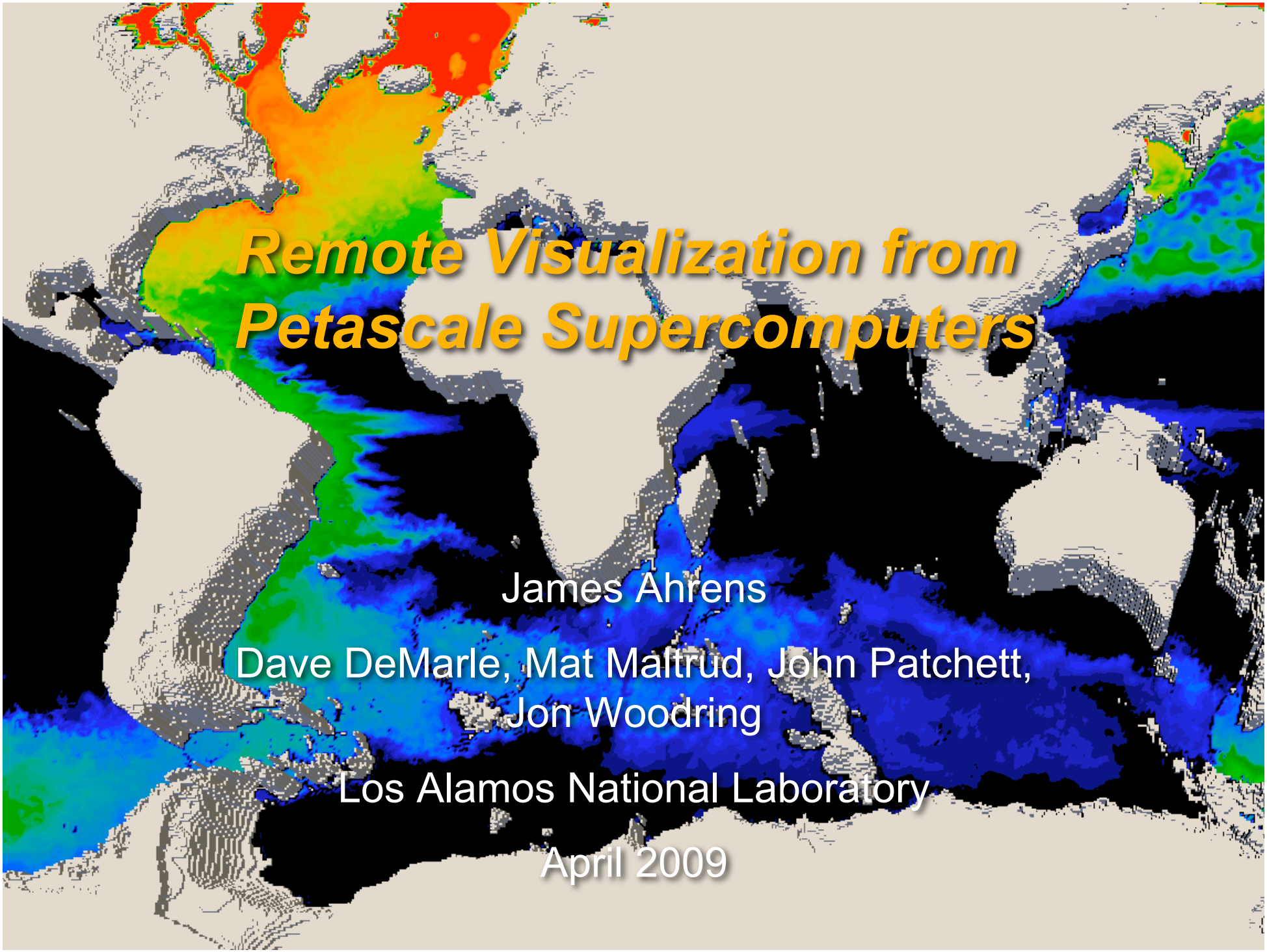


Remote Visualization from Petascale Supercomputers

James Ahrens

Dave DeMarle, Mat Maltrud, John Patchett, Jon Woodring

Los Alamos National Laboratory

April 2009

Remote visualization is an extremely challenging and important problem to solve for SC scientists

- **The goal of this project is to improve the state-of-art in visualization and analysis of massive data at remote supercomputing sites.**
  - Scientists are limited in their access to their remote data by network bandwidth and performance issues.

- **Approach and motivation**
- **Visualization**
  - Exploratory, interactive process
- **Driving problems**
  - Scale and distance
- **Systems-based approach**
  - Framework and abstractions

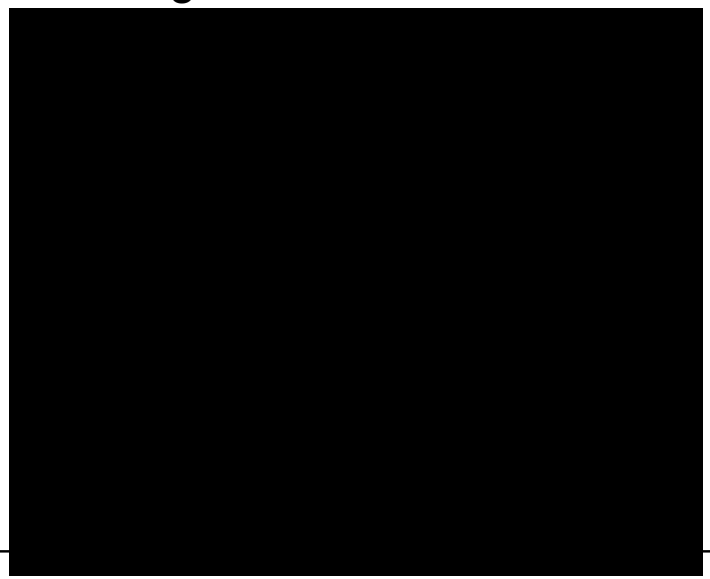
Use trends in supercomputing to guide visualization solutions

Prefix	Mega	Giga	Tera	Peta	Exa
10^n	10^6	10^9	10^{12}	10^{15}	10^{18}
Technology	Displays, networks →		Data sizes and machines →		

Visualization research with a real-world impact -- Addressed the large scale data visualization problem

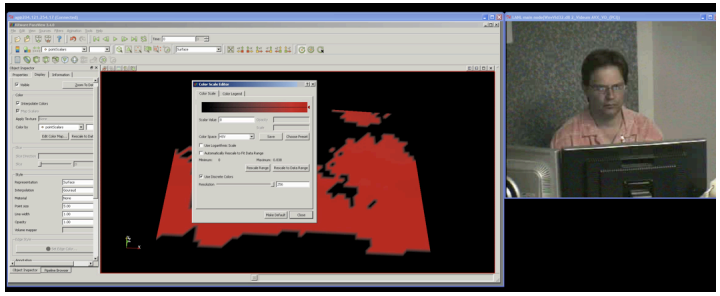
- **A decade ago - Large scale data, no visualization solutions**
- **Los Alamos/Ahrens led project to go:**
  1. From VTK - An open-source object-oriented visualization toolkit - www.vtk.org
  2. To Parallel VTK
  3. To ParaView - An open-source, scalable visualization application - www.paraview.org

- **Key concepts**
  - Streaming is the incremental processing the data as pieces
  - Streaming enables parallelism
    - Pieces processed independently
  - Applied to all operations in the toolkit
    - Contouring, cutting, clipping, analysis
  - Advantage of systems-approach over algorithms



Remote visualization – Real-world Office of Science application requirements

- **Mat Maltrud works at LANL on the Climate team and runs climate simulations at Oak Ridge**
  - Mat is responsible for generating and analyzing the simulations

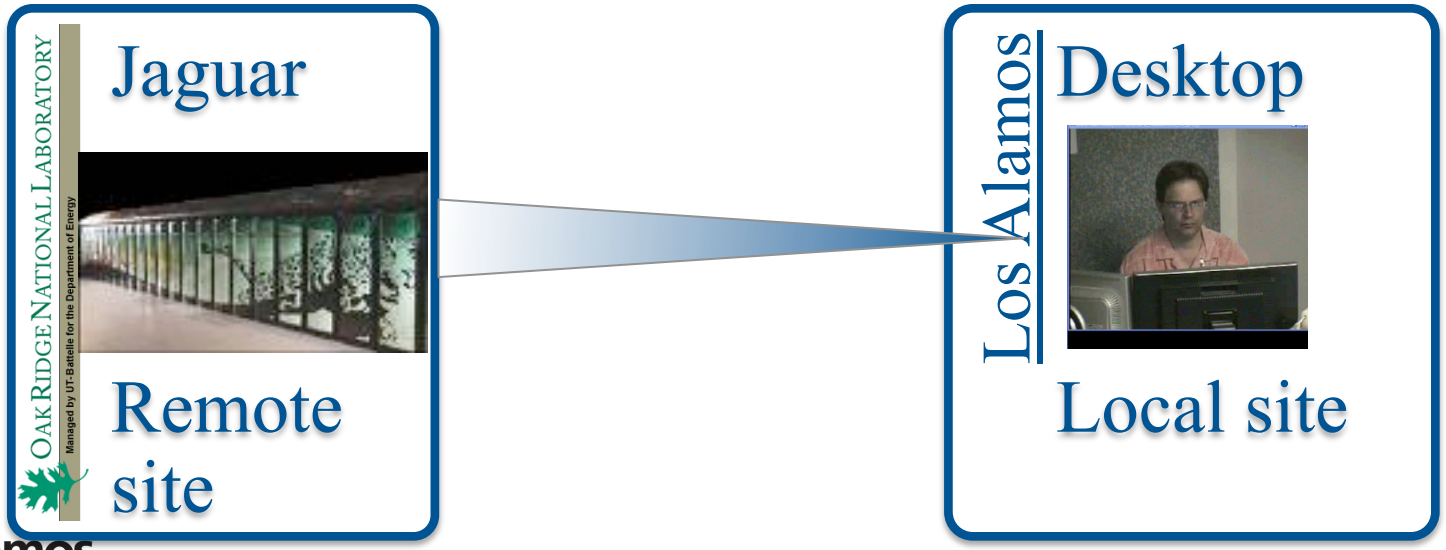


- **Requirements**
  - Using 100 TFLOPS / ORNL machine
    - 6 fields at 1.4GB each 20x a day
      - 3600 x 2400 x 42 floats
    - Transfer to LANL would take >74 hours
    - ~5 Mbps from LANL to/from ORNL's
    - Transferring all the data from ORNL to LANL will not work!
  - 250 TFLOPS
    - 12 fields
  - 1 PFLOP
    - 24 fields & 40x a day
    - 740 hours to transfer!

Overview of remote visualization approaches

- **Transfer images, visualization and rendering at remote site**
  - Advantages – images are low bandwidth
  - Disadvantages – always dependent on remote server and network

- **Transfer subsets of data or geometry, visualize and render locally**
  - Advantages – fast local visualization and rendering, not as dependent on the remote server
  - Disadvantages – transfer of subsets of data or geometry can be slow



Remote Visualization – Initial Results

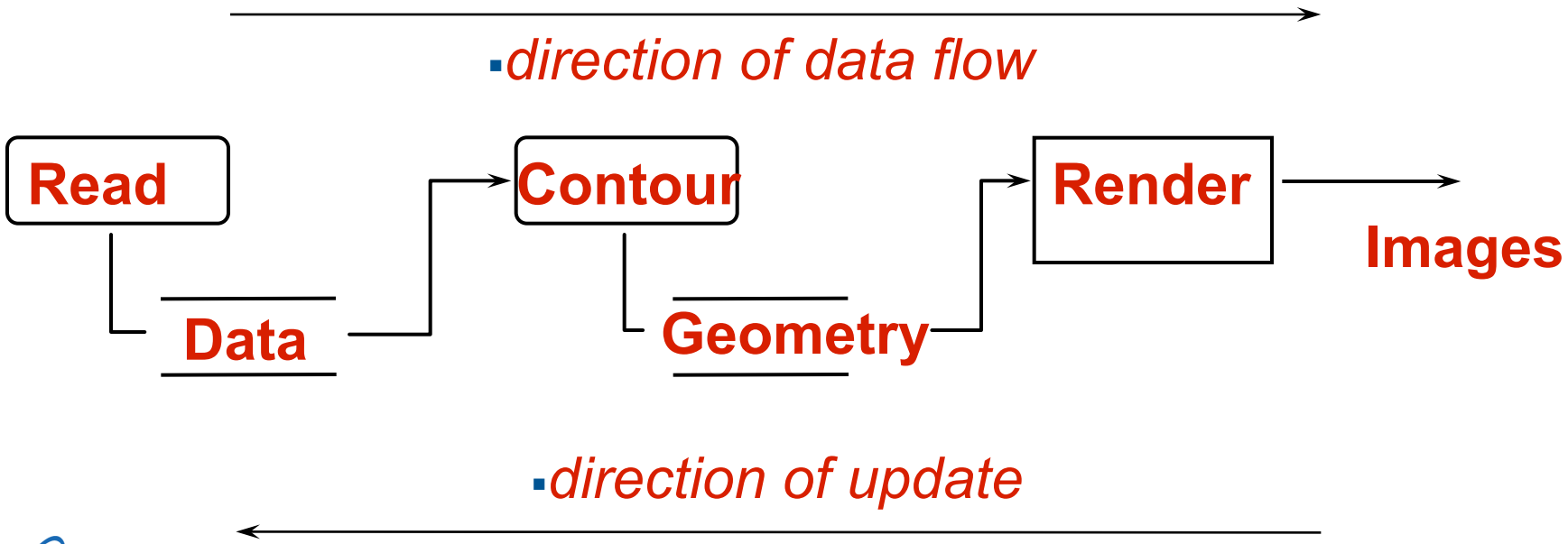
- **Remote Visualization at ORNL**
  - Port ParaView to Jaguar XT4/5
  - Tested to 2048 processors
  - Goal: Interactive visualization
    - Currently focus on image transfer techniques

Our Remote Visualization Approach - Improving remote visualization via prioritization

- **Prioritize data and geometry transfers by sending important pieces of information over the network first.**
- **Once this information is local, scientists can do visualization and analysis locally, offloading these tasks from flagship supercomputing resources.**
- **Progressively refine visualized result over time**

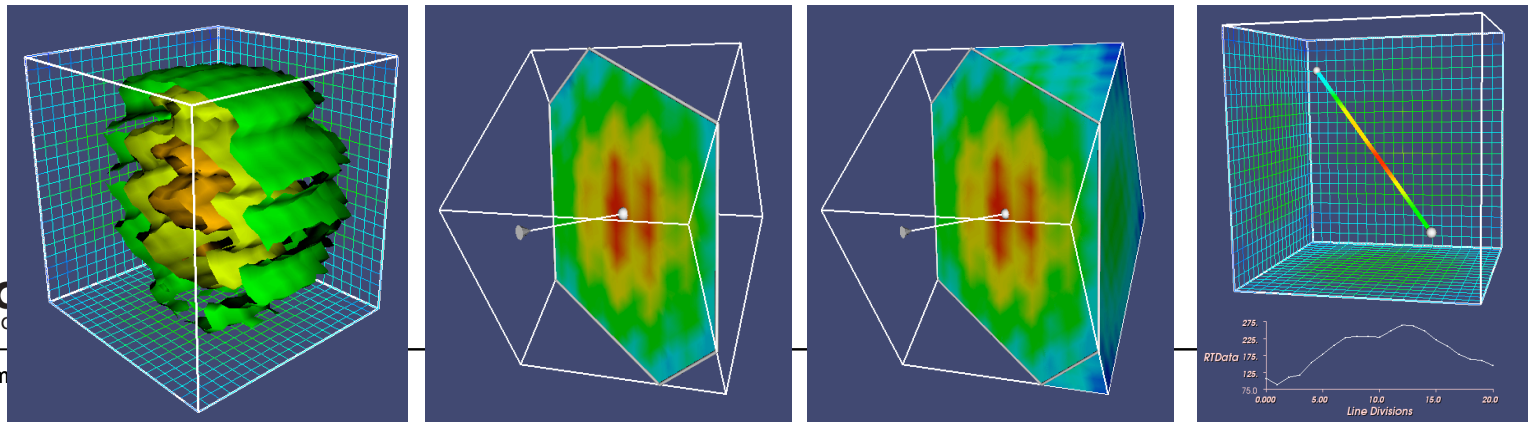
Background -
The data-flow visualization pipeline

- A sequence of process objects that operate on data objects to generate geometry that can be rendered by the graphics engine
- Typically a data reduction process
- Data, meta-data, execution model



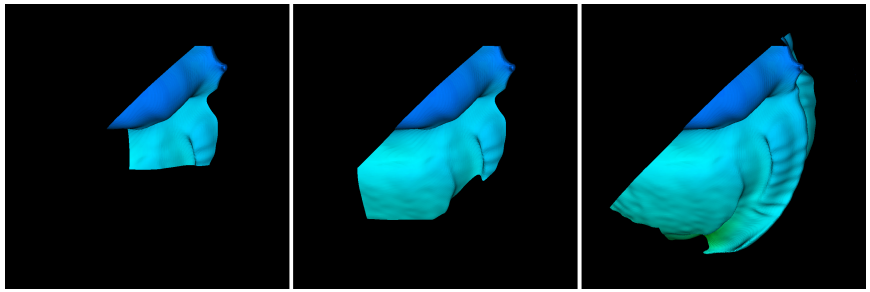
Our Remote Visualization Approach – Streaming with culling and prioritization

- Each module in the pipeline can cull and prioritize…
  - **Culling – remove pieces**
    - Based on spatial location
      - Spatial clipping
      - Cutting
      - Probing
      - Frustum culling
      - Occlusion culling
    - Based on data value
      - Contouring
      - Thresholding
  - **Prioritization – order piece processing**
    - Based on spatial location
      - View dependent ordering
    - Based on features
    - Based on user input

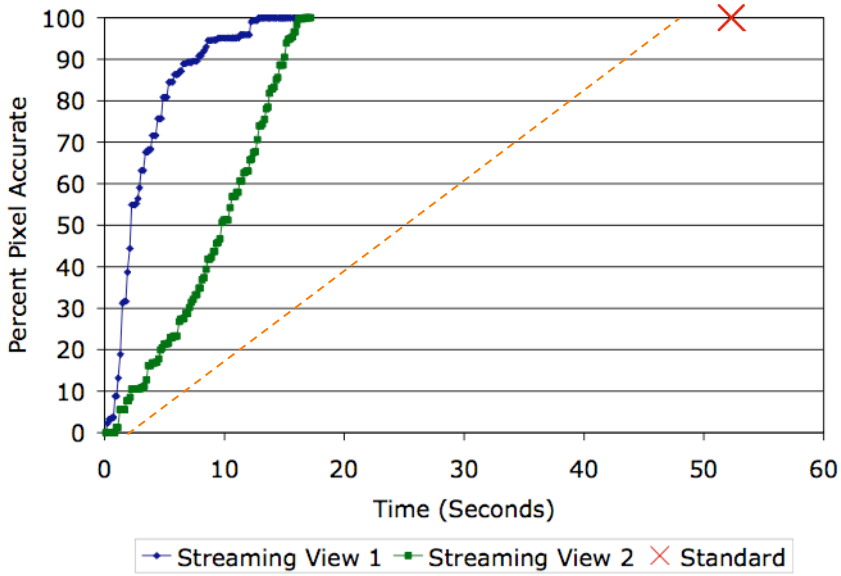


Our Remote Visualization Approach – Benefits of prioritization and progressive refinement

- **Results displayed progressively**
  - Finished in 2%, 4% and 8% of the time it takes the standard architecture to generate the final image (25%, 50%, 75% pixel accurate)



- Prioritized progressive streaming provides accurate results quickly



Our Remote Visualization Approach - Improving remote visualization via multi-resolution

- **Pieces of data may have varying resolutions controlled by the prioritization.**
  - This supports making tradeoffs between improved resolution or improved performance.

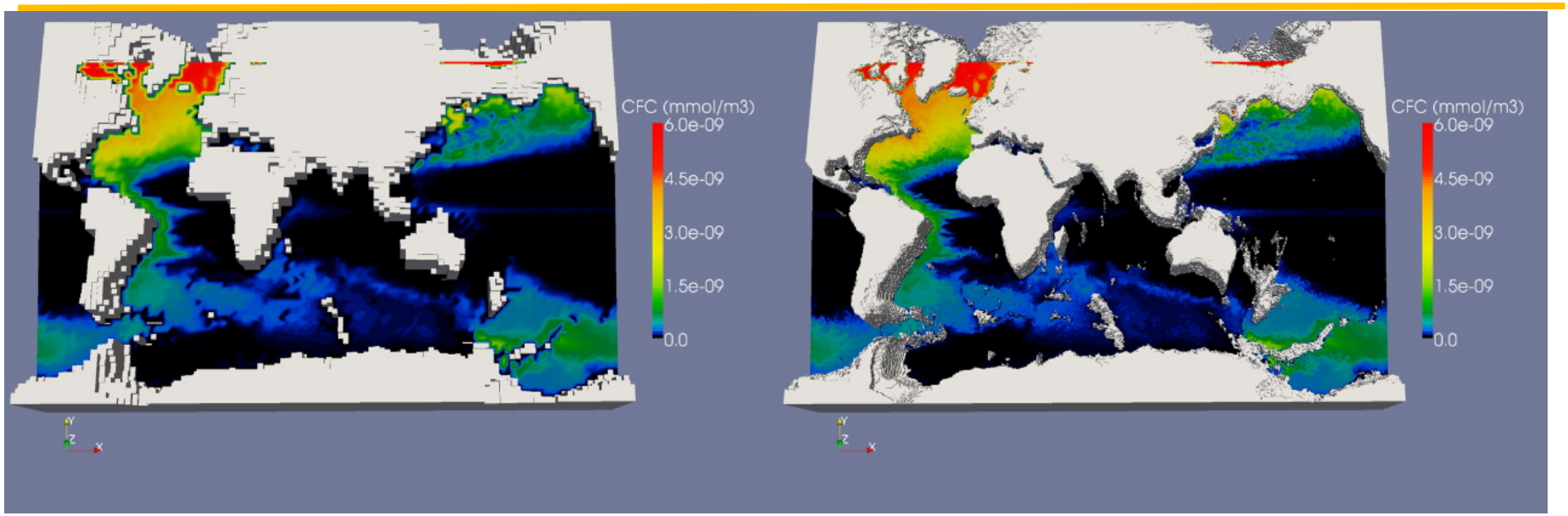
- **Sample from disk with strided reads of structured grid data**
  - Read in coarse-resolution initially
  - In prioritized areas, read in fine resolution

- **Does not require special formatting/data structures on disk**
  - Read from binary Fortran dumps
  - Creates AMR representation of data in visualization framework from disk

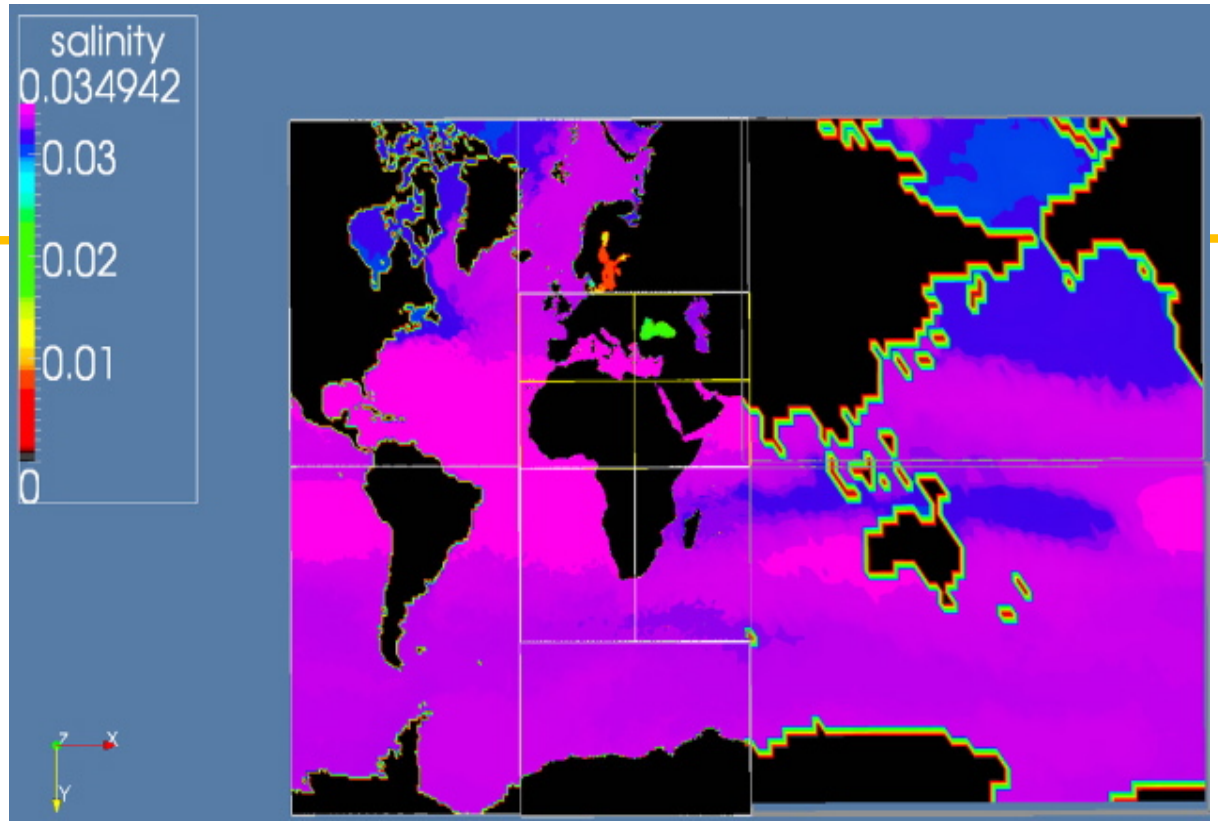
Our Remote Visualization Approach – Summary

- **Data reduction**
  - ① Subsetting the data and culling
  - ② Sampling the data from disk to create multi-resolution representation
  - ③ Visualization and analysis modules in pipeline - highlighting property of the dataset
    - For example - isosurface, cut plane, clipping
- **Prioritization**
  - Which piece, what resolution, what visualization operation…
  - Combine multiple priorities
  - Processing most important data first
- **Continuously improve visualized results over time**
- **Think of progressive refinement approach of 2D images on the web…**
  - Our solution provides a prioritized 3D progressive refinement approach that works within a full-featured visualization tool…

Using our remote techniques to visualize CFC11s from a POP global 1/10-degree model run generated at ORNL

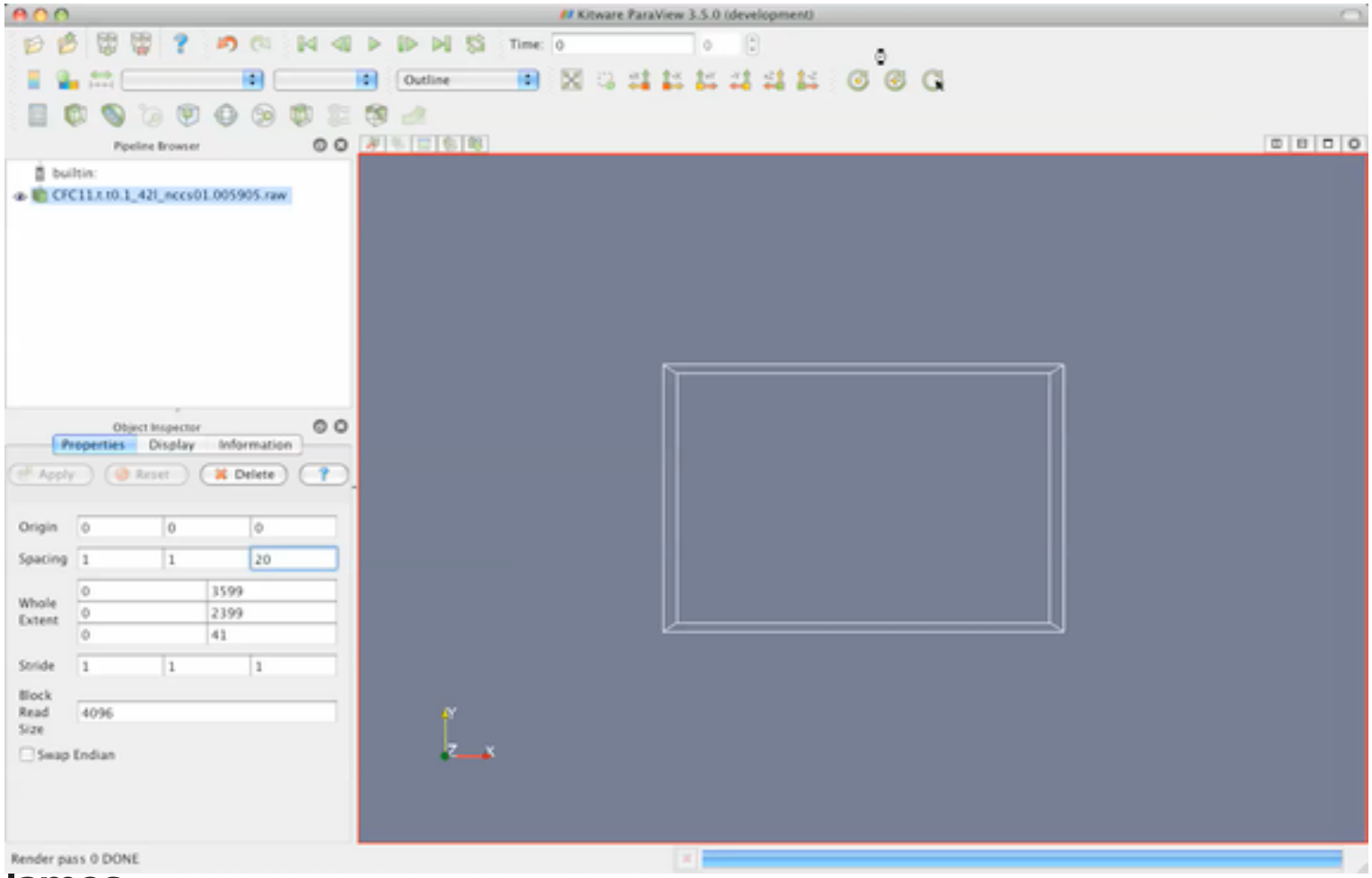


- **A LANL climate scientist made these images by first prioritizing the creation of a low-resolution result. Due to the efficient performance when visualizing the low-resolution data they were able to quickly create the type of visualization they wanted. Once the type of visualization was specified the scientist changed the prioritization to produce a high-resolution result.**

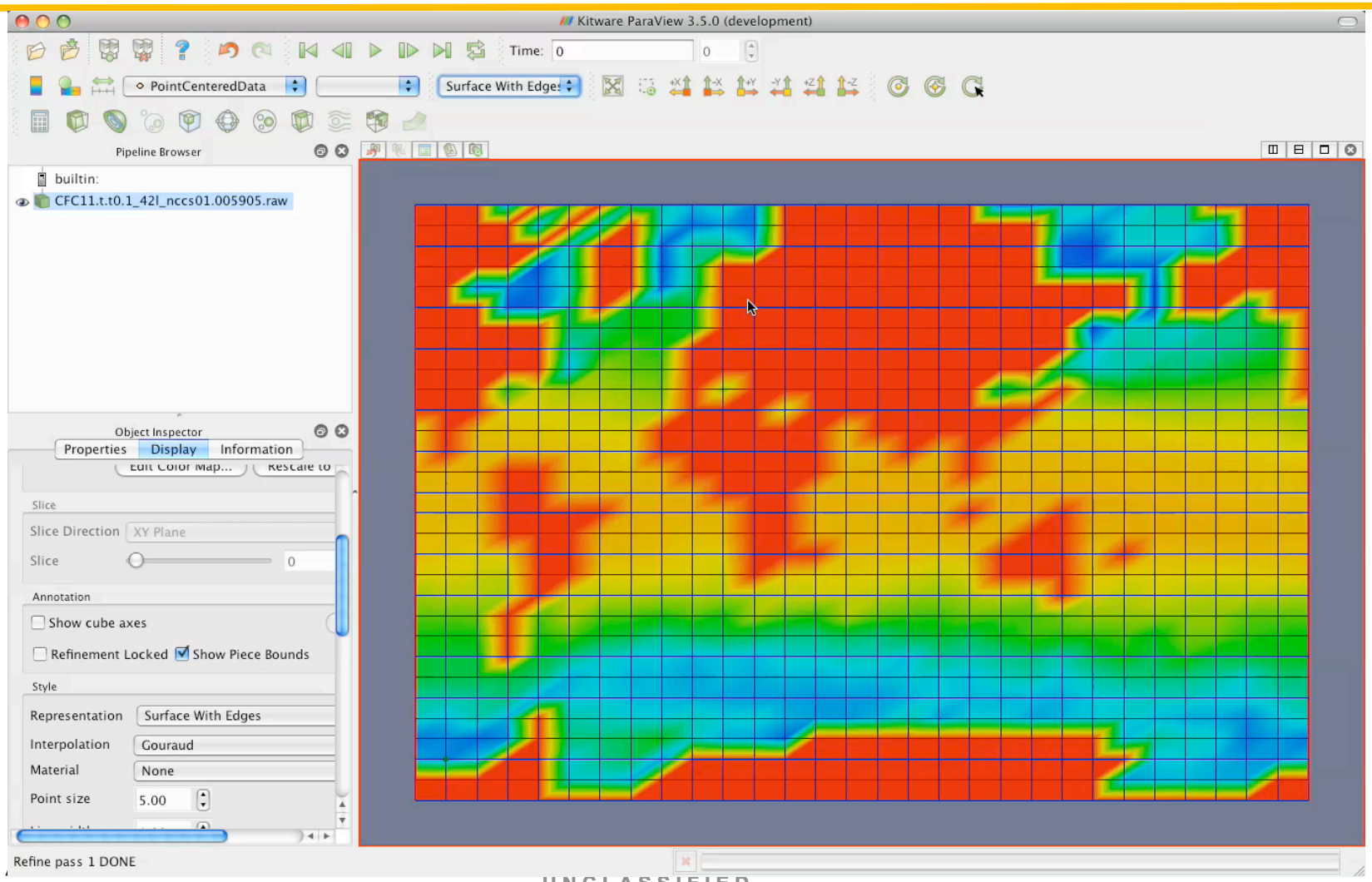


A more complex use of prioritization and multi-resolution techniques. The prioritization is set to refine the area of the Mediterranean Sea. The visualization tool, reads coarse-resolution representations in some areas (such as Asia) while reading the finest resolution data in the area of interest.

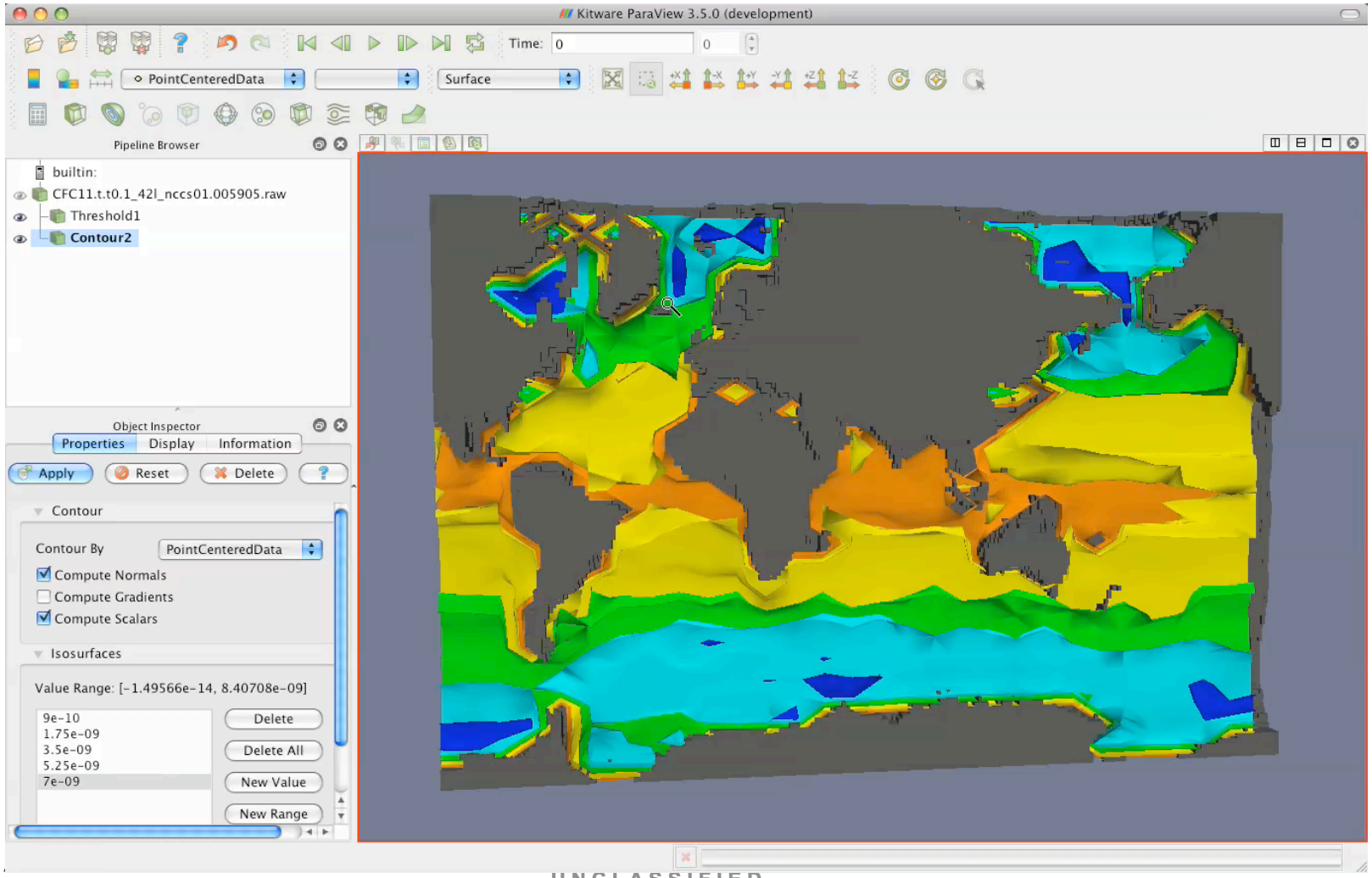
Remote Visualization – Results



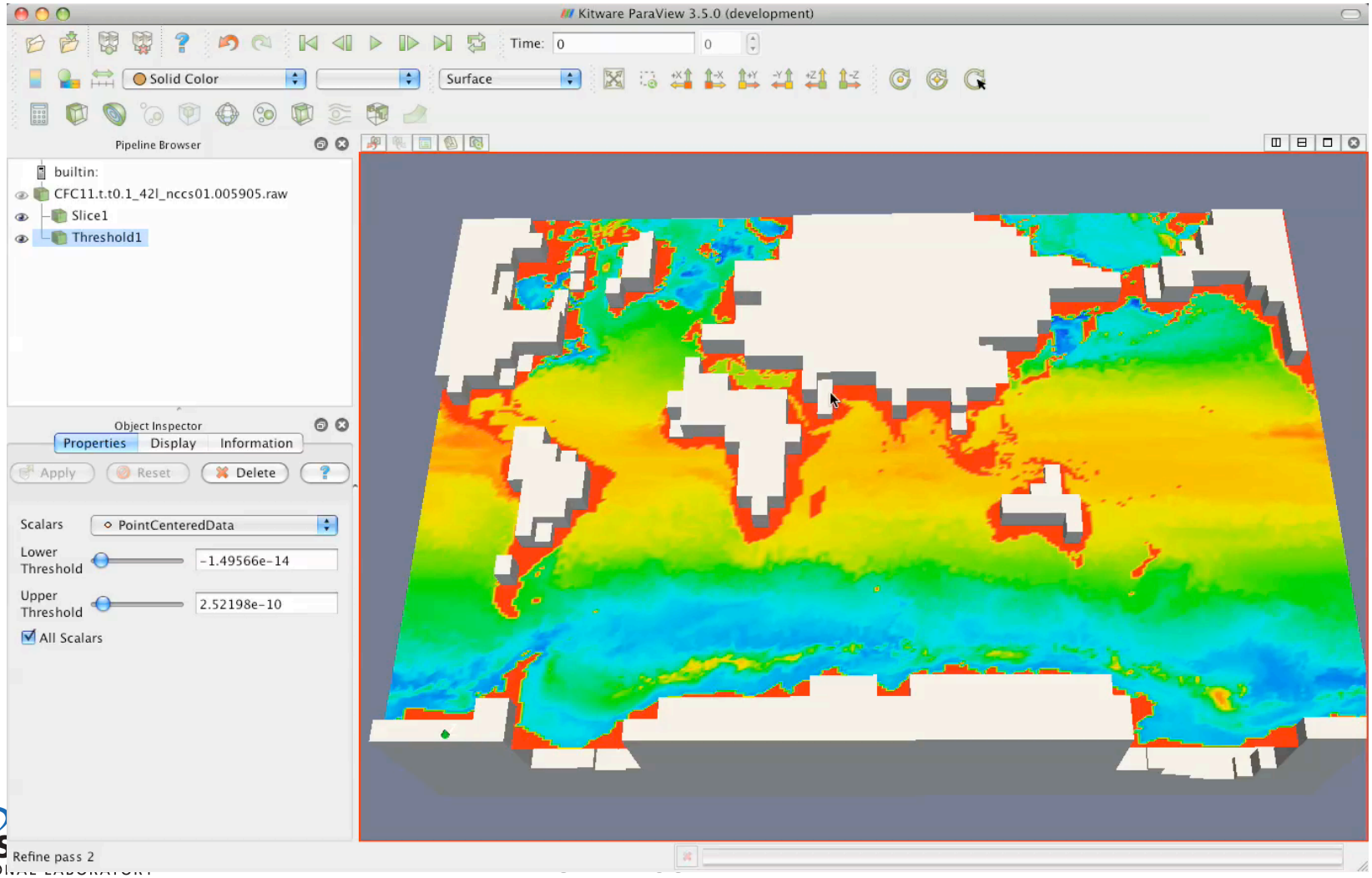
Remote Visualization – Results



Remote Visualization – Results



Remote Visualization – Results



Our approach directly addresses the fundamental ultrascale visualization challenge

- By creating a visualization framework for data reduction and prioritization
  - Provides an approach for the petascale/exascale visualization problem

Prefix	Mega	Giga	Tera	Peta	Exa
10^n	10^6	10^9	10^{12}	10^{15}	10^{18}
Technology	Displays, networks →		Data sizes and machines →		

Conclusions

- **Cutting edge R&D in remote visualization and analysis techniques**
- **Available as part of ParaView, a scalable open-source visualization tool**
- **We have focused on a specific real-world Office of Science problem for testing our work**
  - Helping LANL climate modelers analyze massive datasets created on Jaguar at ORNL