Visualization of Large Data

- Can't do it on my computer.
- The data is over THERE.
- It's too slow when I click it.
- Can't do it on any computer.

This is Kelly Gaither's talk repurposed to discuss large data visualization. Any errors are Drew's.

Answers

- Understand speeds and feeds – where do bandwidth and latency matter? How much do you need?
- Where am I making choices?
- What happened that I'm seeing nothing, or it looks askew, or it blinks at me?

Three Pipelines

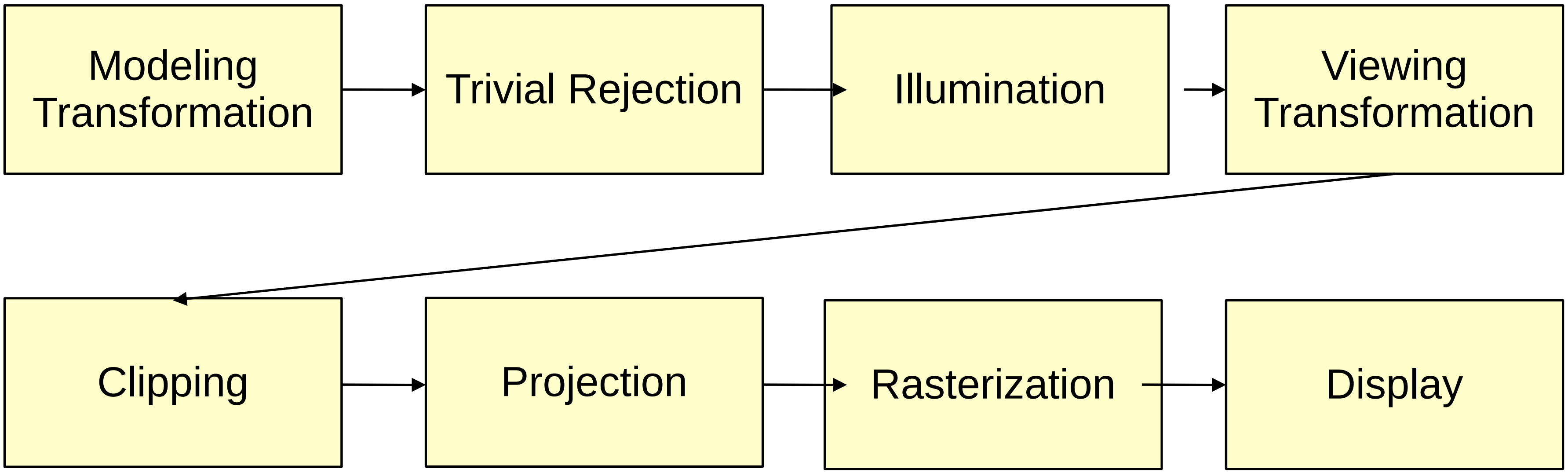
- Visualization Pipeline -> Data to Polygons
- Graphics Pipeline -> Polygons to Pixels
- Human Perception -> Pixels to Insight

Graphics Pipeline

- Given polygons, show them on the screen.
- GL does this for you

```
glColor3f(0.0, 1.0, 0.0); // blue
glBegin(GL_QUAD);
  glVertex2f(0.0, 0.0);
  glVertex2f(1.0, 0.0);
  glVertex2f(1.0, 1.0);
  glVertex2f(0.0, 1.0);
glEnd();
glTranslate(-1.5, 0.0, 0.0); // move
object
```

Graphics Pipeline



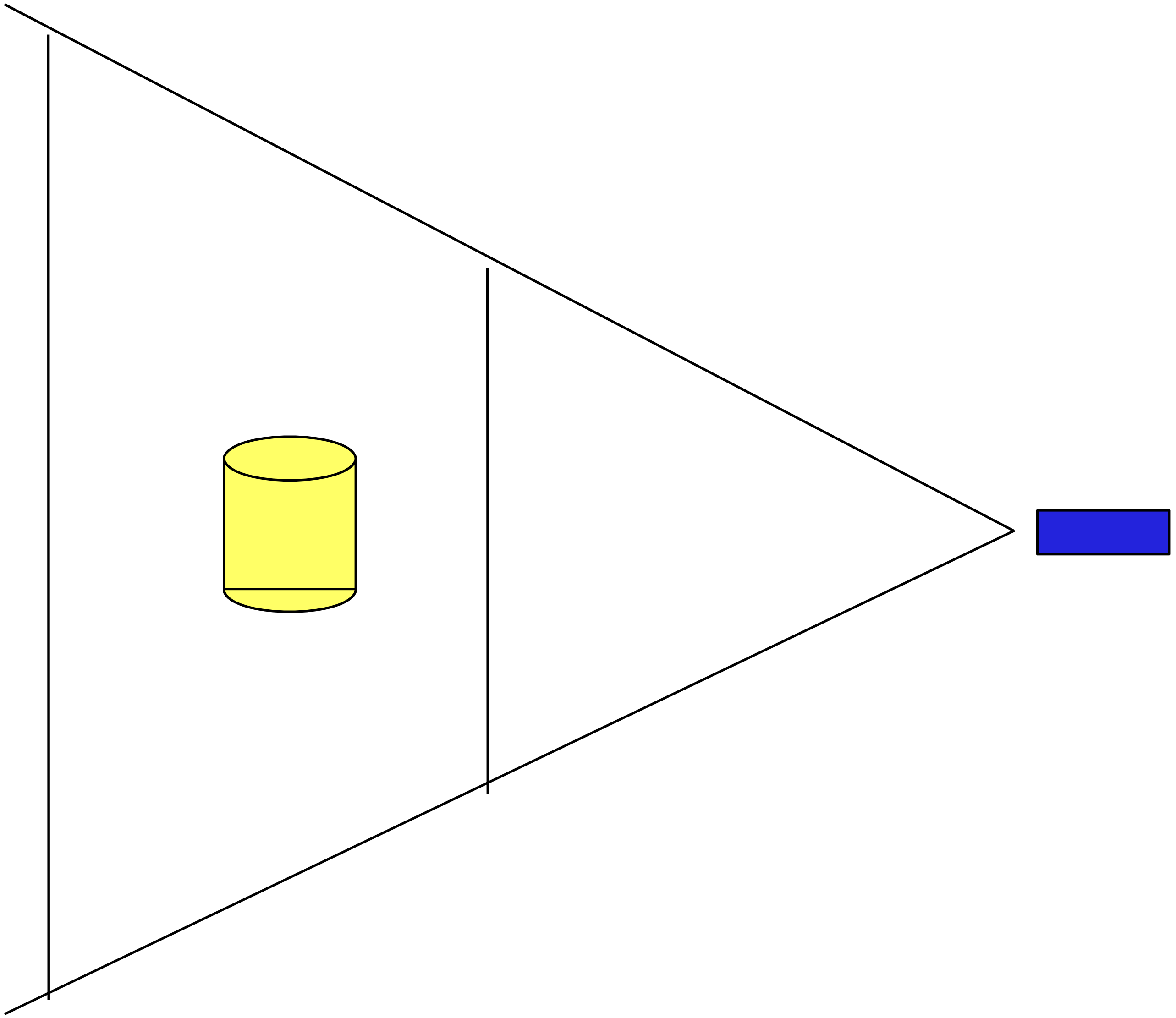
Modeling Transformation

- Modeling transformation takes to unit coordinates for conversion to screen
- Given a simulation of a tsunami in the Indian Ocean, you find the simulation freezes or disappears, but tests of simpler input look fine. What's going wrong?
- As you turn a model, some of the polygons flip from one color to another and back again. Why?

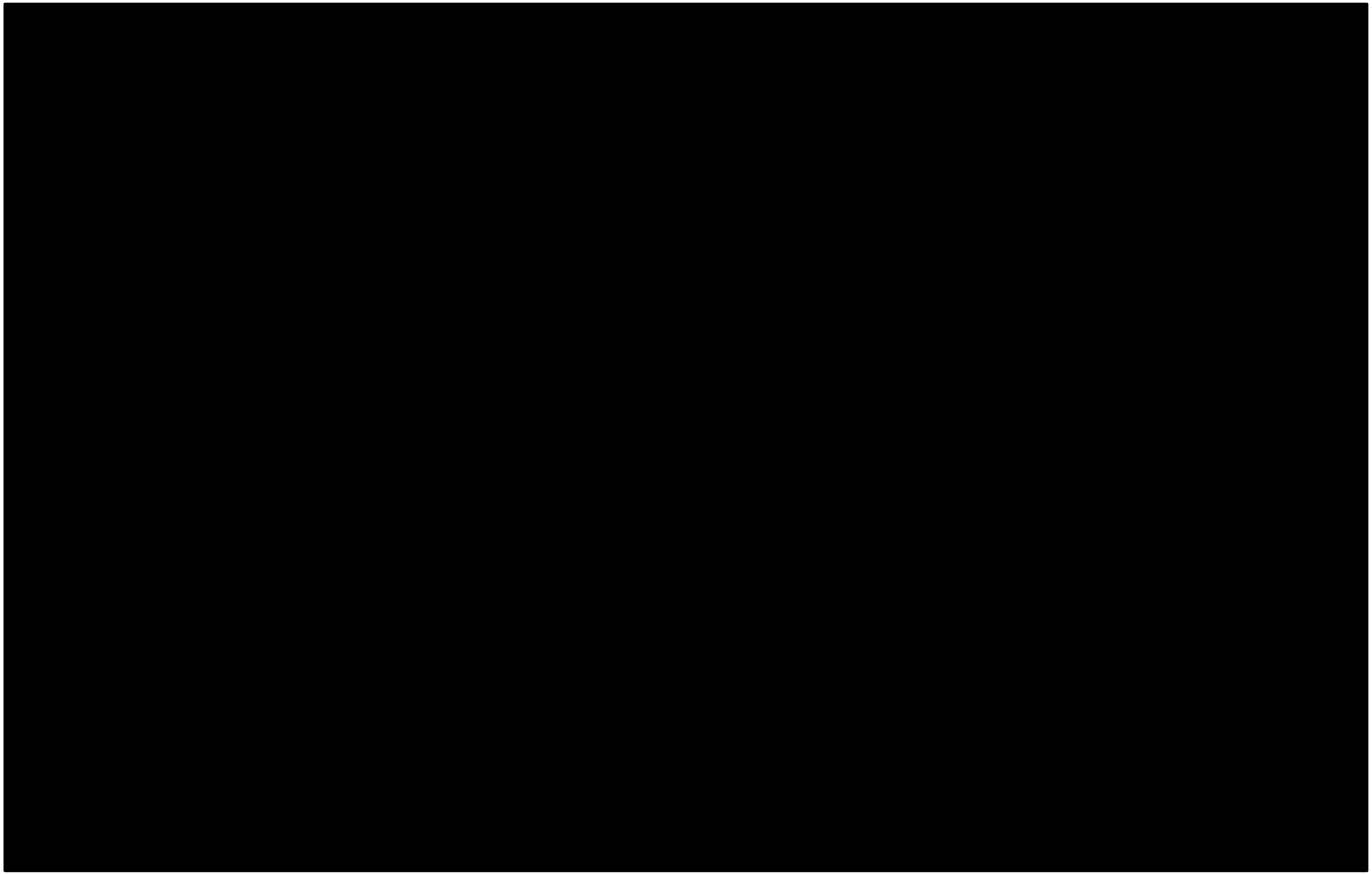
Illumination

- Normals of polygons used to calculate color.
- Default to diffuse. You add ambient and specular.
- Key is main (warm) light, fill is (cool) from opposite side, headlight smooths between the two, backlight gives that nice glint.

Clipping



What's Wrong?



Projection

- Turn 3D to 2D.
- Perspective
- Parallel – orthogonal, oblique
- How many operations so far?

Rasterization

- Convert to pixels
- With more or less complicated interpolation.
- May include interpolation for shading (Gouraud or Phong), which is more expensive.
- Memory buffer access rate is main limit.
- 1024x768 pixels * 24 bits per pixel * frame rate

Computational Intensity

- Example from Foley & van Dam
- 10,000 triangles, 100 pixels/triangle, illumination model, shading, 1280x1024 screen, 10fps
- Geometry: 2.2M mults, 1.5 add/sub = 36.9M flops
- Rasterization: 42.5M additions, 51M frame-buffer accesses/s

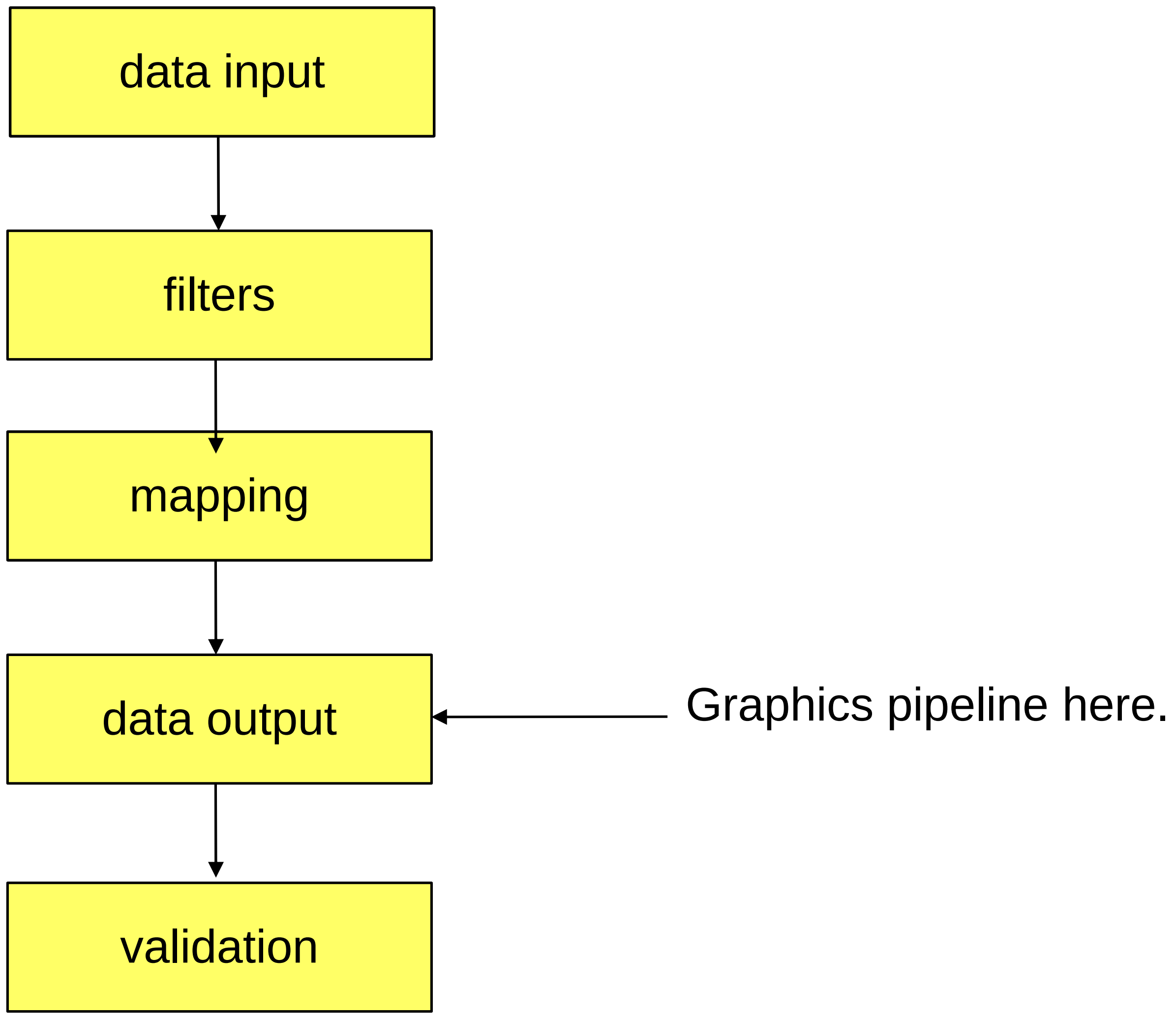
Parallelism

- Geometry processing just needs pixels, but how geometry is split affects rasterization.
- Depth-sort lets you paint the back first.
- Z-buffer stores, per pixel, how far things are.
- Transparency can get muddled.

Bandwidth and Latency

- So what do you send over the wire to somebody in Madagascar?
- Points, lines, polygons?
- OpenGL commands?
- Pixels?
- Give them their own copy of the input data?

Visualization Pipeline



Data Lives in Spaces

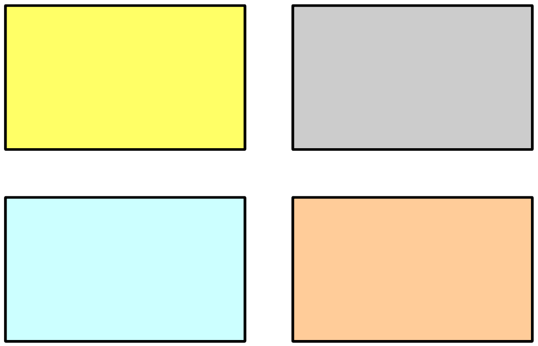
- Scalar, vector, tensor
- At a point, in 1D, 2D, 3D, N-dimensions with metrics in those dimensions
- With connectivity
- Becomes fields of data defined on points or cells
- Connectivity of points, meshes, or unstructured

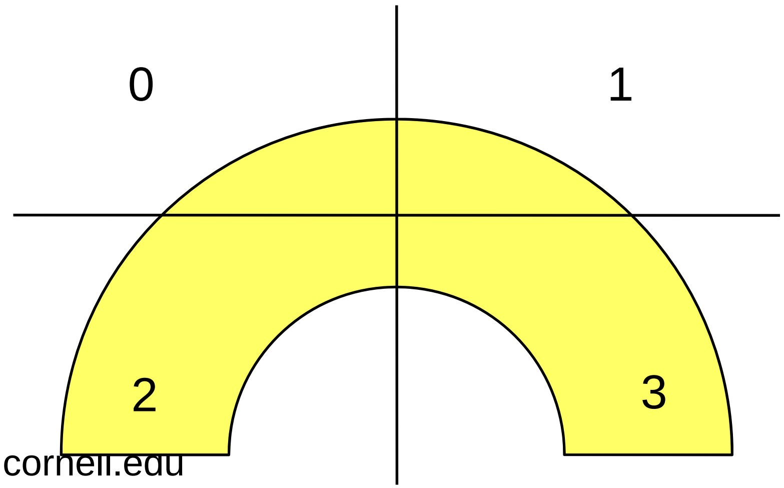
Name That Dataset Type

- Confocal microscope gives 3D view of slime.
- Computational fluid dynamics
- Crystallographic rotations from electron backscattering.
- Stream level monitoring stations

Parallel Filters

- Balance is important.
- Task and/or data decomposition.
- Ghost cells allow filters to share with neighbors.
- Unstructured grids may need to reallocate.
- How do filters differ from simulations?





Mapping

- Refers to choice of a something visual to represent data. A polygonal surface, a set of glyphs, a volume.
- In VTK, the mapper is what takes points, lines, polygons, and calls OpenGL.
- Fluid flow could be contours of energy, streamlines along flow, colors for vorticity, textures to show flow direction, glyphs to point.

Validation

- Verify the product is a reasonable approximation of the raw data.
- Errors may result from how the dat is read and manipulated by the visualization tool, bugs in the tool, or problems in the data itself.
- Most tools today do not include a mechanism to automatically tell you if there is a problem.

Human Pipeline

- Can you really see all those little triangles?
- 1x1 pixel has no color
- 3D + colors + movement? Funny glasses?
- Depth of understanding of complex data comes more from speed of interaction, ability to shift focus, than from increasing complexity of what we see.

Paraview

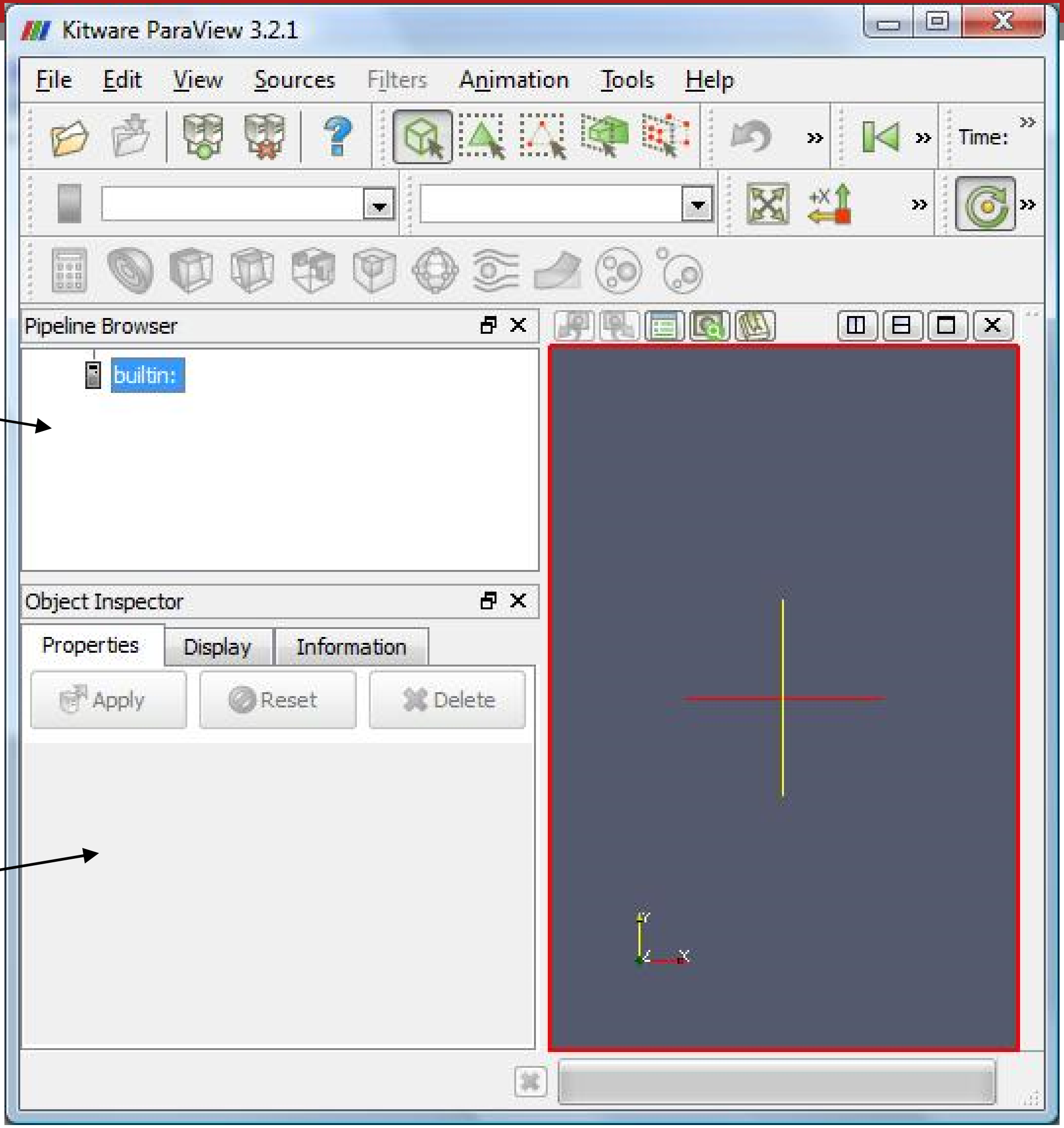
- Excellent for computational fluid dynamics
- Straightforward to use
- Runs in parallel several ways
- Download at http://www.paraview.org/New/download.html

Paraview Lab

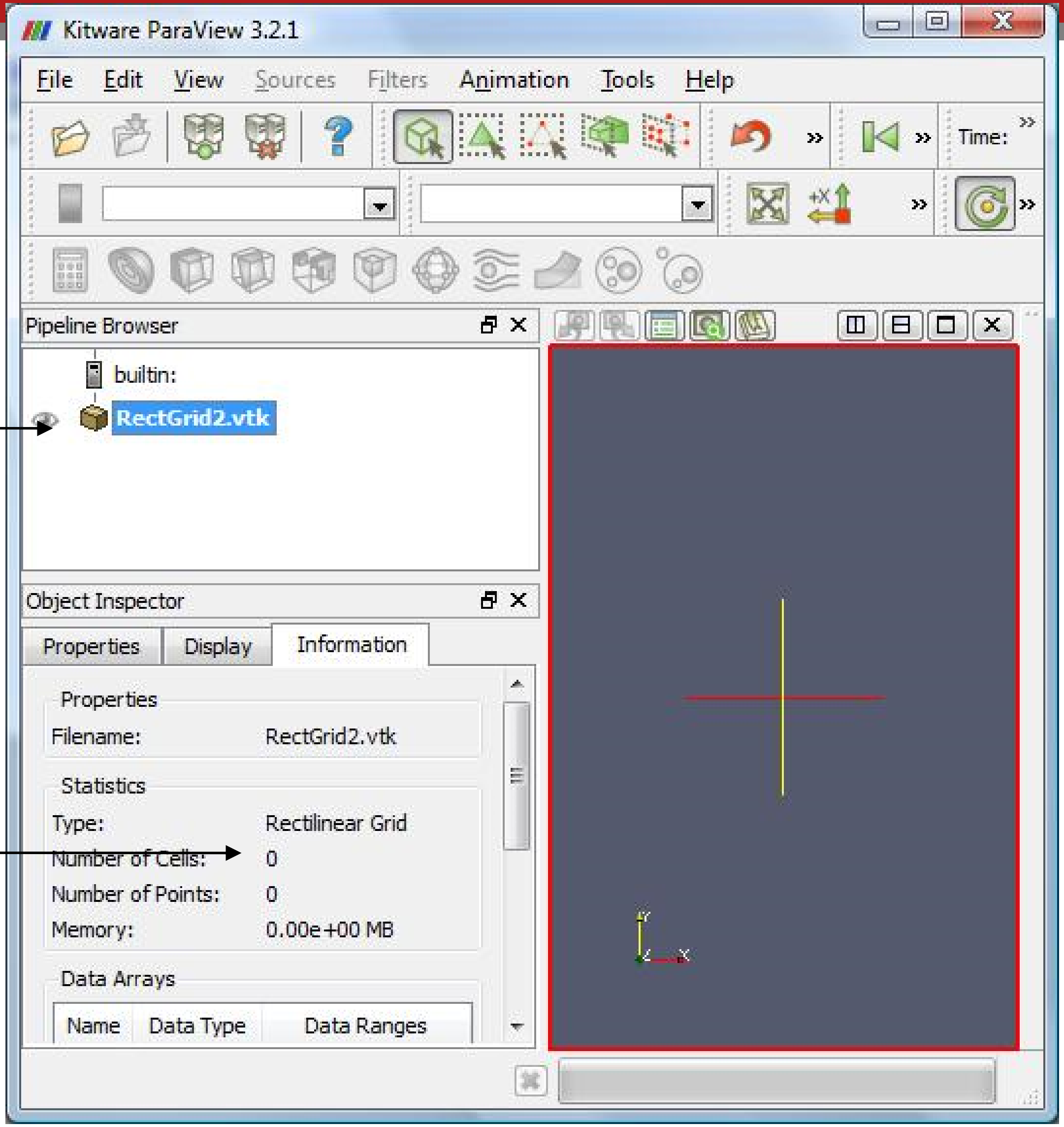
- Goal is to see basic functions – contour, streamline, volume rendering – on a local machine before we go to Spur.
- Copy RectGrid2.vtk to your machine.
- Start Paraview from the Windows menu.

Pipeline browser

Object inspector



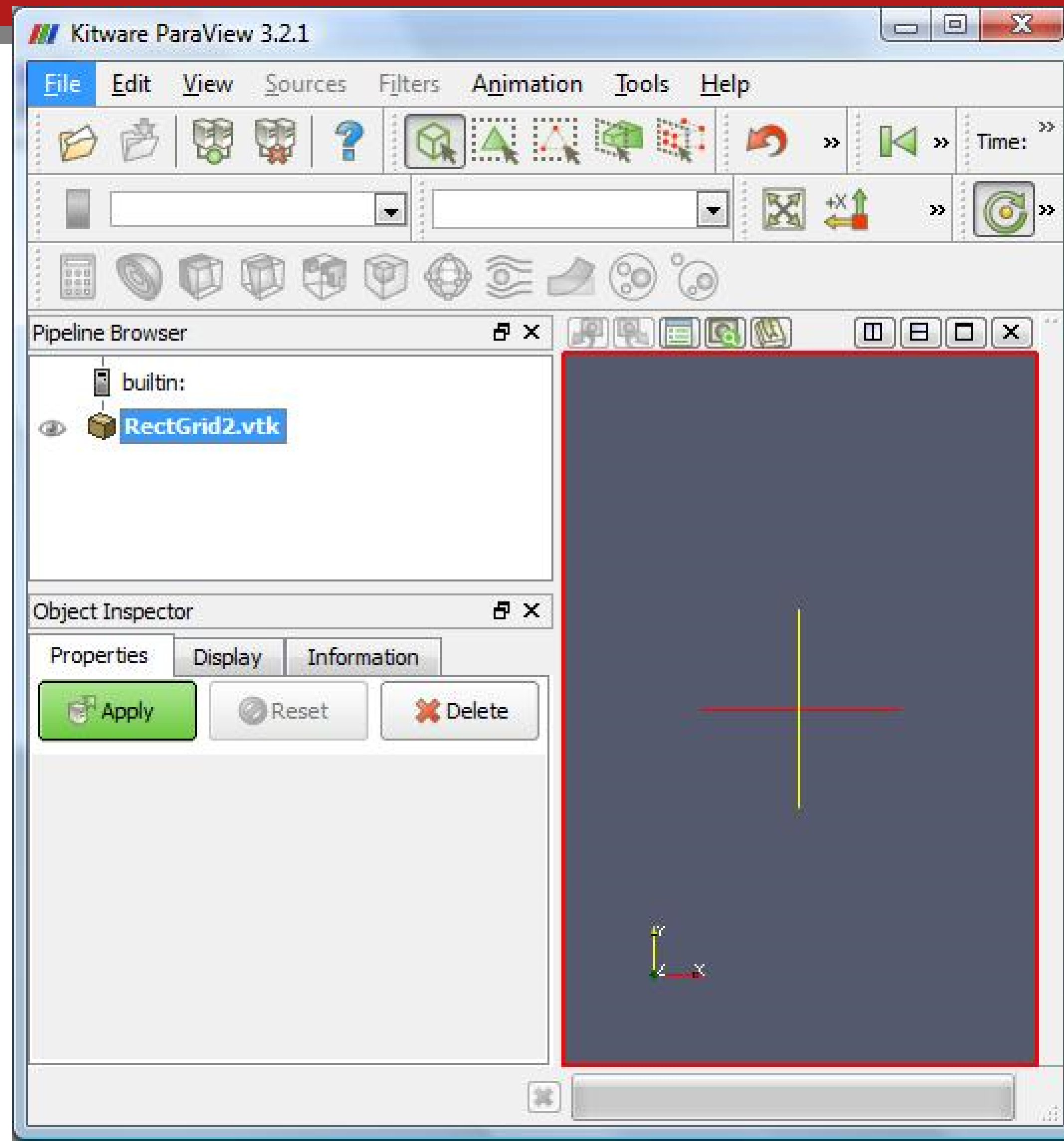
Load RectGrid2.vtk using File->open.

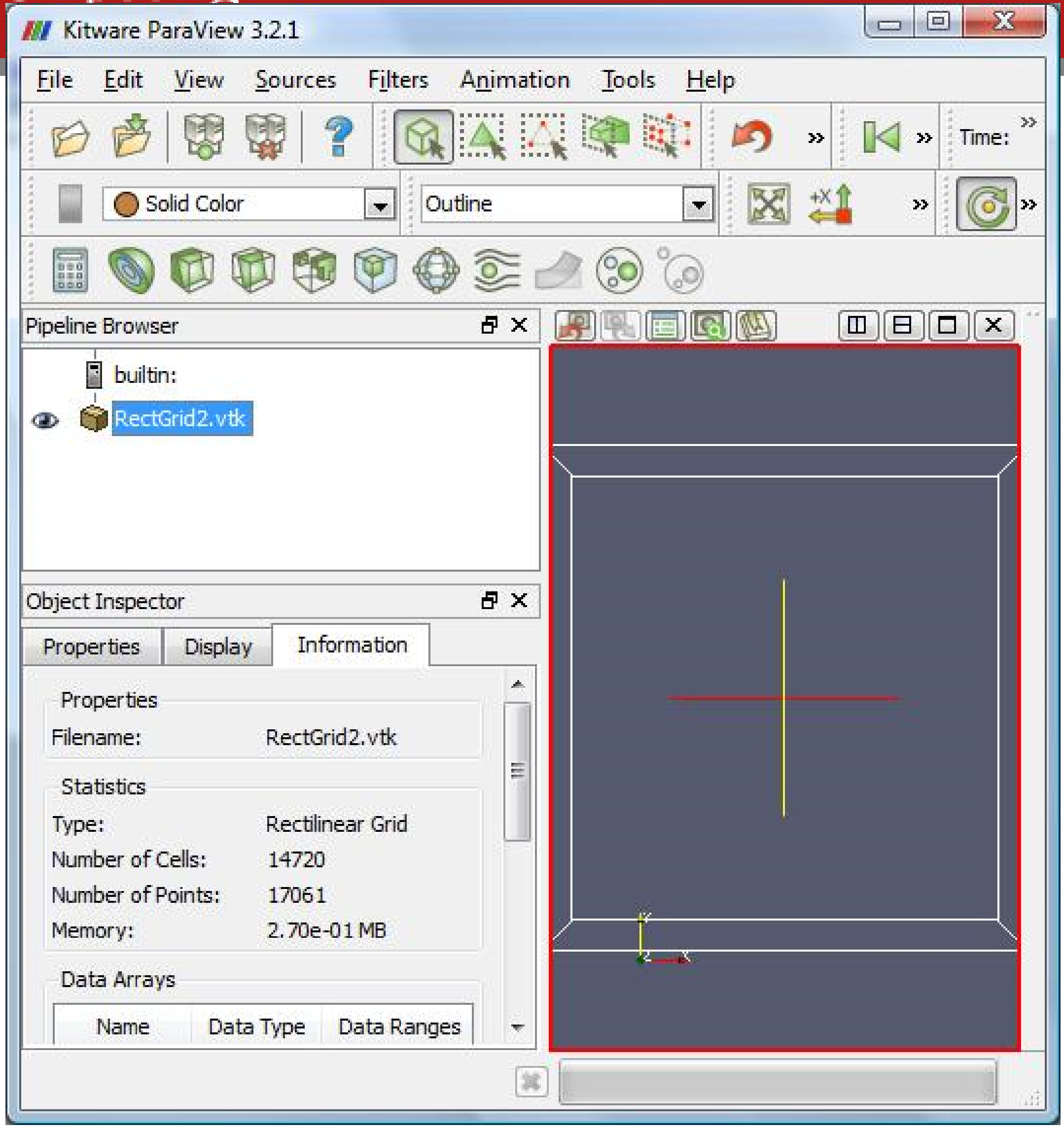


Your file at the top of the pipeline.

It thinks there are no cells or points in the file.

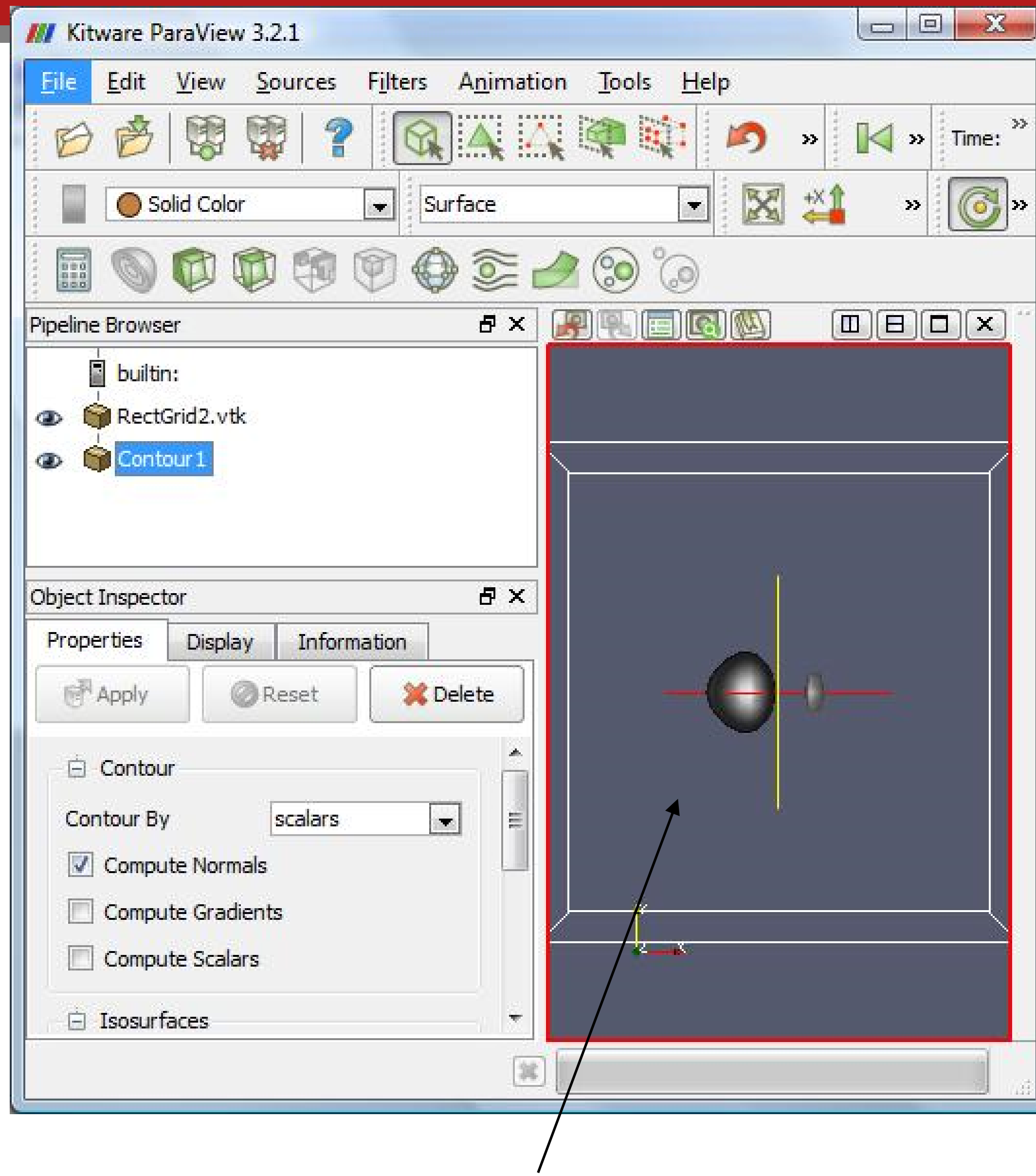
Hit “Apply” to load the file.





Cells!
Points!

1. Select dataset
2. Find Contour filter in the Filters menu.
3. Hit Apply, as usual.

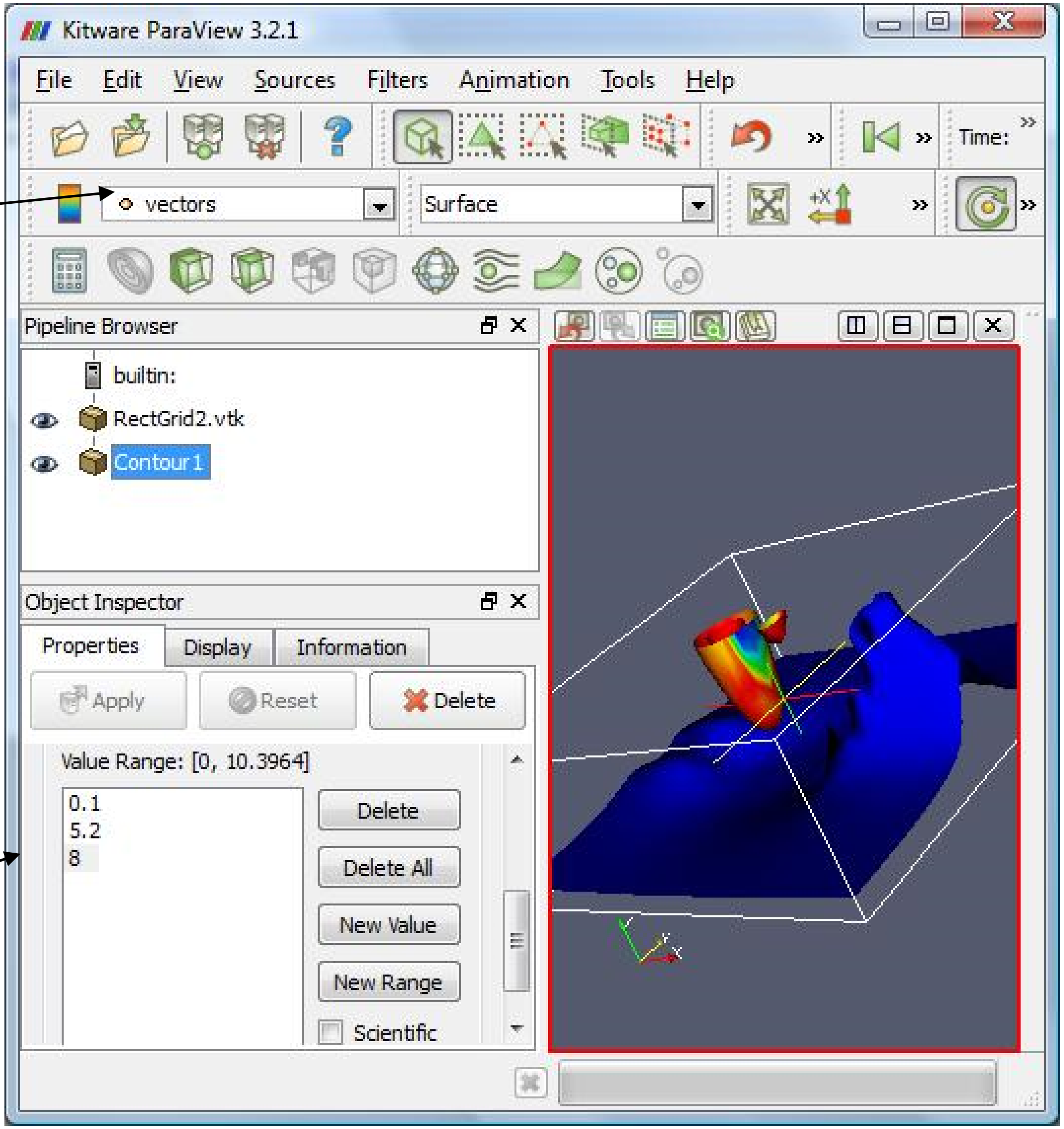


Click and drag. Try ctrl, shift, middle-click, right-click.

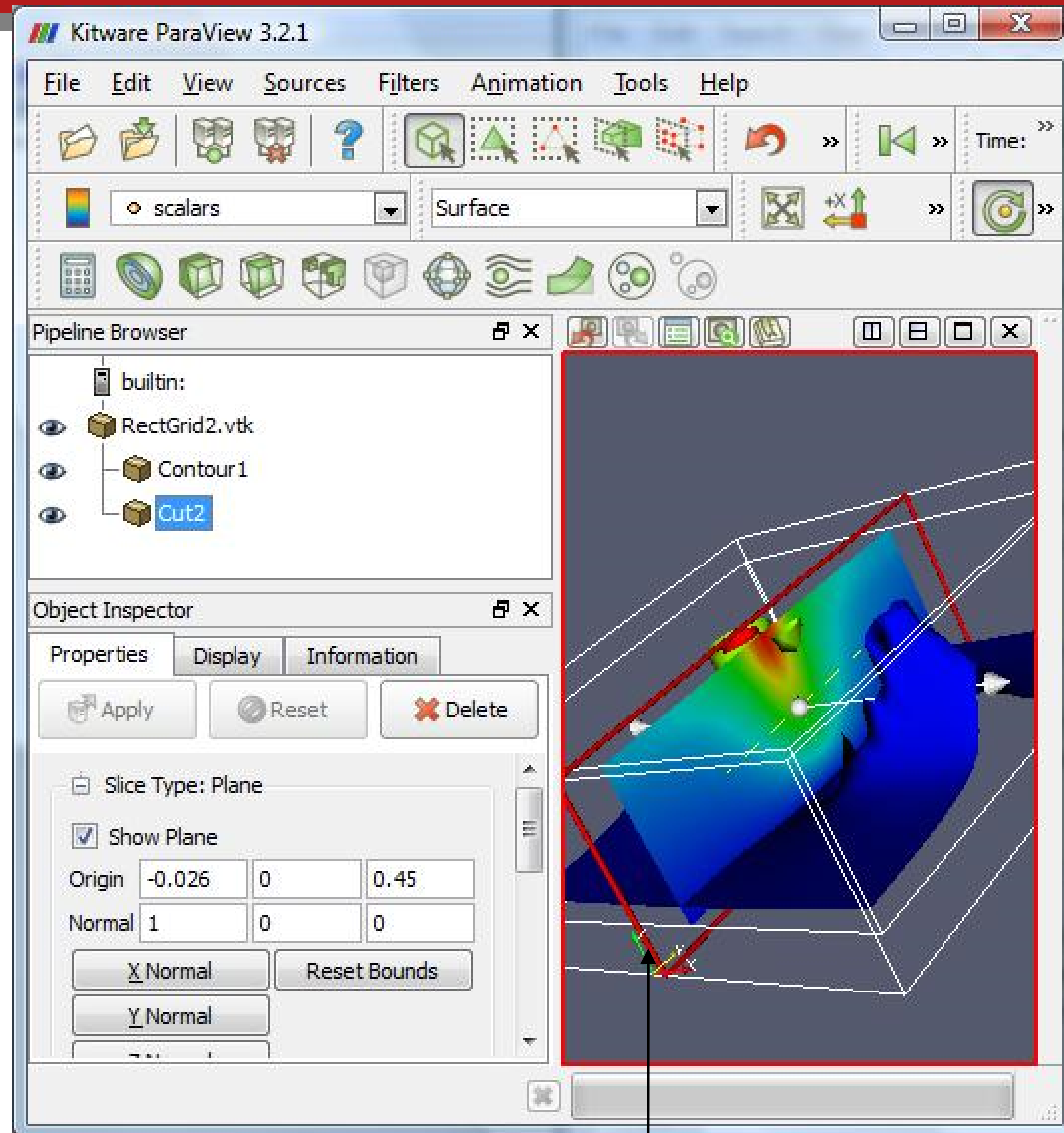
(Change navigation in Edit menu->Settings->Render View->Camera.)

Add colors with the drop-down list box. The choices come from fields of the dataset.

Change contours



1. Select RectGrid2.vtk
2. Add Slice filter.
3. Hit apply, again.

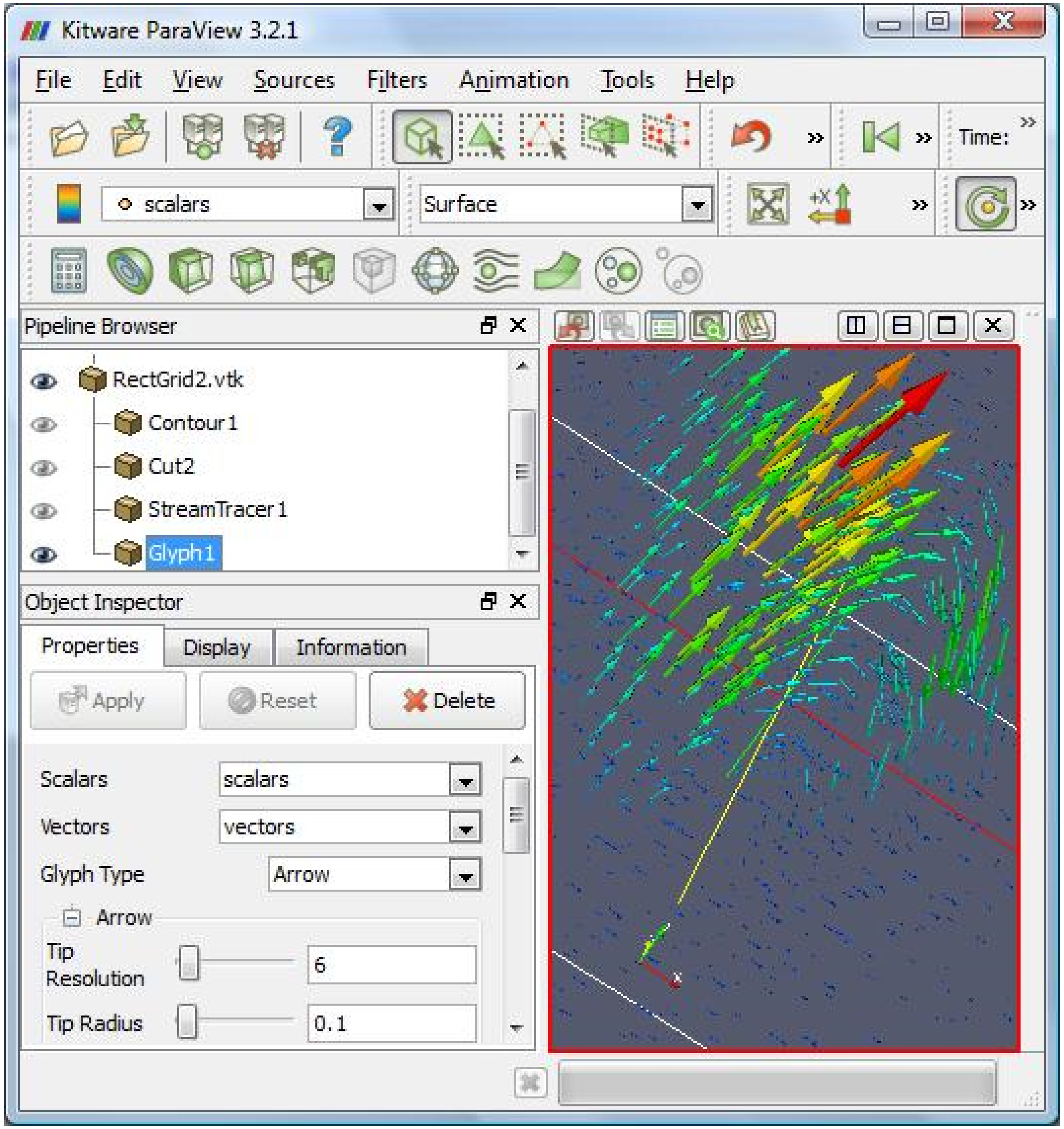


or...
1. Select Contour1
2. Add Slice filter.
3. Apply, apply.

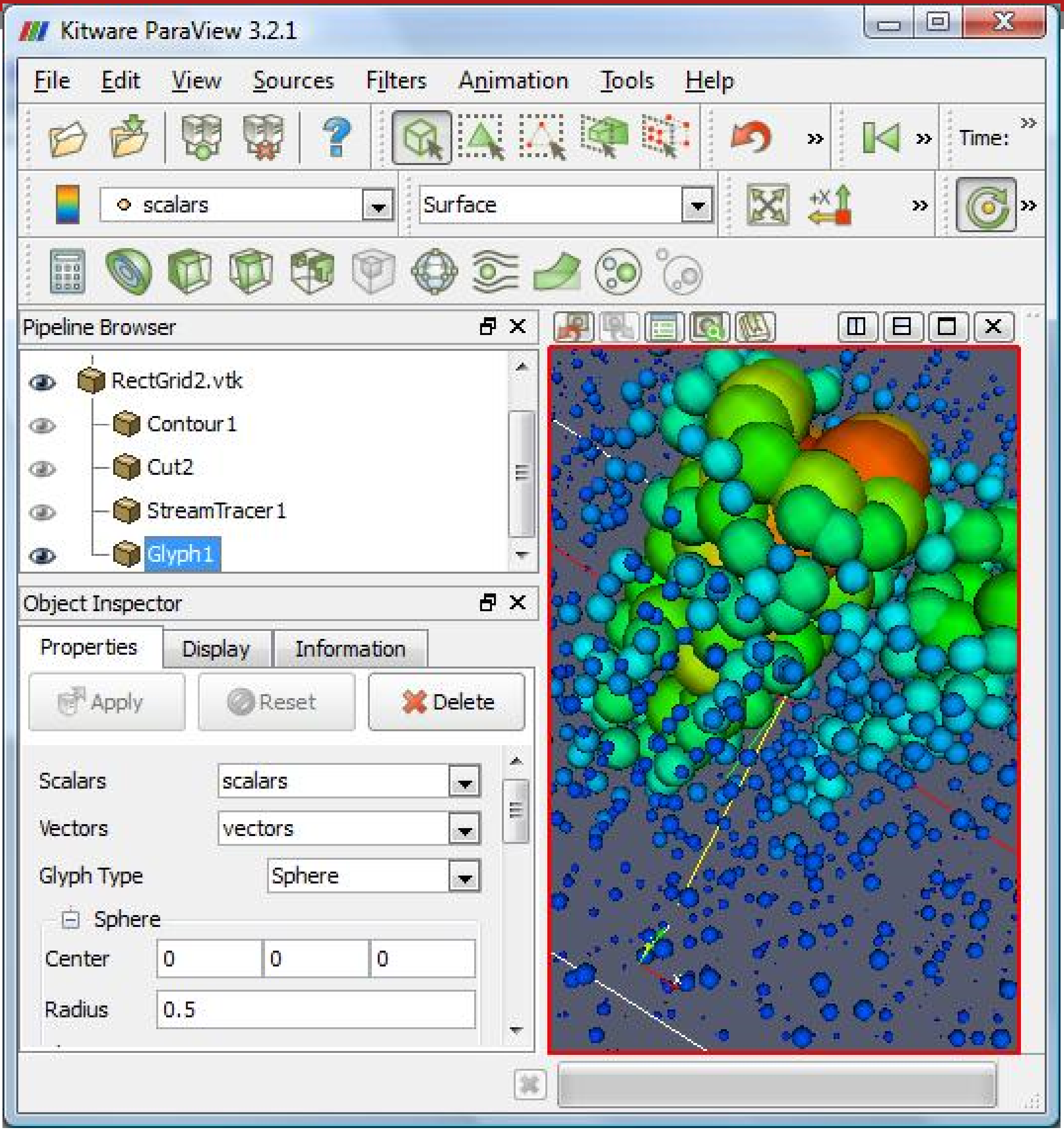
What is the difference?

Click and drag. Try Ctrl, Shift, right-click.

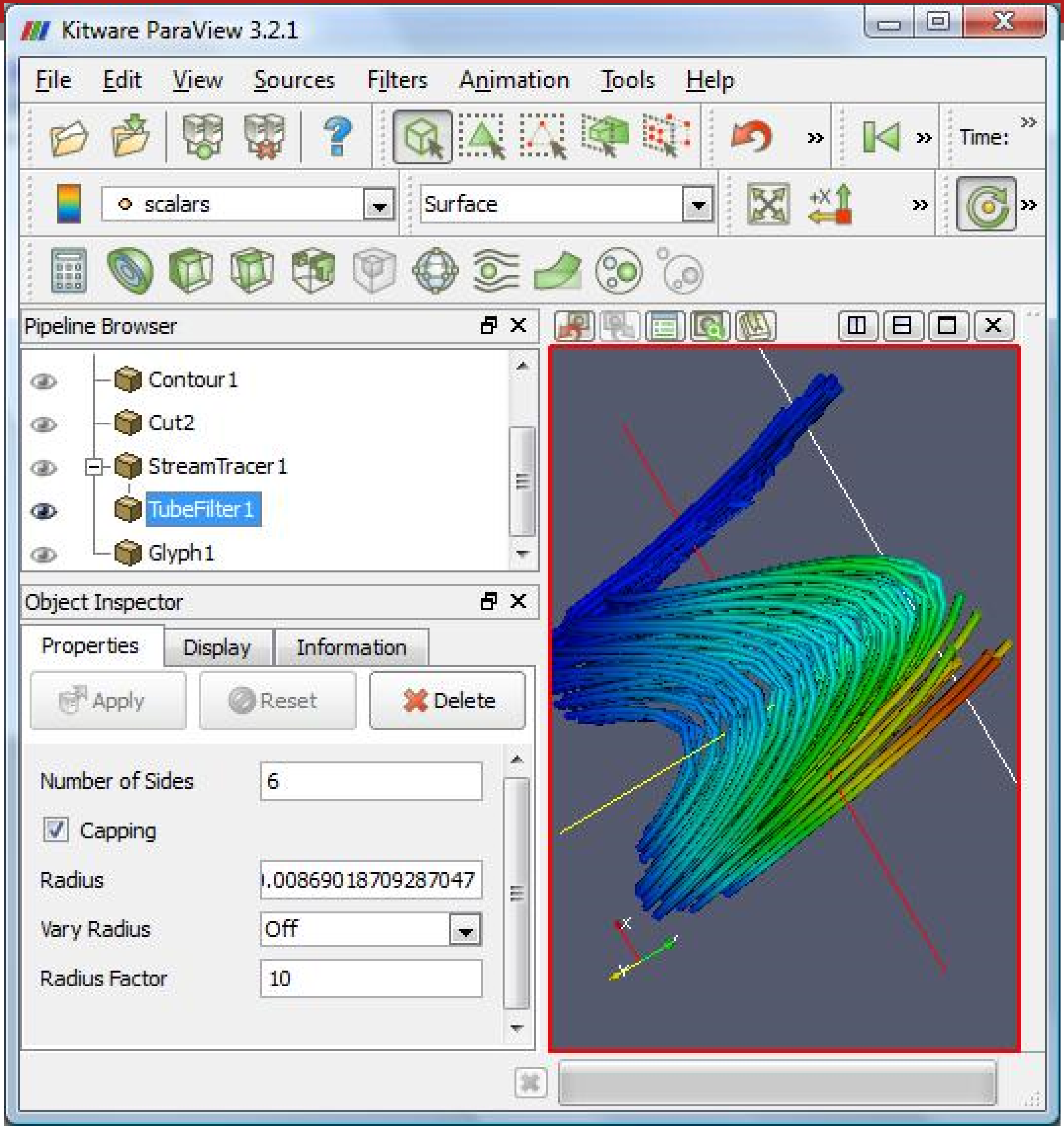
Glyph filter.
Play with the glyph type options.



The human visual system does not know whether to compare sphere size by diameter or volume. There is no good sense of “twice as much.”



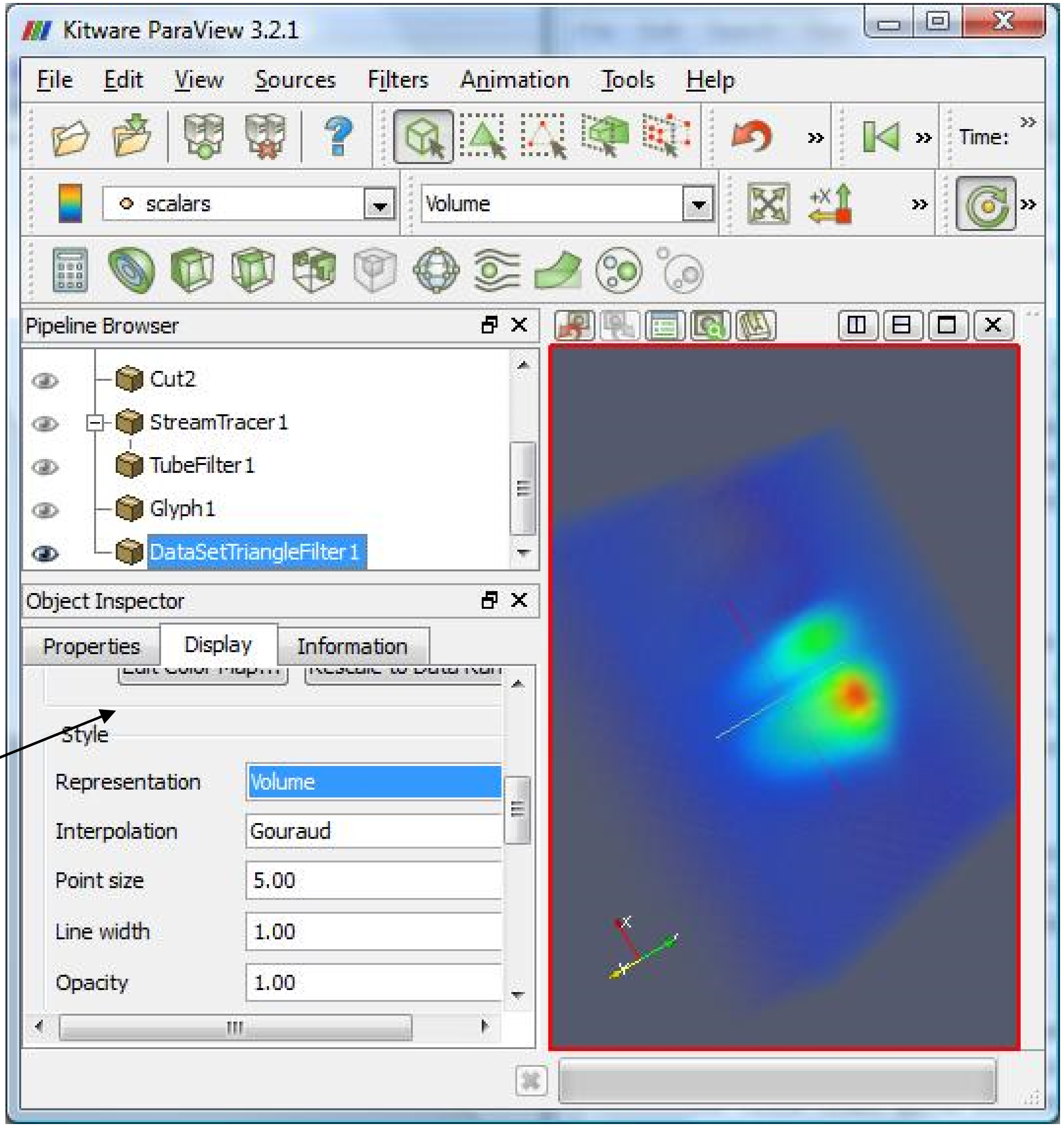
This time, add the tube filter to the StreamTracer, not to RectGrid2.



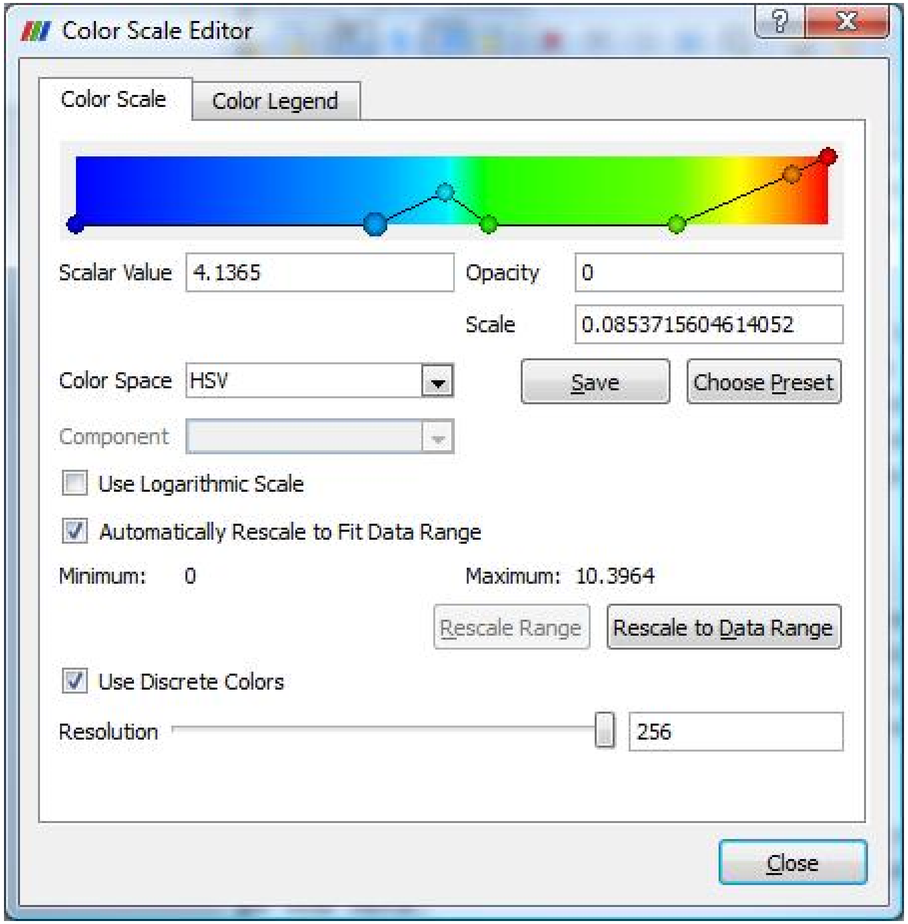
Volume Rendering

First, add “tetrahedalize” filter.
Select Display tab.
Find the Style section.
Change representation.

The colormap is just above
the Style section if you scroll up.



For volume rendering, the y axis of the line determines opacity.



Remote Visualization

- When you cannot move the data
- When the data is sensitive
- When a workstation isn't enough, these visualization machines harness multiple GPUs.

What's in there?

Remote Visualization Challenges

- Bandwidth mostly OK. 1280X1024 pixels of 24 bits at 12 times a second = 360 MBps
- Latency over the network and with the GPU
- Quality of service, user-interface response, ease of use, ease of access, scheduling

TACC's Ranger

Sun Constellation Linux Cluster

System Name: Ranger
ranger
Host Name: ranger.tacc.utexas.edu
IP Address: 129.114.50.163
Operating System: Linux
Number of Nodes: 3,936
Number of Processing Cores:
62,976
Total Memory: 123TB
Peak Performance: 579.4TFlops
Total Disk: 1.73PB (shared)
31.4TB (local)

http://www.tacc.utexas.edu/resources/hpcsystems/

TACC TeraGrid Visualization Hardware

- Locally, a tiled wall and a large new visualization laboratory
- Spur
  - 1 Sun Fire X4600 server (head node) with 2 Nvidia Quadro Plex model 4. Each Quadro Plex model 4 cocntains 2 FX5600's. 8 dual core CPU's and 256 GB of memory.
  - 7 Sun X4440 servers (quad socket, quad core) each connected to a QuadroPlex S4, 128 GB mem.
  - Total: 128 cores, 1TB memory, 32 GPUs

TACC Visualization Software

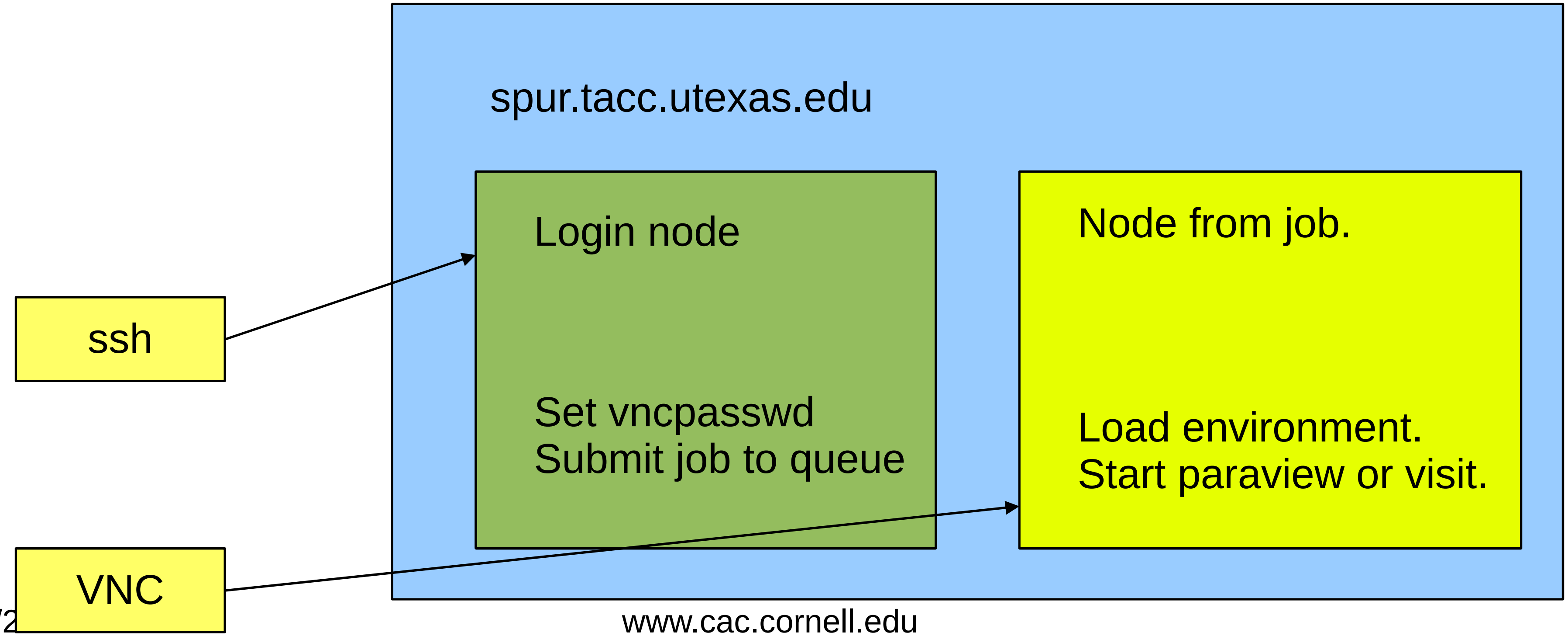
- Applications
  - Paraview, open source, general purpose,
  - Visit free parallel, general purpose,
  - EnSight commercial for CFD,
  - Amira, very good for medical data
- The Visualization Toolkit, OpenGL, OpenInventor, OpenSceneGraph

On Ranger and Spur

- Parallel Visit
- Parallel Paraview
- Mesa 3D graphics (Open Source OpenGL)
- The Visualization Toolkit (VTK) – Large library of visualization algorithms

Remote Visualization on Spur

- VNC = Virtual Network Computing
- Use TurboVNC, UltraVNC, TightVNC, Vinagre



Paraview Remote Lab

- Goal: Experience joys and pains of remove visualization.
- Account is train4xx. It will work for a week.
- Login using putty, a secure-shell client, or ssh.
- Use vncpasswd to set a low-security VNC password.
- Copy files with “cp ~/train400/* .”

From your machine: ssh spur.tacc.utexas.edu using Putty on Windows.

You will vnc to spur, so you need to set a password.
spur% vncpasswd
Submit a job to gain access to a visualization node running vncserver.
spur% qsub -A 20081023DATA /share/sge/default/pe_scripts/RUN.vnc -geometry 1440x900 -l h_rt=1:00:00

Check to see if the job ran and wrote ~/vncserver.out
spur% showq -u
spur% qstat
spur% ls vncserver.out

Options to RUN.vnc.

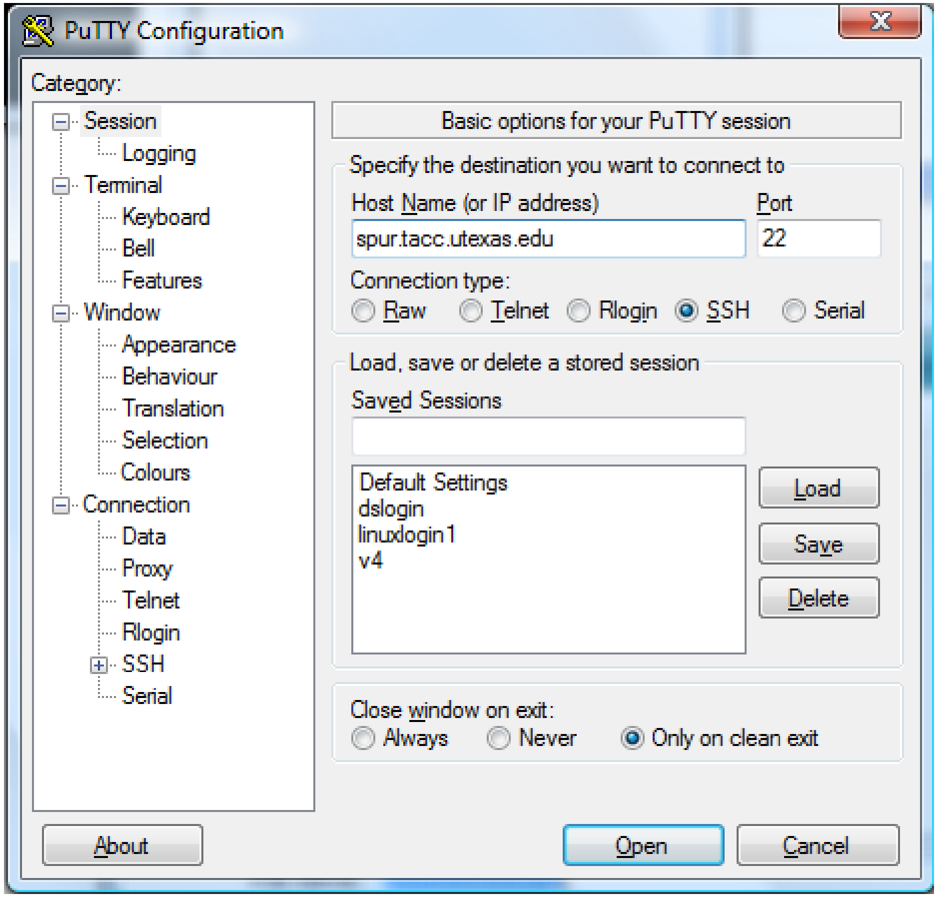
GEOMETRY = window size
-l h_rt=1:00:00 is one hour of run time
-l gfx=1 requests 1 graphics card

Tunnel VNC to Spur

- SSH will encrypt a TCP/IP connection from your machine to spur's VNC server.
- ~/vnc_server.out tells you the remote port.
- On Linux: “ssh -g -L 59xx:spur.tacc.utexas.edu:59yy spur.tacc.utexas.edu”

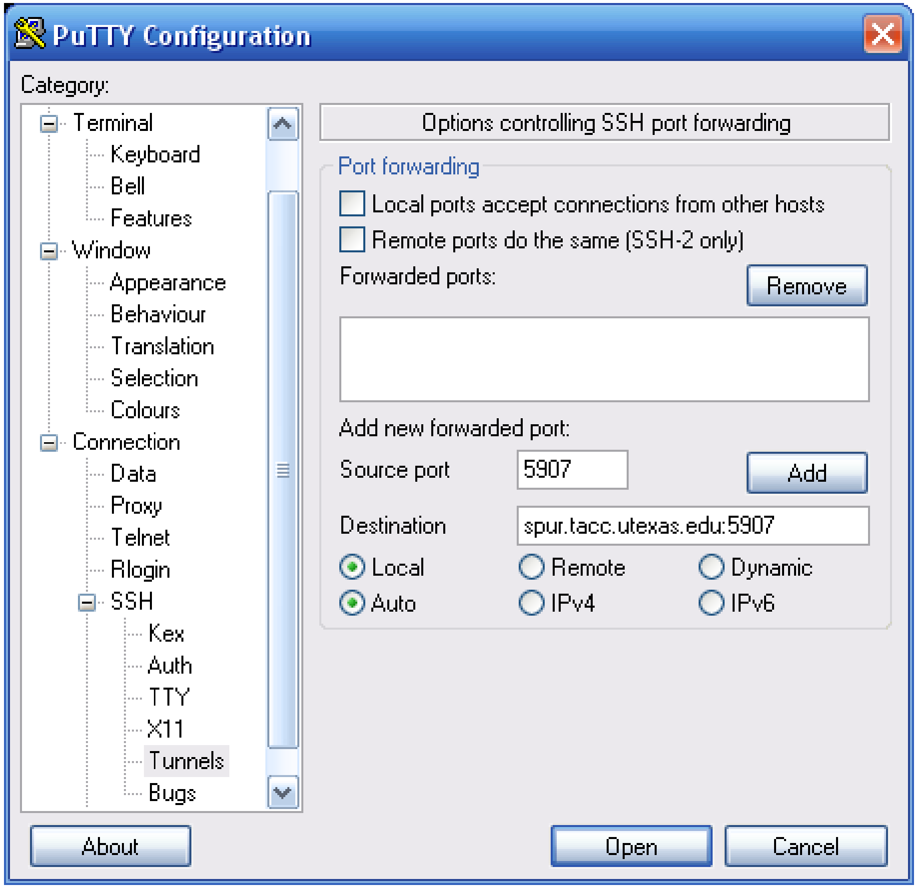
Port-forwarding in Putty

First type the hostname on the opening page.



Then selectSSH->Tunnels on the left.
Pick a random local port number.
Type the spur host:port below.
Click Add.
Then click open and enter your username and password.

Start VNC and connect to localhost:59xx.



Once VNC Starts

- In the xterm in VNC,
- Don't exit the black window until you're done.
- spur% module load vis
- spur% module load paraview
- “vglrun” is used for all OpenGL programs on VNC.
- spur% vglrun paraview
- Play with RectGrid2.vtk.

Visit

- Open Source, parallel visualization from LLNL
- Scalars, tensors, vectors
- Support for AMR and CSG meshes
- Quantitative analyses (expressions, queries, picking, lineout)
- GIS support
- Annotation for publication and presentations
- Built on VTK

Visit Advanced Features

- Geometry export to Curve, Alias Wavefront...
- Animation and movie generation
- Scripting interface with Python
- API interface with C++ and Java
- Dynamically extensible through plugins
- Parallel and distributed for large datasets
- Multiple database correlation / visualization

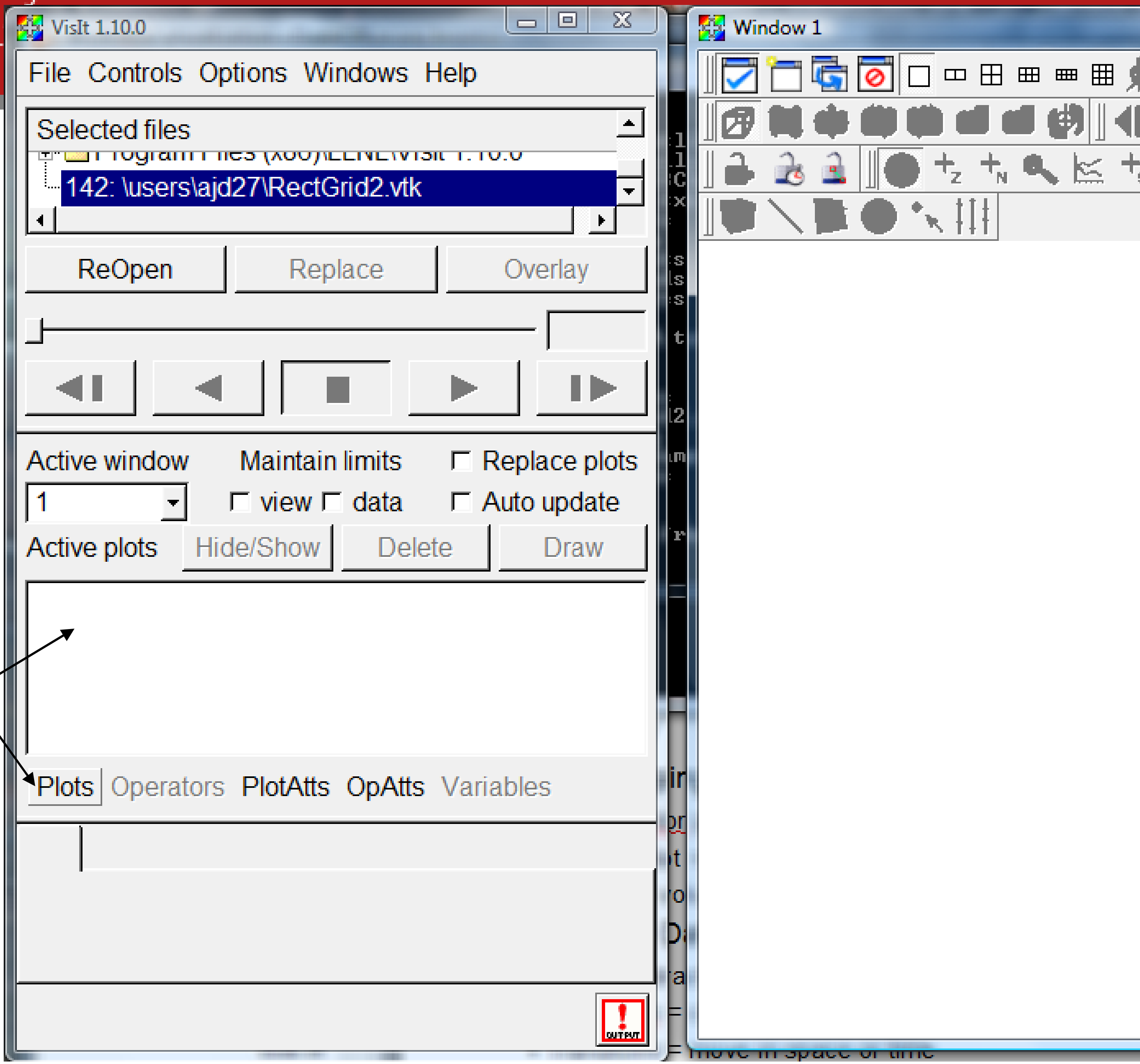
Visit on Spur

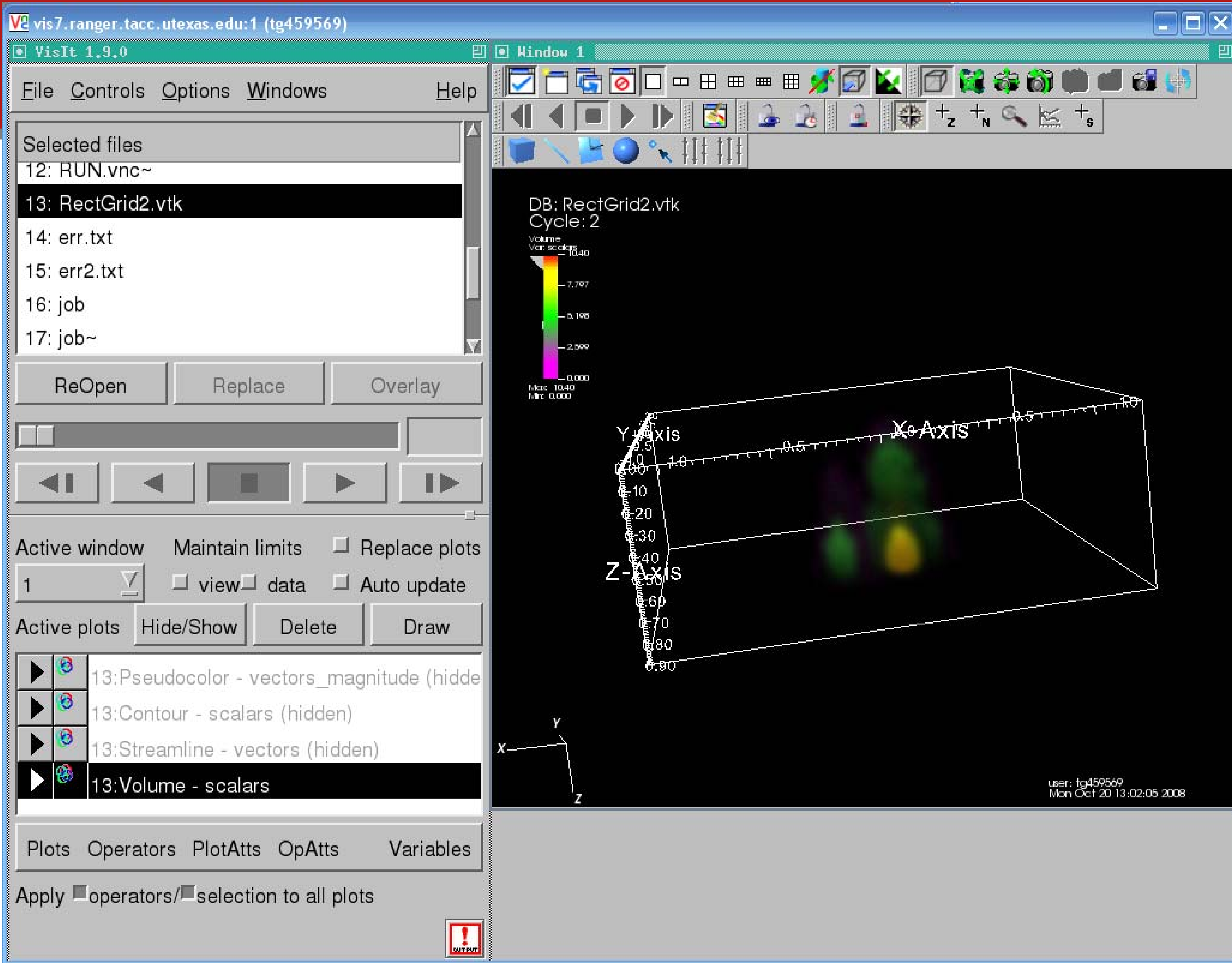
- Run it like Paraview, except “load module visit.”
- Terminology
  - Plot = Mapping algorithm
    - Pseudocolor plot = scalar color map
    - Surface plot = 3D isosurface of 2D data
    - Volume = volume rendered in 3D
  - Operator = Data manipulation algorithm
    - Slice = extract data
    - Resample = change data resolution
    - Transform = move in space or time

List of files in current directory. Select RectGrid2.

Click Plots to do something to the data.

Your plot will show as a line up here. Double-click the line to set plot properties.





vis7.ranger.tacc.utexas.edu:1 (tg459569)
VisIt 1.9.0
Window 1
File Controls Options Windows Help
Selected files
12: RUN.vnc~
13: RectGrid2.vtk
14: err.txt
15: err2.txt
16: job
17: job~
ReOpen
Replace
Overlay
Active window
Maintain limits
Replace plots
1
view
data
Auto update
Active plots
Hide/Show
Delete
Draw
13:Pseudocolor - vectors_magnitude (hidde
13:Contour - scalars (hidden)
13:Streamline - vectors (hidden)
13:Volume - scalars
Plots Operators PlotAtts OpAtts Variables
Apply operators/selection to all plots
DB: RectGrid2.vtk
Cycle: 2
X-Axis
Y-Axis
Z-Axis
user: tg459569
Mon Oct 20 13:02:05 2008

Now Try Some Plots

- Pseudocolor.
  - Select Plot button, choose Pseudocolor->Scalar.
  - Select Draw button to make it show.
  - Right-click in the display window to see View menu which lets you invert the background color from white to black.

Visit Isosurfaces

- Select Plot->Isosurfaces
- Show/Hide the last plot to see the new one.
- Double-click the isosurfaces line on the left to set its properties.

Visit Streamlines

- Select Plot->Streamlines.
- Not much showing? Double-click the line on the lower left to try different sources. Make sure they are inside the bounding box of the dataset. Increase the density of the streamlines or their length.

Volume Rendering

- Volume Rendering is another Plot option.
- The properties for this plot control not only color but opacity, which is crucial for volume rendering. Try different opacity curves.

Task Parallelism

	Timesteps					
Processes		1	2	3	4	5
	1		Read file 1	Isosurface	Cut plane 1	
	2			Read file 2	Streamlines 2	Render
	3	Read file 3	Triangulate 3	Decimate 3	Glyph 3	

Running an MPI-parallel isosurface algorithm on several nodes is less efficient than running it on a single node, although it takes longer. Task parallelism can offer more speed if it is balanced.

Pipeline Parallelism

- If process 1 can read faster, process 3 better at rendering.
- Increases render rate without decreasing efficiency

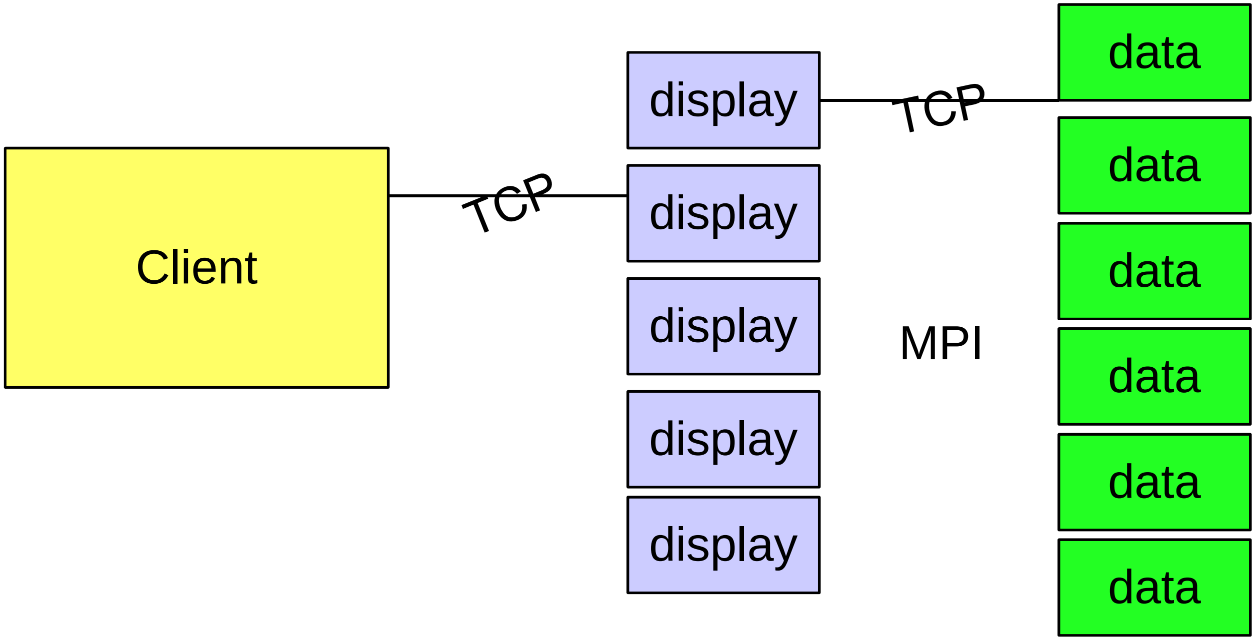
	Timesteps					
Processes		1	2	3	4	5
	1	Read file 1	Read file 2	Read File 3		
	2		Isosurface 1	Isosurface 2	Isosurface 3	
	3			Render 1	Render 2	Render 3

Data Parallel

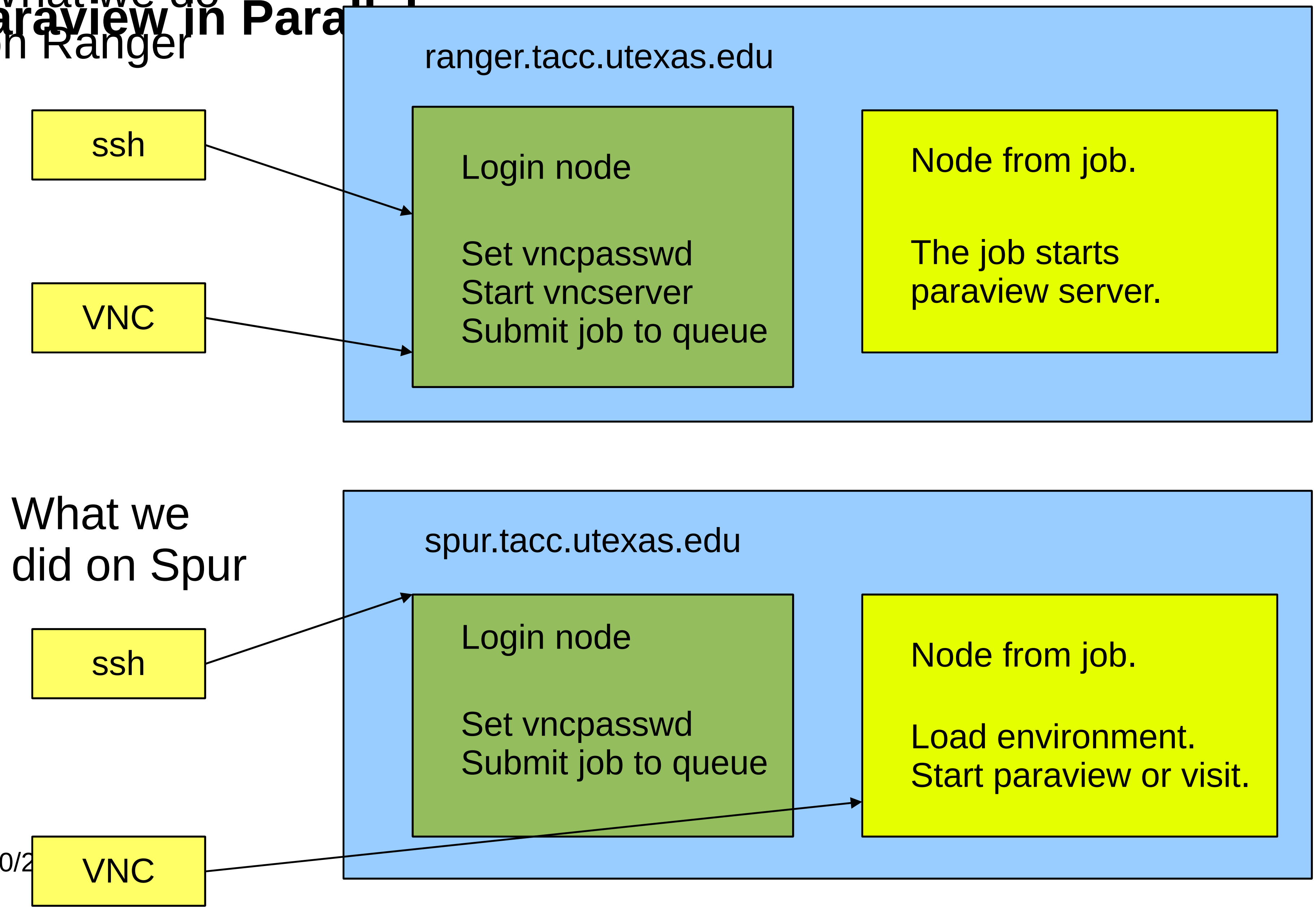
- When one computer can't handle a filter.
- Depends on communication required.

	Timesteps			
Processes		1	2	3
	1	Read partition 1	Isosurface partition 1	Render partition 1
	2	Read partition 2	Isosurface partition 2	Render partition 2
	3	Read partition 3	Isosurface partition 3	Render partition 3

Paraview in Parallel



Paraview in Parallel



Setup VNC Server on Ranger

- Ssh login3.ranger.tacc.utexas.edu
- Vncpasswd
- Vncserver # Look at display number it shows
- Connect with VNC from your terminal.

Note Ranger has a /home, /work, /scratch with cd, cds, and cdw.

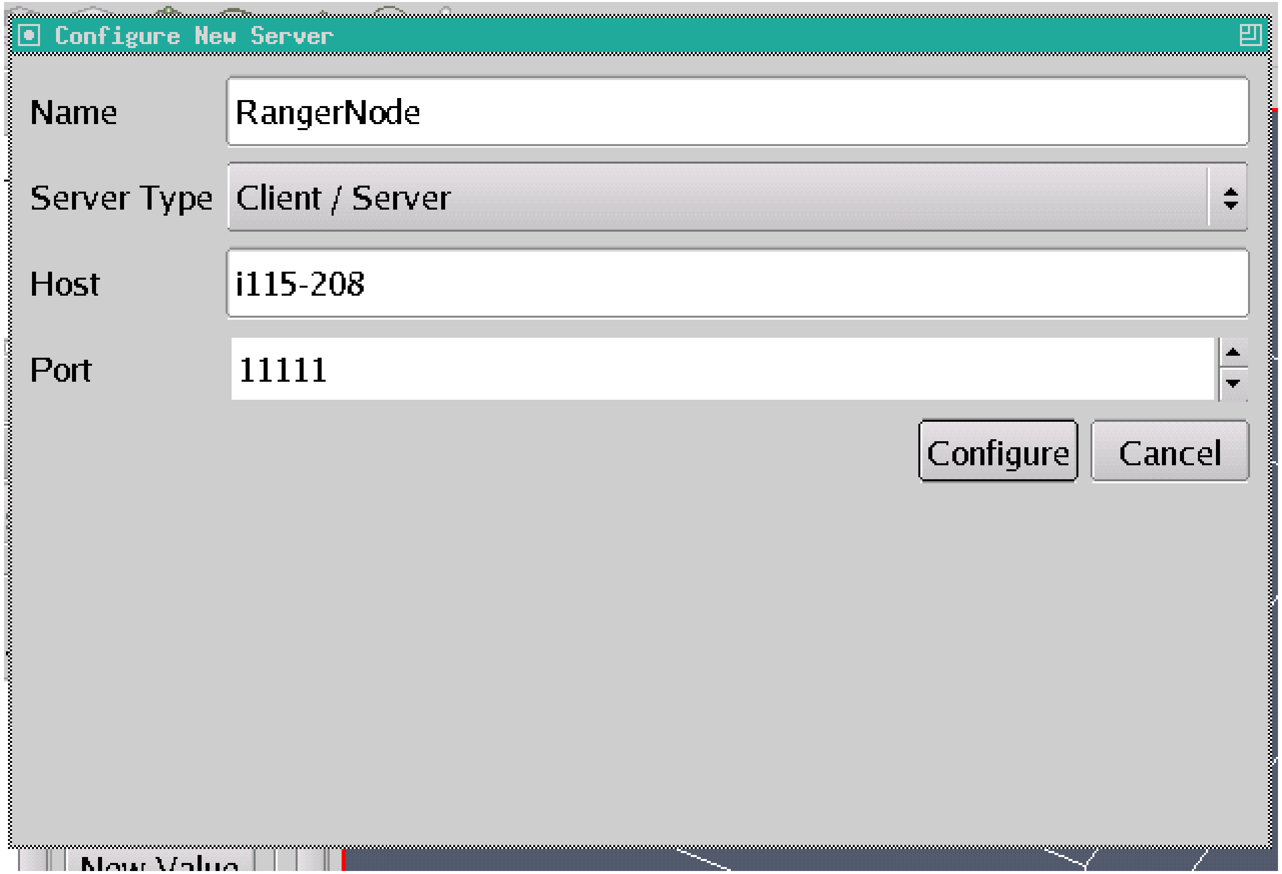
Start Paraview on Login3

- qsub job
- (Try module load paraview out of order.)
- Module delete mpich2
- Module swap pgi intel
- Module load mvapich/0.9.9, then vis, then mesa, then qt, then paraview.

Find Your Ranger Server

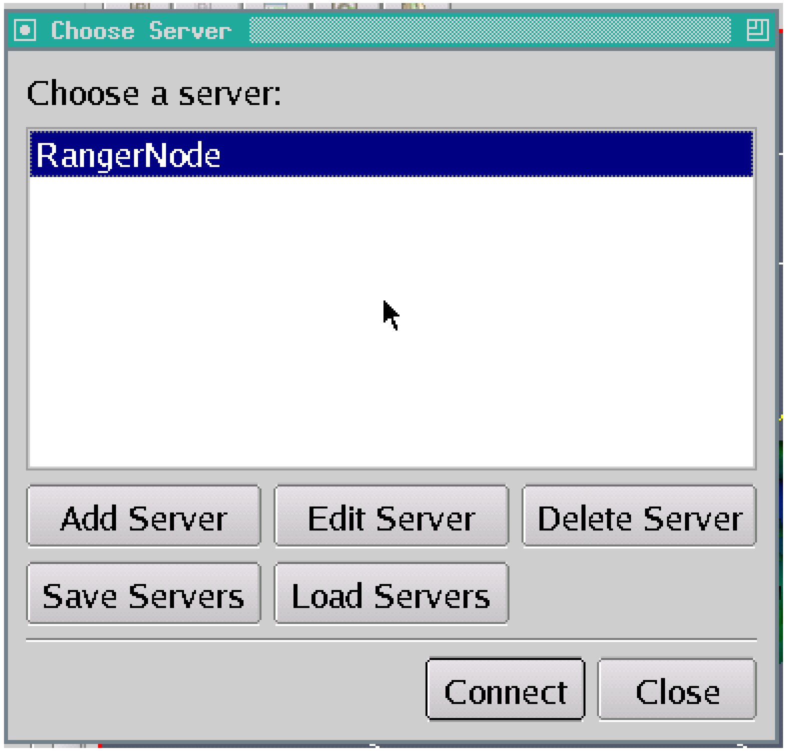
- showq -u # to see when it runs
- qstat   # to see what node you got
- i115-406.ranger.tacc.utexas.edu means i115-406 is the node.
- Ranger% vglrun paraview

In Paraview, go to File->Connect.
Click “Add Server”
Enter the node in “host”.
Click Configure.
Under Configure, select Startup Type: Manual and click Save.



We choose “manual” because the job we submitted already started the server.

Select the server.
Click connect.



Errors! But they are OK.

Open RectGrid2.vtk.