DOE Computer Graphics Forum Site Report

Sandia National Laboratories
Livermore, CA

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Site & Group Mission

- The SNL/CA site is focused on national security issues including nuclear weapon stockpile surety, energy security (transportation energy and the energy grid), and cyber-security.
- Scalable modeling and analysis are the key tasks of our department and are in support of the site goals.

SNL/CA Site Report

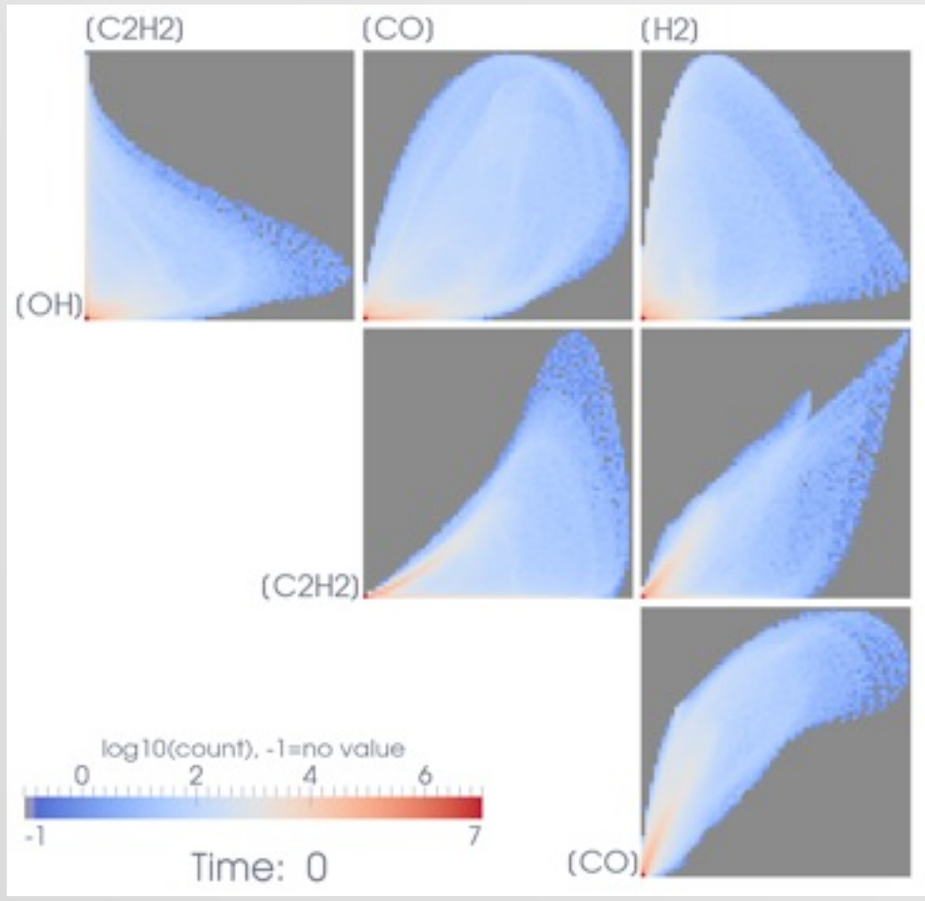
- Hardware
  - New facility: CRCV
  - 6400 ft^2 for people
  - 2000 ft^2 for computers
- Software
  - Working w/ SNL/NM on Titan, NGC, ParaView
  - ASCR MAPD – Combine topology & statistics
  - OVIS – Data collection & analysis for sensor nets
  - Megatux – Emulate & characterize botnets
  - SICAIDA – Cloud-flavored post-processing
  - Automatic Threat Recognition for Advanced Imaging Technology – Human body imaging for DHS

Faculty, Facilities, and Funding

- Staff
  - ~20 people in department, ~10 research
  - Site looking to hire 80 people this year
- HPC
  - Small research cluster w/ vis hardware
  - 256-node dual-core cluster w/o vis hardware
  - Combustion Research Facility has clusters for viz & simulation, but use others' machines for large runs
- Funding includes
  - DOE: NNSA (ASC VIEWS/CSSE), ASCR
  - SNL: LDRD, CSRF

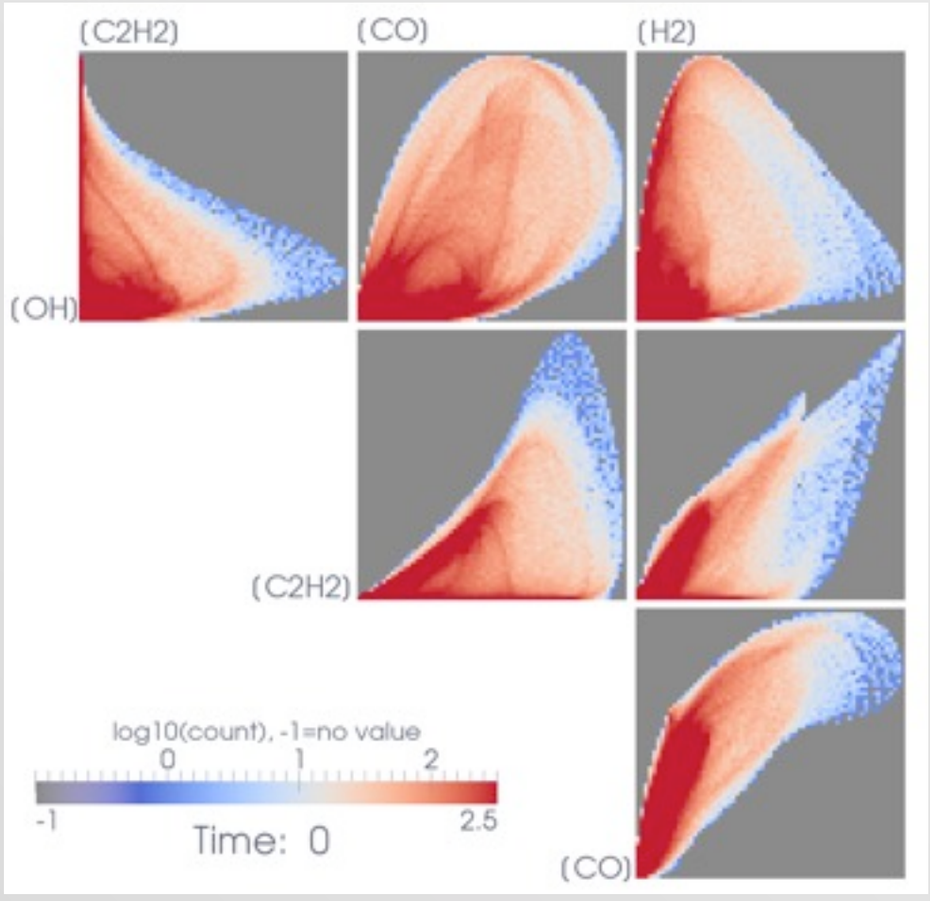
Mathematics of Petascale Data (MAPD)

- Global statistical models average out interesting behavior
- Topological tools generate noisy segmentations
- Can using both at once lead to sharp statistical models on clean topological features?



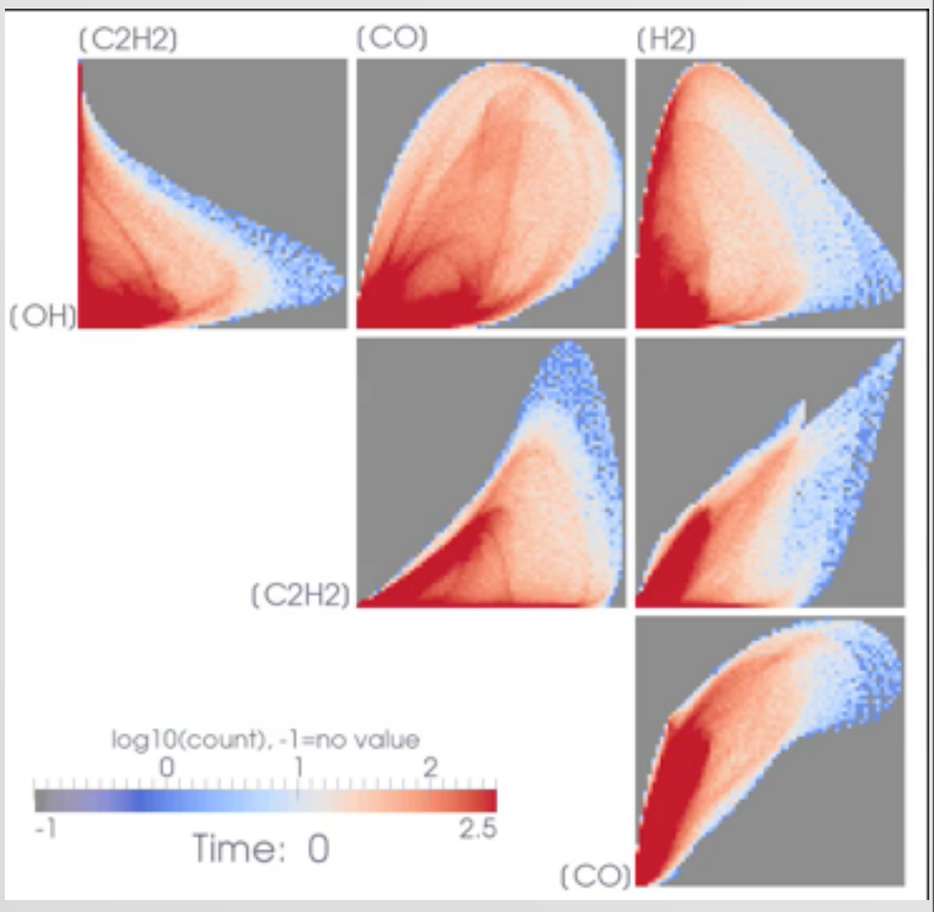
Mathematics of Petascale Data (MAPD)

- Global statistical models average out interesting behavior
- Topological tools generate noisy segmentations
- Can using both at once lead to sharp statistical models on clean topological features?



Mathematics of Petascale Data (MAPD)

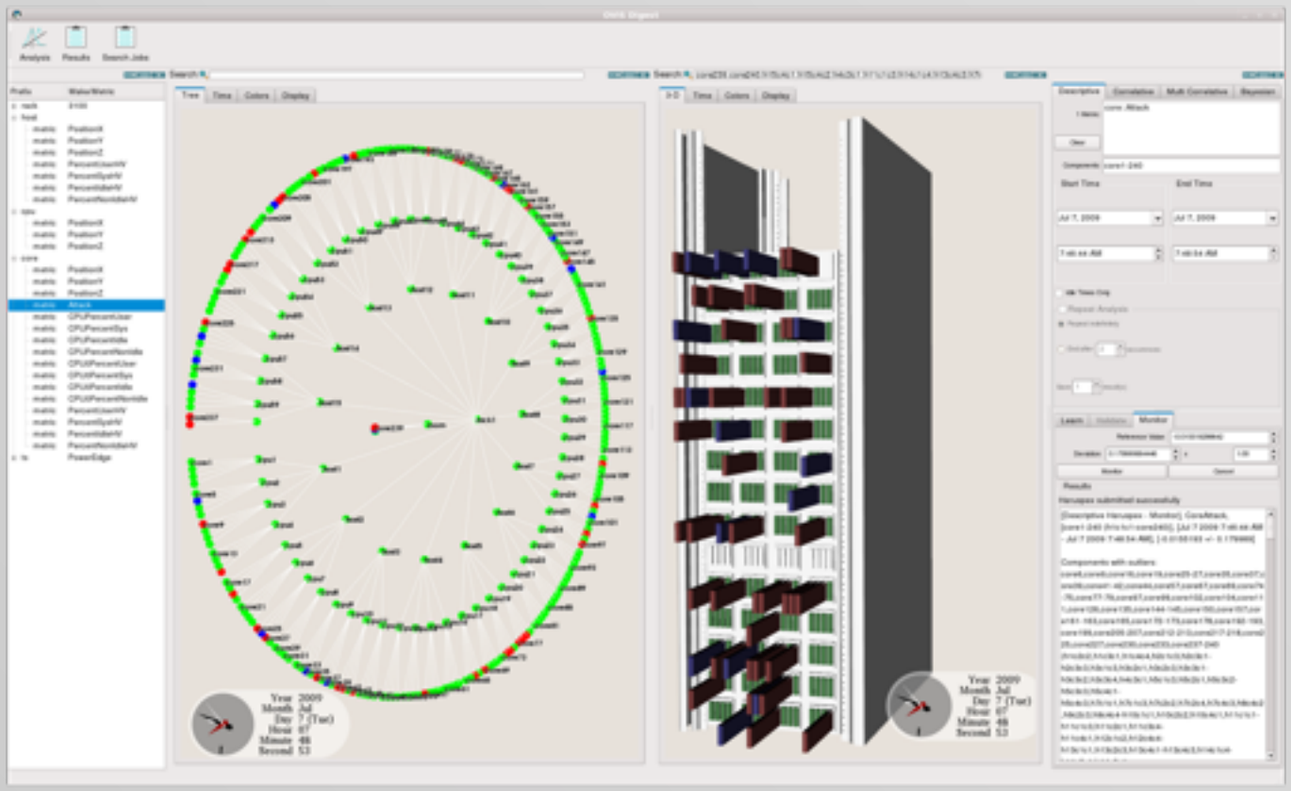
- Global statistical models average out interesting behavior
- Topological tools generate noisy segmentations
- Can using both at once lead to sharp statistical models on clean topological features?

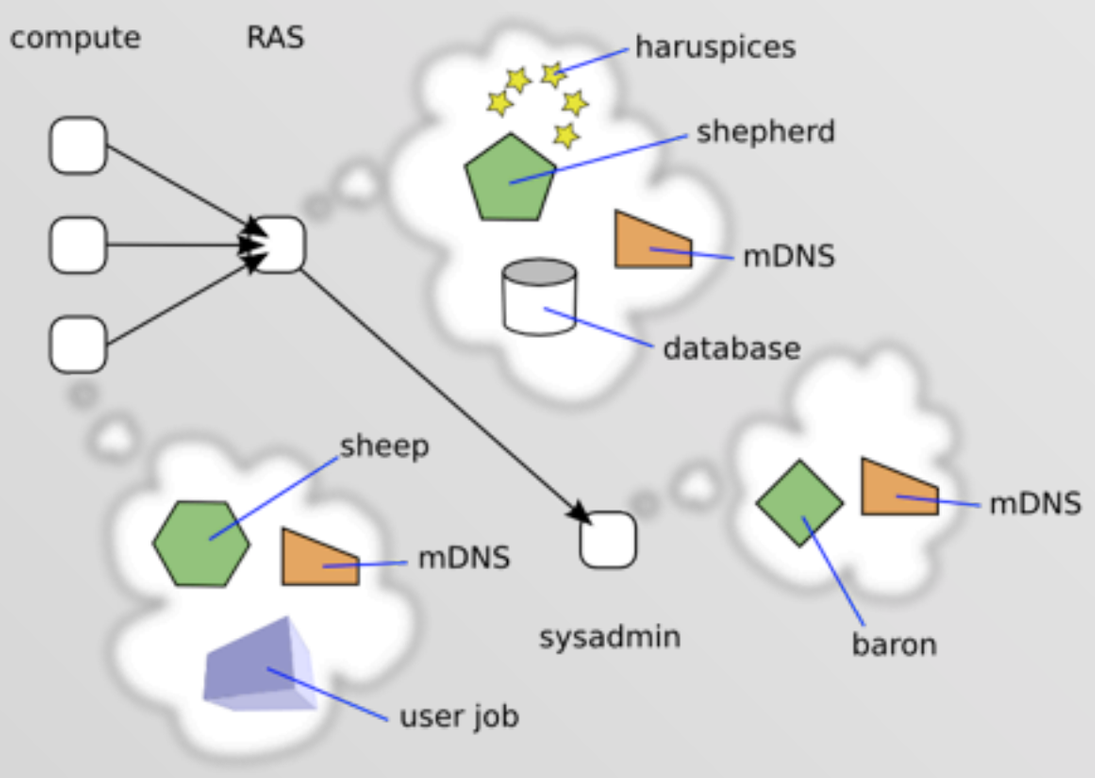


OVIS

http://ovis.ca.sandia.gov/

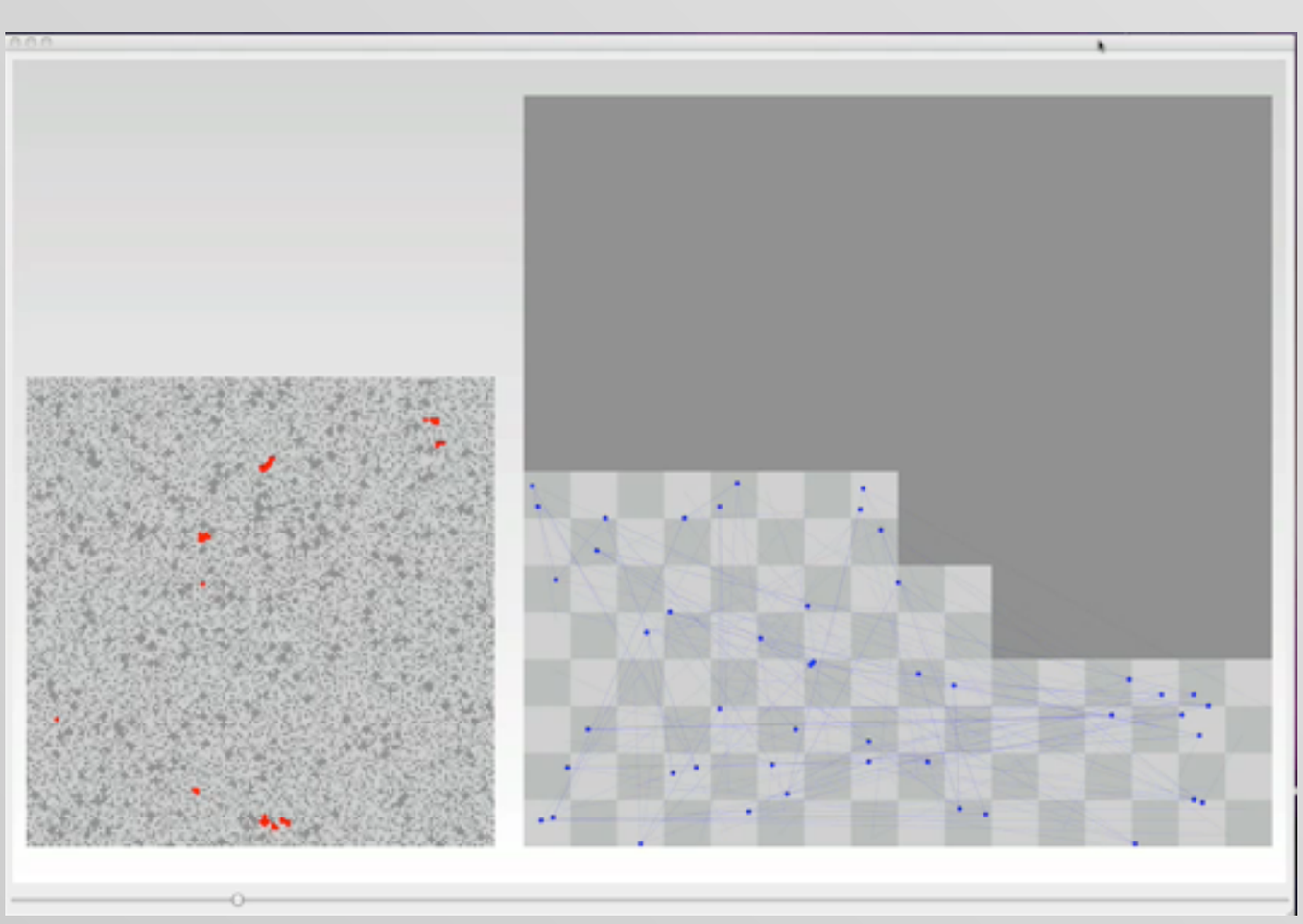
- Sensor net data collection, analysis, & vis
- Can we predict HPC node failure in time to react? Can we partition nodes by failure rate?





Megatux

- Can we emulate a large-scale botnet?
- How can we quantify interactions to aid detection and membership estimation?



~1M VMs to date
2-800 VMs per node
LGuest
Sandpile simulation
Soon: virus emulation

SICAIDA

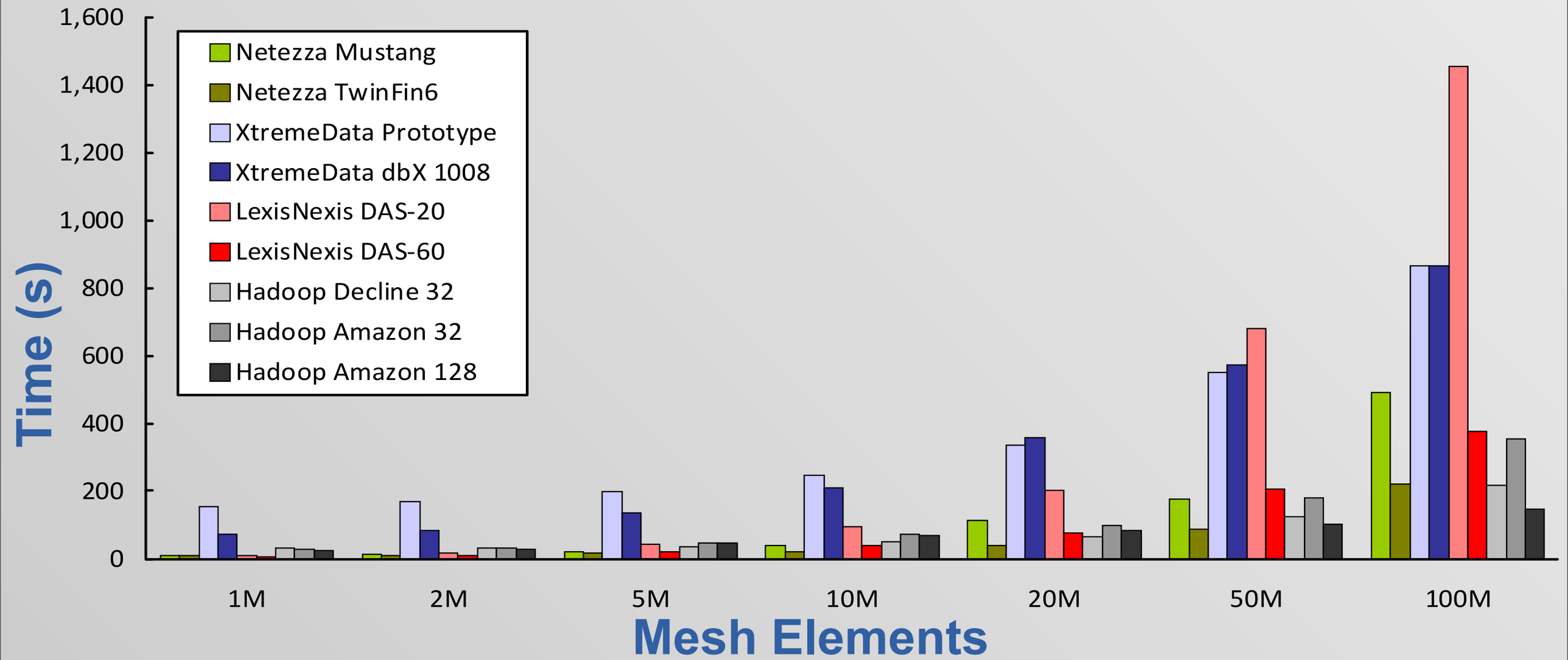
Architectures for Large Data

- Massive datasets plague many communities
  - Scientific Simulation: Petascale = 10s to 100s TBs/simulation
  - Observational Science: Large Synoptic Survey Telescope = 18 TB/night
- Out-of-Core data analysis has always been challenging
  - Performance depends on storage system
  - Reliability: daily disk failures, data rot, the RAID race
  - APIs: How do we write analysis functions that scale?
- Industry: Data Warehouse Appliances (DWAs)
  - Massive arrays of compute/storage blades
  - Data-parallel languages: SQL, MapReduce
  - System-level handling of reliability
- Can we leverage DWAs in scientific applications?
  - Storage-Intensive Computing Architectures for In-situ Data Analysis

SICAIDA

Mesh Analysis on Data Warehouse Appliances

- Ported two mesh analysis algorithms to multilpe platforms
  - Traditional SQL Parallel Database: Netezza, XtremeData
  - "NoSQL" Platforms: LexisNexis DAS, Hadoop (Local + Amazon)



Threat Recognition

- Collection of images from multiple systems
- Developed common data format for image data with XML based metadata
- Researchers outside of Sandia to use images for ATR development
- Sandia to test and evaluate algorithms
- Sponsored by DHS S&T Explosives Division