

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

Kelly Gaither

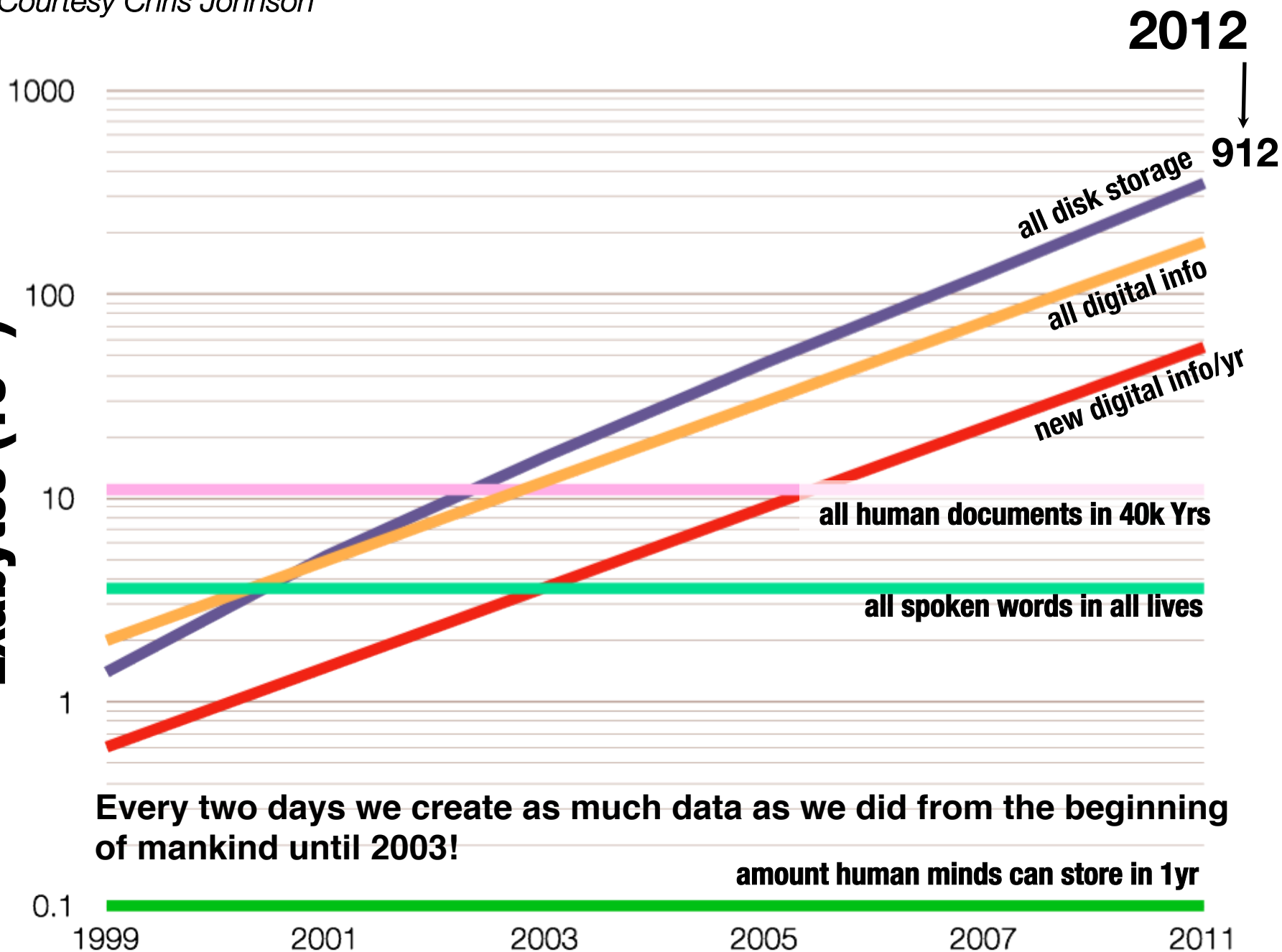
Director of Visualization

Senior Research Scientist

Texas Advanced Computing Center

The University of Texas at Austin

**Exabytes ( $10^{18}$ )**



# How Much is an Exabyte?



How many trees does it take to print out 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

Slide Courtesy Chris Johnson

# How Much is an Exabyte?



How many trees does it take to print out 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.

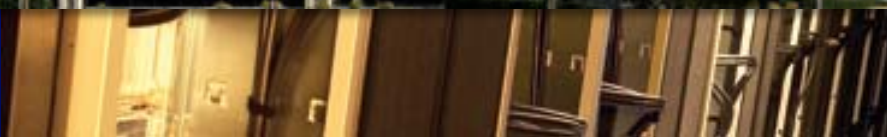
Slide Courtesy Chris Johnson



# 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.

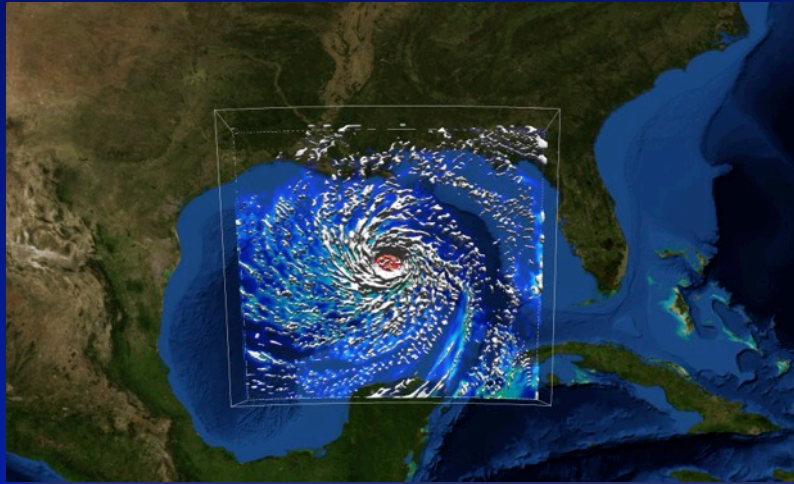
# TACC Visualization Group



- 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



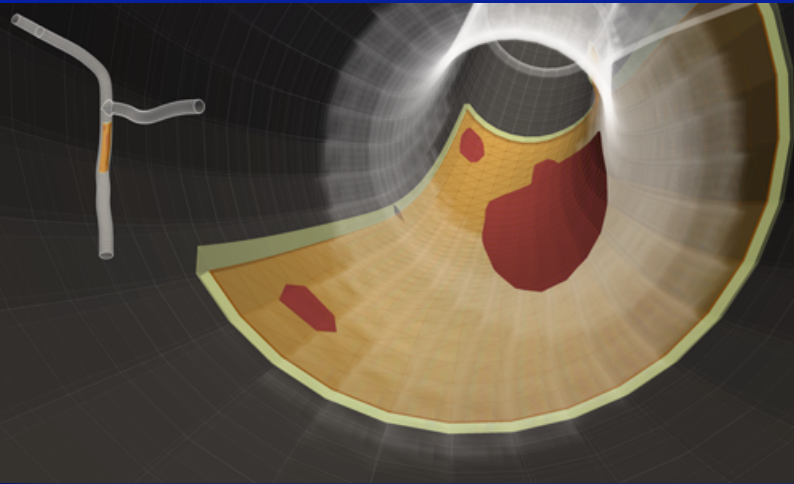
# Maximizing Scientific Impact



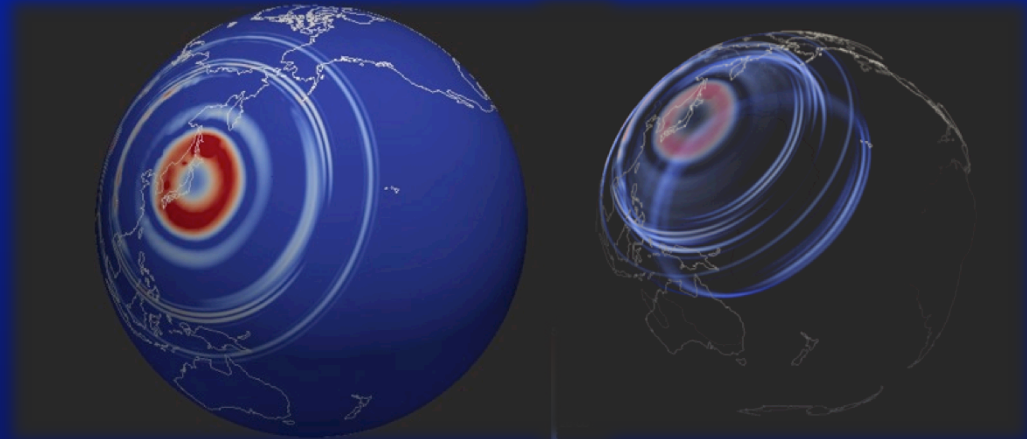
*Image: Greg P. Johnson, Romy Schneider, TACC*



*Image: Adam Kubach, Karla Vega, Clint Dawson*



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



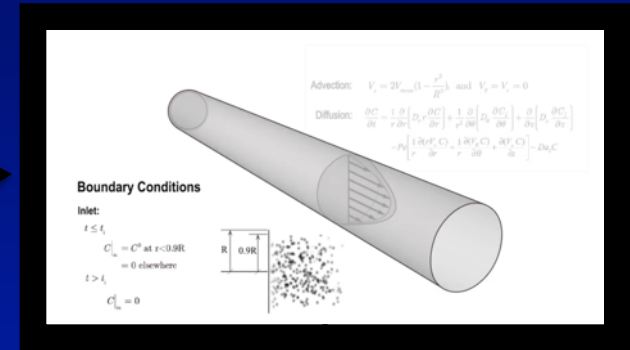
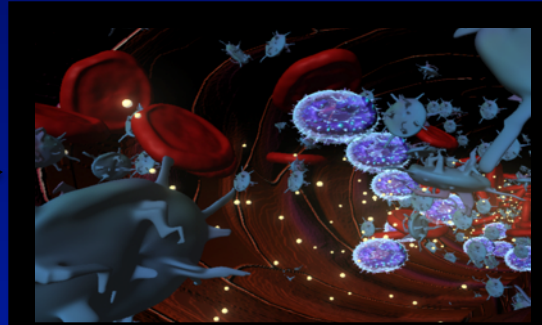
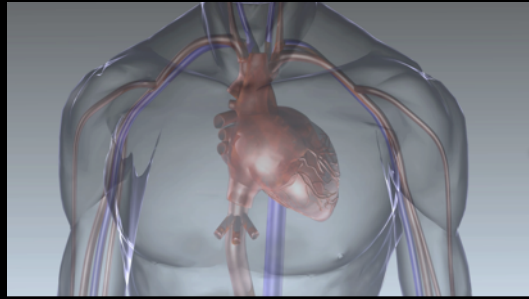
*Greg Abram, Carsten Burstedde, Georg Stadler, Lucas C. Wilcox, James R. Martin, Tobin Isaac, Tan Bui-Thanh, and Omar Ghattas*



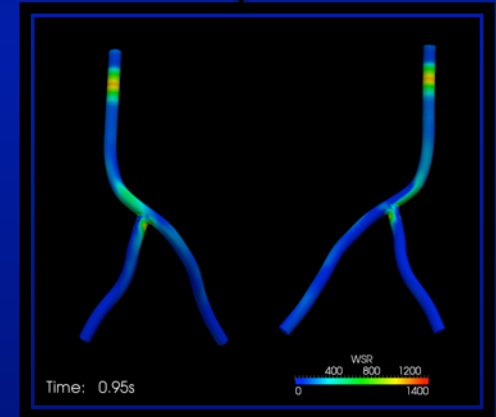
# Scientific and Information Visualization

# Coronary Artery Nano-particle Drug Delivery Visualization

*Ben Urick, Jo Wozniak, Karla Vega, TACC; Erik Zumalt, FIC; Shaolie Hossain, Tom Hughes, ICES.*



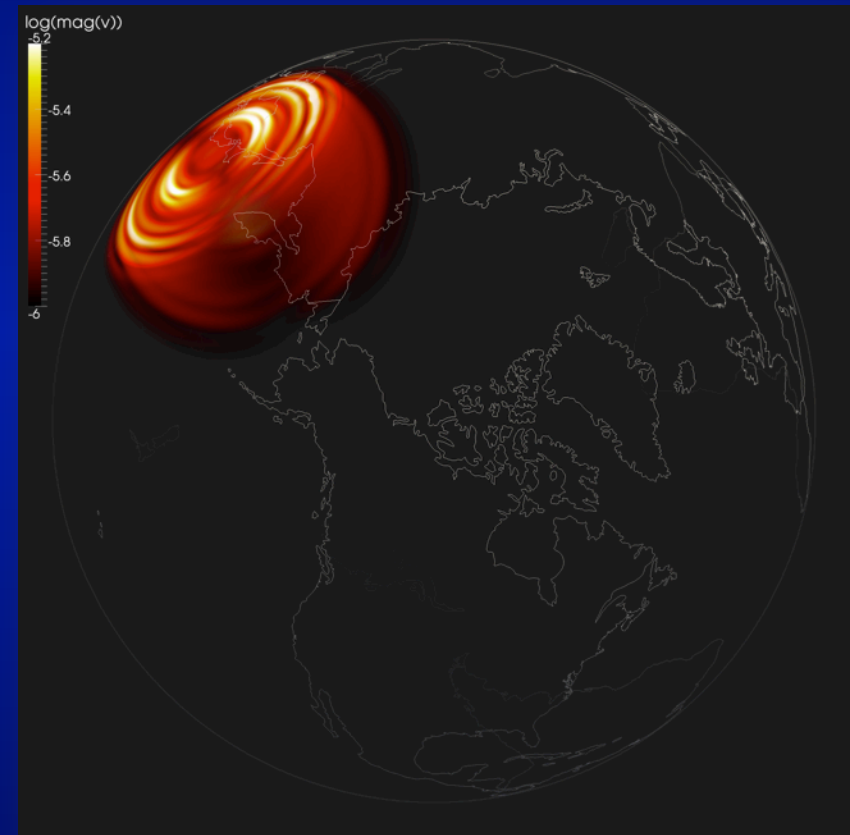
- A computational tool-set was developed to support the design and analysis of a catheter-based local drug delivery system that uses nanoparticles as drug carriers to treat vulnerable plaques and diffuse atherosclerosis.
- The tool is now poised to be used in medical device industry to address important 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

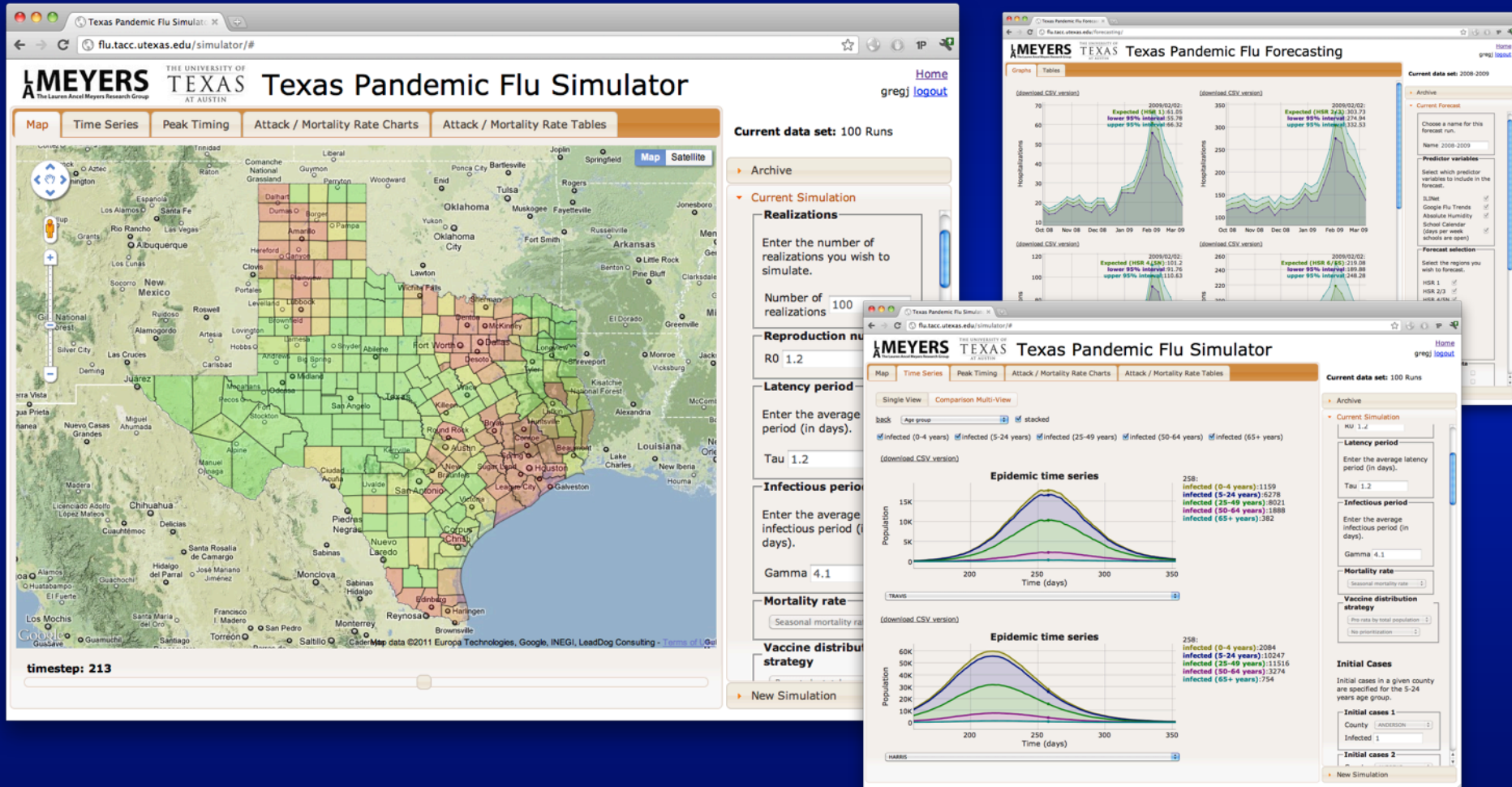
*Carsten Burstedde, Omar Ghattas, James Martin, Georg Stadler and Lucas Wilcox, ICES; Greg Abram, TACC*

- 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.

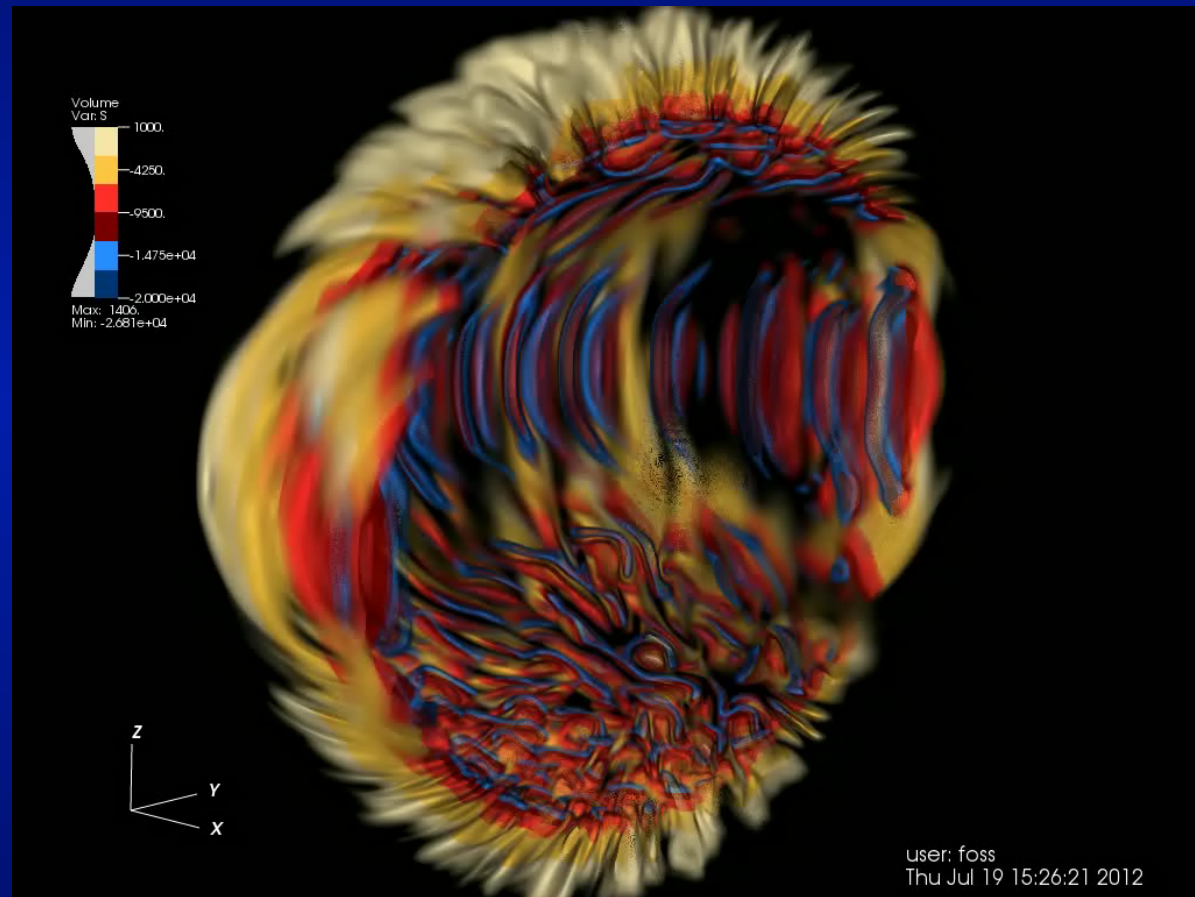




# Stellar Magnetism

*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



Maverick  
Sun Fire E25K



Spur – 8 node Sun  
AMD NVIDIA cluster



Longhorn – 256  
node Dell Intel  
NVIDIA cluster



Longhorn  
replacement  
cluster

2004

2008

2012

2014

same data center

same interconnect fabric

Ranger – 8 node  
Sun AMD NVIDIA subsystem



Lonestar – 16 node  
Dell Intel NVIDIA subsystem



Stampede – 128 node  
Dell Intel NVIDIA subsystem



# 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

For larger data, move vis back to HPC cluster!

- - 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



# 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*

The screenshot shows the Longhorn Visualization Portal homepage. At the top, there's a navigation bar with 'Home', 'Allocations', 'Resources', and 'Help'. Below this, a 'Select a Resource' section allows users to choose a resource (Longhorn), project (PECOS), and session type (VNC or EnVision guided visualization). A 'Number of nodes' dropdown is set to '1 (8 slots)'. A 'Select' button is present. Below this, an 'Available Resources' section provides details about the Longhorn cluster, including its size (2048 compute cores, 14.5 TB aggregate memory, 512 GPUs) and a link to the 'Longhorn User Guide'. A 'Queue information' section shows a pie chart indicating that 180 nodes are available out of 250 total. At the bottom, there's a table of 'ACTIVE JOBS' with columns for JOBID, JOBNAME, USERNAME, STATE, CORE, REMAINING, and STARTTIME. The table lists several jobs in various states (Running, Pending).

**Longhorn Visualization Portal**

Home | Allocations | Resources | Help

**Select a Resource**

Resource:

Project:

Session type: ☒ VNC ☐ EnVision guided visualization

Number of nodes:

Note: increasing the number of nodes will only increase performance for parallel applications (e.g. ParaView or VisIt)

**Available Resources**

Longhorn (portal.longhorn.tacc.utexas.edu), TACC's Dell XD Visualization Cluster, contains 2048 compute cores, 14.5 TB aggregate memory and 512 GPUs. Longhorn has a QDR InfiniBand interconnect and has an attached Lustre parallel file system. Longhorn is connected by 10GigE to Ranger's Lustre parallel file system thus making it more convenient to work on datasets generated on Ranger. Longhorn has 256 nodes + 2 login nodes, with 240 nodes containing 49GB of RAM, 8 Intel Nehalem cores (@ 2.5 GHz), and 2 NVIDIA Quadro FX 5800 GPUs. Longhorn also has an additional 16 large-memory nodes containing 144GB of RAM, 8 Intel Nehalem cores (@ 2.5 GHz), and 2 NVIDIA Quadro FX 5800 GPUs. For more detailed information on Longhorn, please see the [Longhorn User Guide](#).

**Queue information:**

updated at January 31, 2010, 7:50:22 pm [refresh](#)

Available: Used: The Longhorn queues are open. 180 nodes available out of 250 total.

**ACTIVE JOBS**

JOBID	JOBNAME	USERNAME	STATE	CORE	REMAINING	STARTTIME
5175	job.tr.256	gnov	Running	128	22:15:58	Sun Jan 31 18:06:20
5177	CUPLA_GEMM	figuall	Running	128	00:11:13	Sun Jan 31 18:31:35
5181	CUPLA_GEMM	figuall	Running	128	04:30:58	Sun Jan 31 18:51:20
5182	CUPLA_GEMM	figuall	Running	128	04:31:13	Sun Jan 31 18:51:35
5185	CUPLA_GEMM	figuall	Pending	32	04:33:58	Sun Jan 31 18:54:20
5186	CUPLA_GEMM	figuall	Running	64	04:33:58	Sun Jan 31 18:54:20

6 active jobs : 76 of 250 hosts ( 30.40 %)

**WAITING JOBS**

JOBID	JOBNAME	USERNAME	STATE	CORE	WCLIMIT	QUEUE TIME
-------	---------	----------	-------	------	---------	------------

The screenshot shows the Longhorn Visualization Portal interface within a Mozilla Firefox browser window. The browser address bar shows the URL 'https://portal.longhorn.tacc.utexas.edu/'. The portal has a navigation bar with 'Home', 'Allocations', 'Jobs', 'Rendering', 'Help', 'Admin', and 'Vislab'. Below this, there's a 'Selected Files' section listing files like '1: .ICEauthority', '2: .Xauthority', '3: .bash\_history', '4: .envision\_address', '5: .envision\_display', and '6: .envision\_job\_id'. A 'State' section shows 'Dataset Information' for 'Isotropic.vtk' with dimensions 'X: min: 0 max: 127', 'Y: min: 0 max: 127', and 'Z: min: 0 max: 127'. A 'Plots' section shows 'Added Algorithms' with a '3D' plot selected. A 'Toolbox' on the right contains various visualization tools like 'Camera', 'Settings', '3D', '2D', 'Wireframe', 'Surface', 'Volume', 'Color', 'Opacity', 'Grid', 'Axis', 'Legend', 'Export', and 'Print'. The main visualization area shows a 3D plot of a dataset, which appears to be a complex, branching structure, possibly a molecular model or a network graph, rendered in green and yellow.

**Longhorn Visualization Portal - Mozilla Firefox**

File Edit View History Bookmarks Tools Help

https://portal.longhorn.tacc.utexas.edu/

Most Visited | Getting Started | Latest Headlines | Customize Links | Lens Caps and Vein Taps | Salsa Chicken Rice Cass...

Longhorn Visualizatio... | TACC User Portal

**Longhorn Visualization Portal**

Home | Allocations | Jobs | Rendering | Help | Admin | Vislab

Disconnect | Options | Clipboard | Send Ctrl-Alt-Del | Refresh

Applications | Places | System

Visit 2.0.1

Window 1

File Controls Options Windows PlotAtts O

Global

Active window 1

Selected Files

localhost

- 1: .ICEauthority
- 2: .Xauthority
- 3: .bash\_history
- 4: .envision\_address
- 5: .envision\_display
- 6: .envision\_job\_id

Open

Time

Plots

Add Operators Delete

Added Algorithms

3D

Apply operators / sel

**Longhorn Visualization Portal**

Home | Allocations | Jobs | Data | Visualization Algorithms | Rendering | Snapshots | Help

TeraGrid:dstanilo logout  
Resource: Longhorn (Job 72027)  
Time left: 5:56:59

**Toolbox**

- Camera
- Settings
- 3D
- 2D
- Wireframe
- Surface
- Volume
- Color
- Opacity
- Grid
- Axis
- Legend
- Export
- Print

# 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

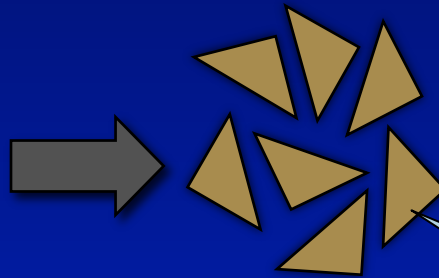
Compute  
Nodes

Lustre File Systems

↔ Read/Write File System Access

→ Job submission

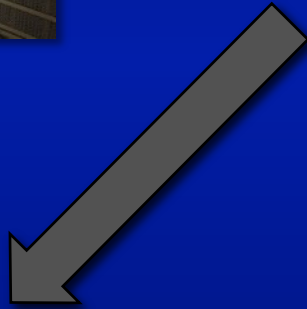
# Current Community Solution: Fat Client – Server Model



- Geometry (or pixels) sent from server to client, user input and intermediate data

Geometry generated on server

- Data traffic can be too high



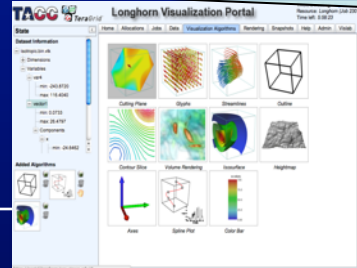
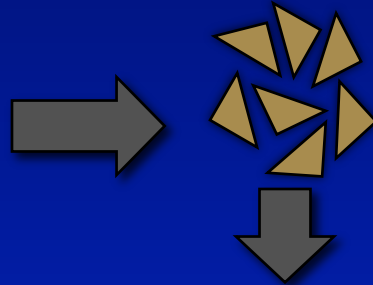
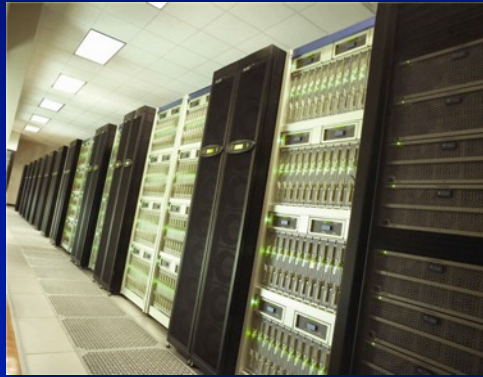
Geometry sent to client running on user machine



Connection options often assume single shared-memory system



# TACC Solution: Thin Client – Server Model



- Run both client and server on remote machine
- Geometry, images and client all remain on server
- Minimizes required bandwidth and maximizes computational resources for visualization and rendering
- Only pixels, mouse and keyboard sent between client and server
- Can use either a remote desktop or a web-based interface

# 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/mathematical analysis

# 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*

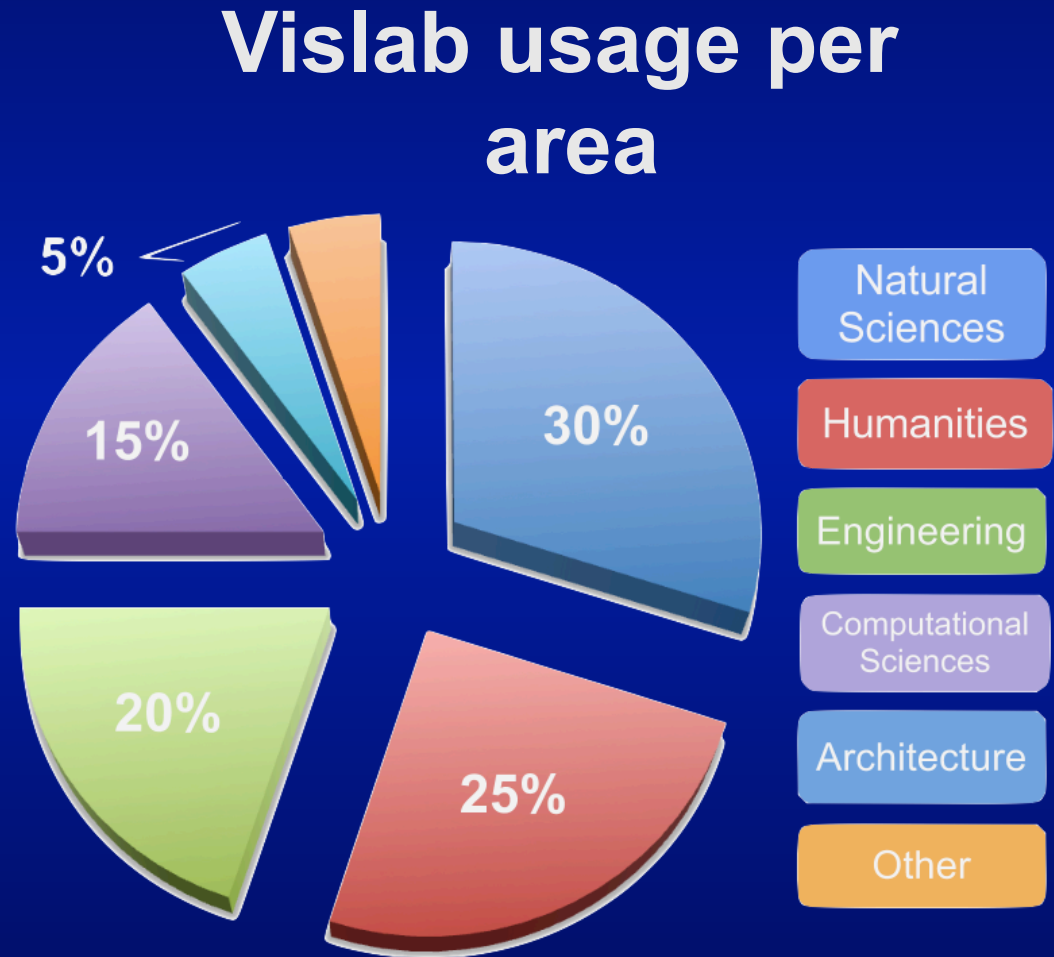
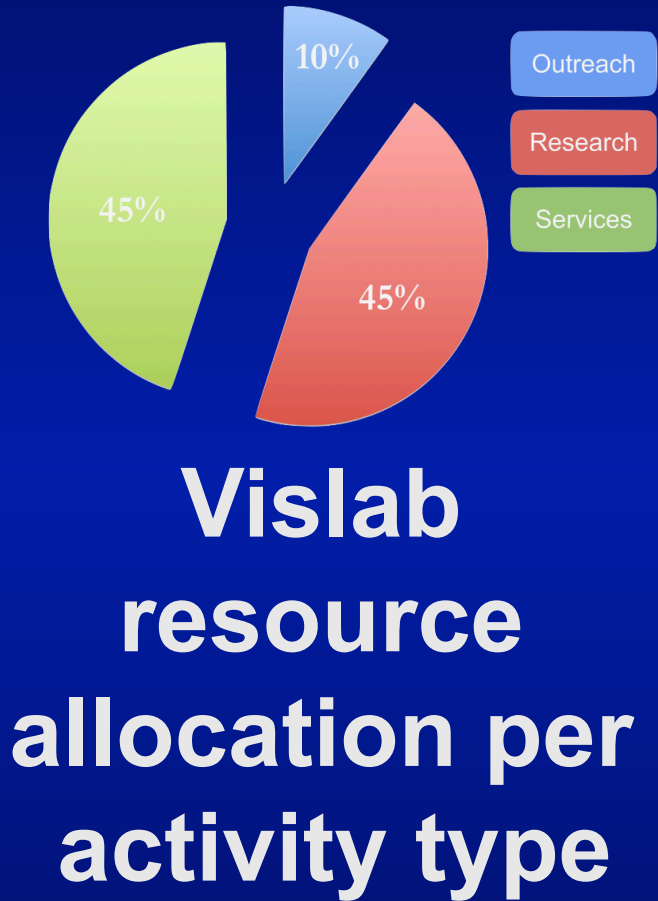
- 3x2 tiled display(1920x1600) – 12M Pixels
- PQ Labs multi-touch overlay, 32 point 5mm touch precision
- 11 mm bezels on the displays



# Vislab Numbers

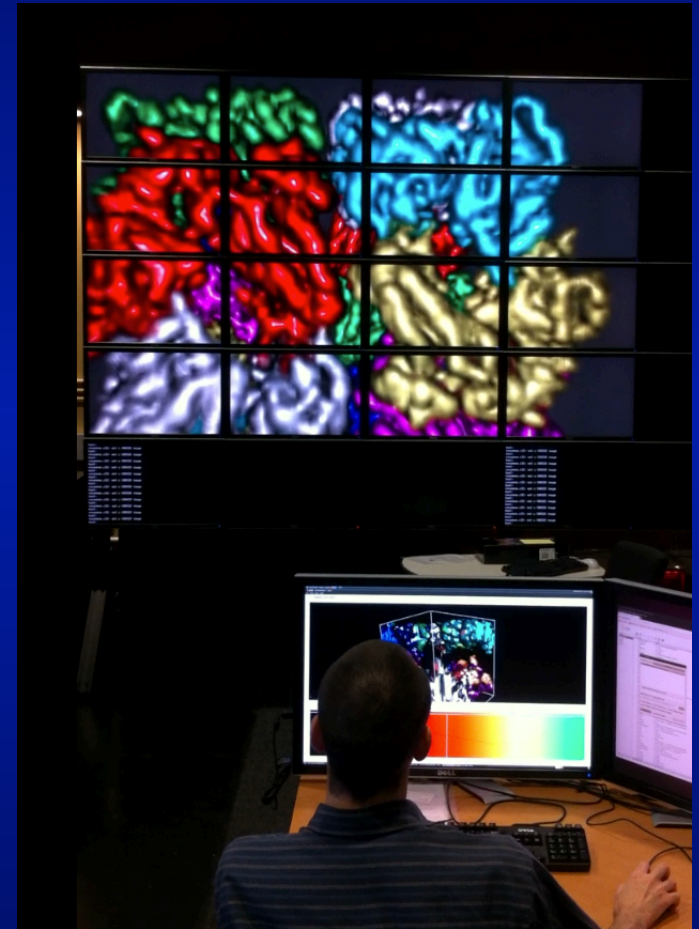
- Since November 2008, the Vislab has seen over 20,000 people come through the door.
- Primary Usage Disciplines – *Physics, Astronomy, Geosciences, Biological Sciences, Petroleum Engineering, Computational Engineering, Digital Arts and Humanities, Architecture, Building Information Modeling, Computer Science, Education*

# Vislab Stats



# Sample Use Cases – Biological Sciences

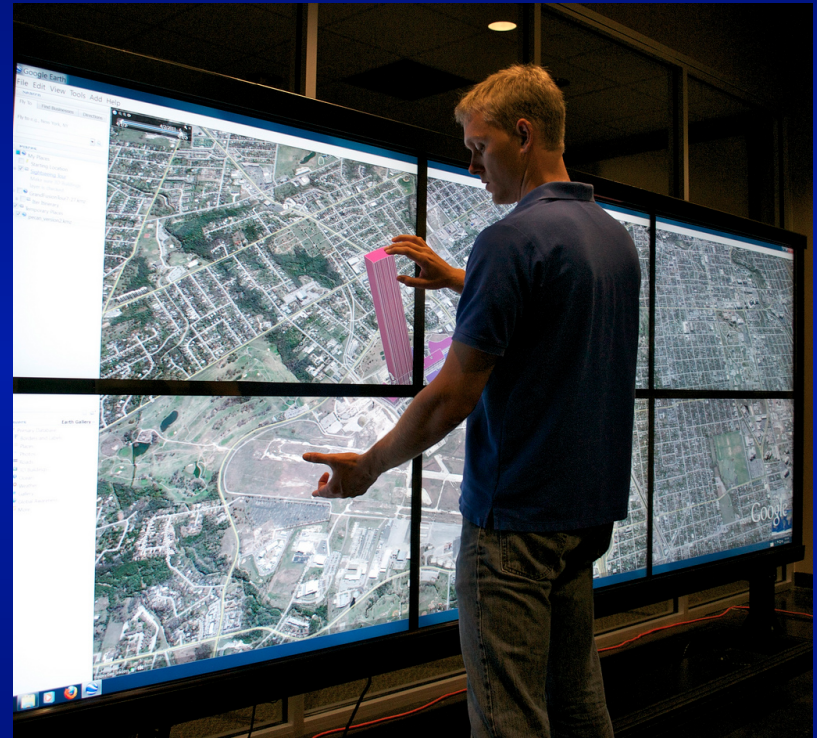
- **Research Motivation:** understand the structure of the neuropil in the neocortex to better understand neural processes.
- **People:** Chandra Bajaj et. al. UT Austin
- **Methodology:** Use Stallion's 328 Mpixel surface to view neuropil structure.





# 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

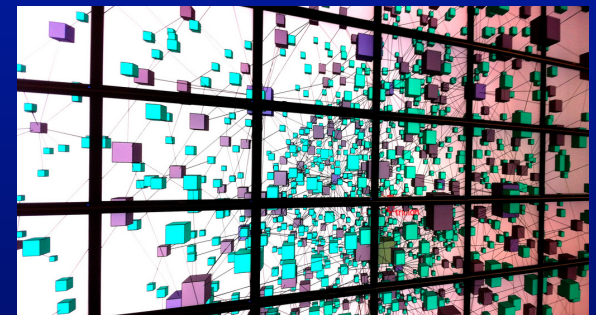
- **Motivation:** use advanced visualization resources as a tool for the arts and humanities
- **People:** TACC Vislab, UT Austin Digital Humanities
- **Methodology:** create software to allow non-trained users the ability to take advantage of distributed systems and graphics technology. Develop Massive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



Faces of Mars



Moving Pixels



Text Universe



# 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



