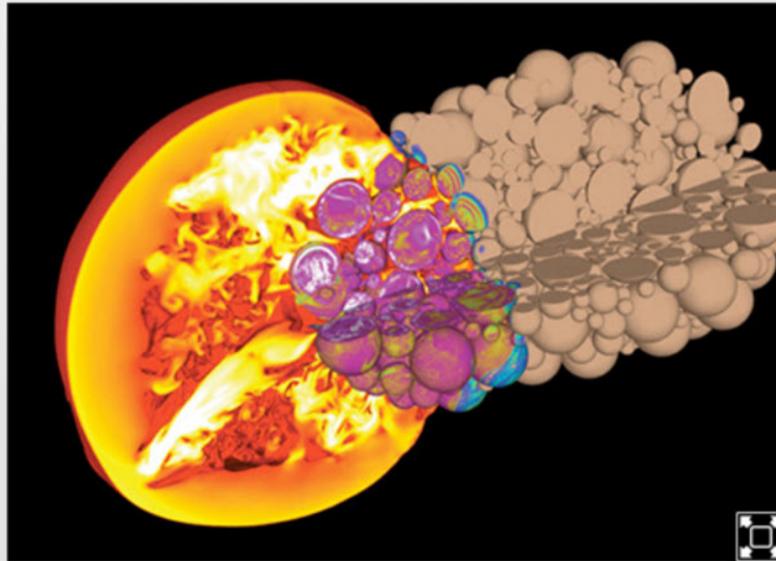
BOOTH SCHEDULE

HOME

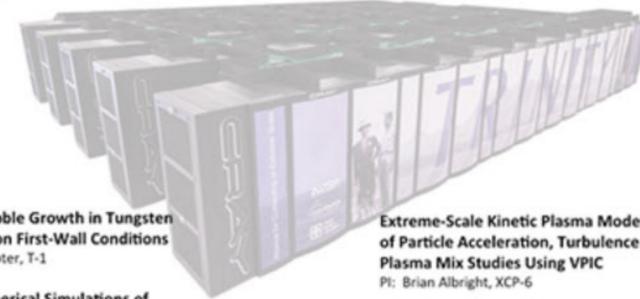


Trinity Center of Excellence & Early Science Projects Setting the stage for science exploration on Trinity

Collaborating closely with vendor partners at Cray and Intel, LANL applications integral to both ASC and Office of Science research undergo code developments in preparation for Trinity.



Six LANL Open Science projects were chosen to be early users of Trinity, to optimally use the new Intel Knights Landing (KNL) processors, high-bandwidth memory and burst buffer. Project areas include materials modeling, plasma physics, neural networks, turbulence modeling, and molecular dynamics.



Helium Bubble Growth in Tungsten under Fusion First-Wall Conditions
PI: Arthur Voter, T-1

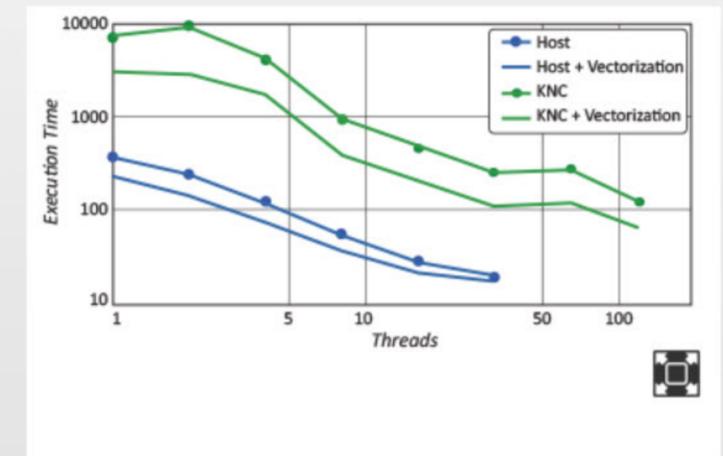
Direct Numerical Simulations of Magnetic Rayleigh-Taylor Instability
PI: Daniel Livescu, CCS-2

Materials Dynamics via Large-Scale Molecular Dynamics and Embedded Scale-Bridging Simulations
PI: Tim Germann, T-1

Extreme-Scale Kinetic Plasma Modeling of Particle Acceleration, Turbulence and Plasma Mix Studies Using VPIC
PI: Brian Albright, XCP-6

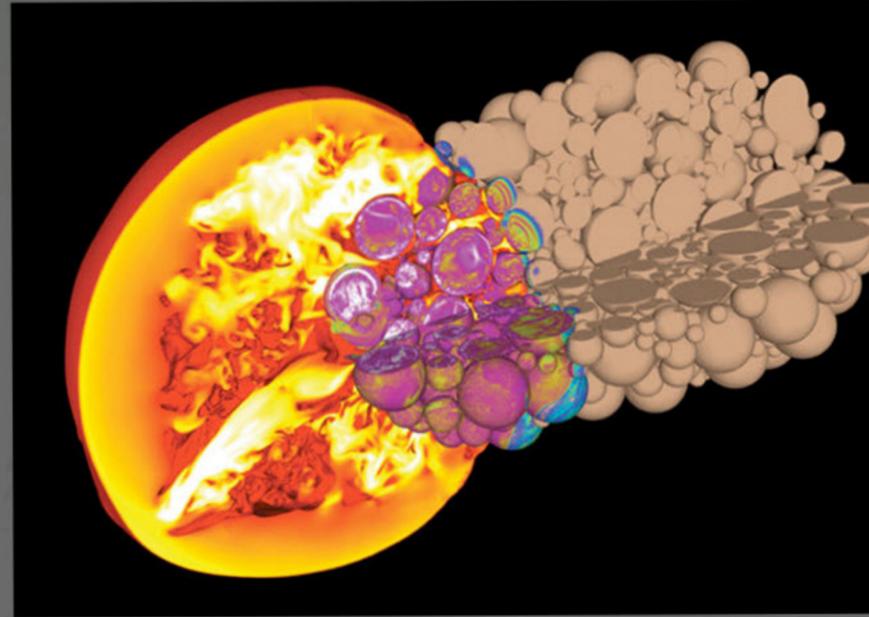
Deep Sparse Columnar Neural Network (dSCANN)
PI: Garrett Kenyon, P-21

Advancing Regenerative Medicine with Trinity: Defining a New State-of-the-Art for Biomolecular Simulation
PI: Karissa Sanbonmatsu, T-6



BACK TO MAP

FUNDING & CREDITS

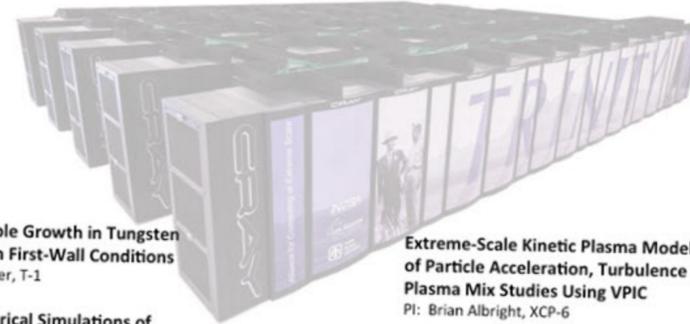


The xRAGE hydro-code developed at LANL is a key Trinity Center of Excellence application. Moving multi-physics codes to new platforms requires close collaborations between domain, computational, and computer scientists at LANL and hardware/software vendor partners. Image Caption: xRAGE has been used to model a 1-Mton energy source on the surface of an asteroid to investigate mitigation strategies for potential hazardous objects in space.

NOTE: the previous reference we used for this was Robert P. Weaver, XTD-6, LANL from 2012 – this was done on Cielo



Six LANL Open Science projects were chosen to be early users of Trinity, to optimally use the new Intel Knights Landing (KNL) processors, high-bandwidth memory and burst buffer. Project areas include materials modeling, plasma physics, neural networks, turbulence modeling, and molecular dynamics.



Helium Bubble Growth in Tungsten under Fusion First-Wall Conditions
PI: Arthur Voter, T-1

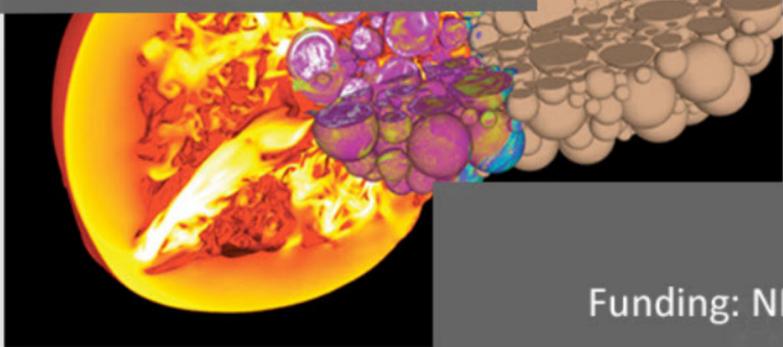
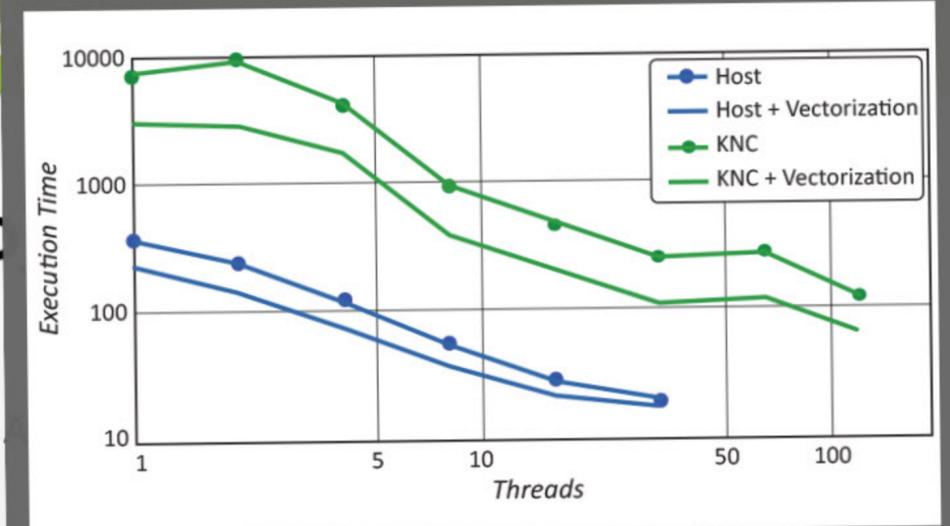
Direct Numerical Simulations of Magnetic Rayleigh-Taylor Instability
PI: Daniel Livescu, CCS-2

Materials Dynamics via Large-Scale Molecular Dynamics and Embedded Scale-Bridging Simulations
PI: Tim Germann, T-1

Extreme-Scale Kinetic Plasma Modeling of Particle Acceleration, Turbulence and Plasma Mix Studies Using VPIC
PI: Brian Albright, XCP-6

Deep Sparse Columnar Neural Network (dSCANN)
PI: Garrett Kenyon, P-21

Advancing Regenerative Medicine with Trinity: Defining a New State-of-the-Art for Biomolecular Simulation
PI: Karissa Sanbonmatsu, T-6



Helium Bubble Growth in Tungsten under Fusion First-Wall Conditions
PI: Arthur Voter, T-1

Extreme-Scale Kinetic Plasma Modeling of Particle Acceleration, Turbulence and Plasma Mix Studies Using VPIC
PI: Brian Albright, XCP-6



Funding: NNSA ASC Program

BACK TO MAP