Remote and Collaborative Visualization at Scale — Gaining Insight Against Insurmountable Odds

Kelly Gaither
Director of Visualization
Senior Research Scientist
Texas Advanced Computing Center
The University of Texas at Austin
Every two days we create as much data as we did from the beginning of mankind until 2003!

Sources: Lesk, Berkeley SIMS, Landauer, EMC, TechCrunch, Smart Planet
How Much is an Exabyte?

• 1 Exabyte = 1000 Petabytes -> approximately 500,000,000,000,000 pages of standard printed text
• It takes one tree to produce 94,200 pages of a book
• Thus it will take 530,785,562,327 trees to store an Exabyte of data

Sources: http://www.whatsabyte.com/ and http://wiki.answers.com
How Much is an Exabyte?

• In 2005, there were 400,246,300,201 trees on Earth
• We can store .75 Exabytes of data using all the trees on the entire planet.

Sources: http://www.whatsabyte.com/ and http://wiki.answers.com
Texas Advanced Computing Center (TACC)

*Powering Discoveries that Change the World*

- **Mission:** Enable discoveries that advance science and society through the application of advanced computing technologies
- **Over 12 years in the making, TACC has grown from a handful of employees to over 120 full time staff with ~25 students**
TACC Visualization Group

• Provide resources/services to local and national user community.

• Research and develop tools/techniques for the next generation of problems facing the user community.

• Train the next generation of scientists to visually analyze datasets of all sizes.
TACC Visualization Group

- 9 Full Time Staff,
- 2 Undergraduate Students, 3 Graduate Student
- Areas of Expertise: Scientific and Information Visualization, Large Scale GPU Clusters, Large Scale Tiled Displays, User Interface Technologies
Maximizing Scientific Impact

Image: Greg P. Johnson, Romy Schneider, TACC

Image: Adam Kubach, Karla Vega, Clint Dawson

Image: Karla Vega, Shaolie Hossain, Thomas J.R., Hughes

Scientific and Information Visualization
A computational tool-set was developed to support the design and analysis of a catheter-based local drug delivery system that uses nanoparticles as drug carriers to treat vulnerable plaques and diffuse atherosclerosis.

The tool is now poised to be used in medical device industry to address important design questions such as, "given a particular desired drug-tissue concentration in a specific patient, what would be the optimum location, particle release mechanism, drug release rate, drug properties, and so forth, for maximum efficacy?"

The goal of this project is to create a visualization that explains the process of simulating local nanoparticulate drug delivery systems. The visualization makes use of 3DS Max, Maya, EnSight and ParaView.
Volume Visualization of Tera-Scale Global Seismic Wave Propagation

Carsten Burstedde, Omar Ghattas, James Martin, Georg Stadler and Lucas Wilcox, ICES; Greg Abram, TACC

- Modeling propagation of seismic waves through the earth helps assess seismic hazard at regional scales and aids in interpretation of earth's interior structure at global scales.
- Discontinuous Galerkin method used to for numerical solution of the seismic wave propagation partial differential equations.
- Visualization corresponds to a simulation of global wave propagation from a simplified model of the 2011 Tohoku earthquake with a central source frequency of 1/85 Hz, using 93 million unknowns on TACC’s Lonestar system.
Texas Pandemic Flu Toolkit

Greg Johnson, Adam Kubach, TACC; Lauren Meyers & group, UT Biology; David Morton & group, UT ORIE.
Stellar Magnetism
Greg Foss, TACC; Ben Brown, University of Wisconsin, Madison

- A Sun-like star undergoes magnetic cyclic reversal shown by field lines.
- Shifts in positive and negative polarity demonstrate large-scale polarity changes in the star.
- Wreath-like areas in the magnetic field may be the source of Sun spots.
- Terabytes of data to mine through and visualize.
Remote Visualization at TACC

A Brief History
History of Remote Visualization at TACC

2004
- Maverick
  - Sun Fire E25K

2008
- Spur – 8 node Sun
  - AMD NVIDIA cluster
- Ranger – 8 node
  - Sun AMD NVIDIA subsystem

2012
- Longhorn – 256 node Dell Intel NVIDIA cluster
- Lonestar – 16 node
  - Dell Intel NVIDIA subsystem
- Stampede – 128 node
  - Dell Intel NVIDIA subsystem

Same data center

Same interconnect fabric

2014
- Longhorn replacement cluster
TACC Solution: Integrate Visualization Capability into Cluster

- Keep data in same data center, or on same machine
- Spur – integrated into Ranger
  - 8 nodes, 32 GPUs, 1 TB aggregate RAM
  - shares interconnect and file system

- Longhorn – in Ranger machine room
  - 256 nodes, 512 GPUs, 13.5 TB aggregate RAM
  - local parallel file system, high-bandwidth mount to Ranger

- Lonestar – GPU nodes integrated into system
  - 16 nodes, 32 GPUs, 384 GB aggregate RAM

- Stampede – GPU nodes integrated into system
  - 128 nodes, 128 GPUs in vis queues, 16 nodes, 32 GPUs in largemem
  - Working to utilize Xeon Phis for vis and rendering too

For larger data, move vis back to HPC cluster!
Longhorn Usage Modalities:

- **Remote/Interactive Visualization**
  - Highest priority jobs
  - Remote/Interactive capabilities facilitated through VNC
  - Run on 3 hour queue limit boundary

- **GPGPU jobs**
  - Run on a lower priority than the remote/interactive jobs
  - Run on a 12 hour queue limit boundary

- **CPU jobs with higher memory requirements**
  - Run on lowest priority when neither remote/interactive nor GPGPU jobs are waiting in the queue
  - Run on a 12 hour queue limit boundary
Longhorn Visualization Portal
portal.longhorn.tacc.utexas.edu
Stampede Architecture

High Fidelity Visualization of Scientific Data

- Presenting at 3:15pm today at the HPC round table in the Grand Hyatt
- Also being presented in the Intel booth on Wednesday at 1:30 pm
Current Community Solution: Fat Client – Server Model

- Geometry (or pixels) sent from server to client, user input and intermediate data sent to server.
- Data traffic can be too high for low bandwidth connections.
- Connection options often assume single shared-memory system.
TACC Solution:
Thin Client – Server Model

- Run both client and server on remote machine
- Minimizes required bandwidth and maximizes computational resources for visualization and rendering
- Can use either a remote desktop or a web-based interface

Geometry, images and client all remain on server

Only pixels, mouse and keyboard sent between client and server
Visualization Use Cases

2004 - 2013

Image/Animation Generation - Exploration/Knowledge Discovery

- Batch
  - Non-interactive exploration

- Post-processing
  - Interactive exploration asynchronous from data generation

- In Situ
  - Interactive exploration in conjunction with data generation

- Visual Analytics
  - Exploration in conjunction with visual analysis and statistical/matheamtical analysis

TACC
Large-Scale Tiled Displays
Stallion

- 16x5(15x5) tiled display of Dell 30-inch flat panel monitors
- 328M(308M) pixel resolution, 5.12:1(4.7:1) aspect ratio
- 320(100) processing cores with over 80GB(36GB) of graphics memory and 1.2TB(108GB) of system memory
- 30 TB shared file system
Lasso
Multi-Touch Tiled Display

• 3x2 tiled display (1920x1600) – 12M Pixels
• PQ Labs multi-touch overlay, 32 point 5mm touch precision
• 11 mm bezels on the displays
Vislab Numbers

• Since November 2008, the Vislab has seen over 20,000 people come through the door.

• Primary Usage Disciplines – *Physics, Astronomy, Geosciences, Biological Sciences, Petroleum Engineering, Computational Engineering, Digital Arts and Humanities, Architecture, Building Information Modeling, Computer Science, Education*
Vislab Stats

Vislab resource allocation per activity type

Vislab usage per area

- **Natural Sciences**: 30%
- **Humanities**: 15%
- **Engineering**: 20%
- **Computational Sciences**: 25%
- **Architecture**: 5%
- **Other**: 10%
Sample Use Cases – Biological Sciences

- **Research Motivation:**
  understand the structure of the neuropil in the neocortex to better understand neural processes.

- **People:** Chandra Bajaj et. al. UT Austin

- **Methodology:** Use Stallion’s 328 Mpixel surface to view neuropil structure.
Sample Use Cases – Architecture/BIM

- **Motivation**: developing new tools for building modeling and construction using large multi-touch display surfaces
- **People**: Fernanda Leite, Li Wang, UT Austin
- **Methodology**: develop new CAD interaction methods using gesture and collaborative multi-touch to increase productivity in construction and engineering design.
Sample Use Cases - Humanities

- **Motivation**: use advanced visualization resources as a tool for the arts and humanities
- **People**: TACC Vislab, UT Austin Digital Humanities
- **Methodology**: create software to allow non-trained users the ability to take advantage of distributed systems and graphics technology. Develop Massive Pixel Environment (MPE) and DisplayCluster.
Sample Use Cases – Virtual Worlds

- **Motivation**: digitally reconstructing the past, and live life through another’s eyes
- **People**: Janine Barchas, UT Austin
- **Methodology**: create an immersive model of the iconic British art exhibit (The Reynolds Retrospective), which was a turning point in the history of modern exhibit practices.
DisplayCluster

- A cross-platform software environment for interactively driving tiled displays
- Features:
  - Media display (Images (up to gigapixels in size, movies / animations
  - Pixel streaming (Real-time desktop streaming for collaboration / remote vis)
  - Scriptable via Python interface
  - Multi-user interaction (iPhone / iPad / Android devices, Joysticks, Kinect (in development))
  - Implementation (MPI, OpenGL, Qt, FFMPEG, Boost, TUIO, OpenNI, …)
- Short demonstration: http://www.youtube.com/watch?v=JwTwa46BhcU
Most Pixels Ever: Cluster Edition

- Create interactive multimedia and data visualizations that span multiple displays, at very high resolutions
- Enables extremely high resolution Processing sketches
- Licensed and available for download on GitHub:
  - https://github.com/TACC/MassivePixelEnvironment
Using Visualizations for Knowledge Discovery

• As data scales, it becomes increasingly apparent that visualization or visual analysis becomes key to knowledge discovery

• Managing this bottleneck requires us to have an understanding of:
  – Remote and collaborative visualization (data manipulation)
  – High resolution displays (data synthesis)
Lessons Learned Over the Past 12 Years

• Close collaborations with the science partners are key.
• Minimize data transfers if possible.
• Scale resources effectively based on use cases.
• Easy accessibility to and interaction with technologies encourages participation from diverse communities.
Thoughts Towards Exascale:

• Data will get larger and more unwieldy – we will stop moving it around
• High performance computing environments will become high performance science environments that provide computing and analytics
• Rendering will continue to get less and less expensive.
• We will see a real blend in hardware to support high performance computing and interactive visualizations.
Thank You

Questions?
VISTech Workshop: Visualization Infrastructure & Systems Technology at SC ’13

• Half day workshop aimed at discussing the intersection between human perception and large-scale visual analysis through the study of visualization interfaces and interactive displays.

• Organizers: Kelly Gaither, Brandt Westing, TACC; Jason Leigh, EVL; Kuester Falko, Calit2, UCSD; Eric Wernert, Indiana University; Aditi Majumder, UC Irvine