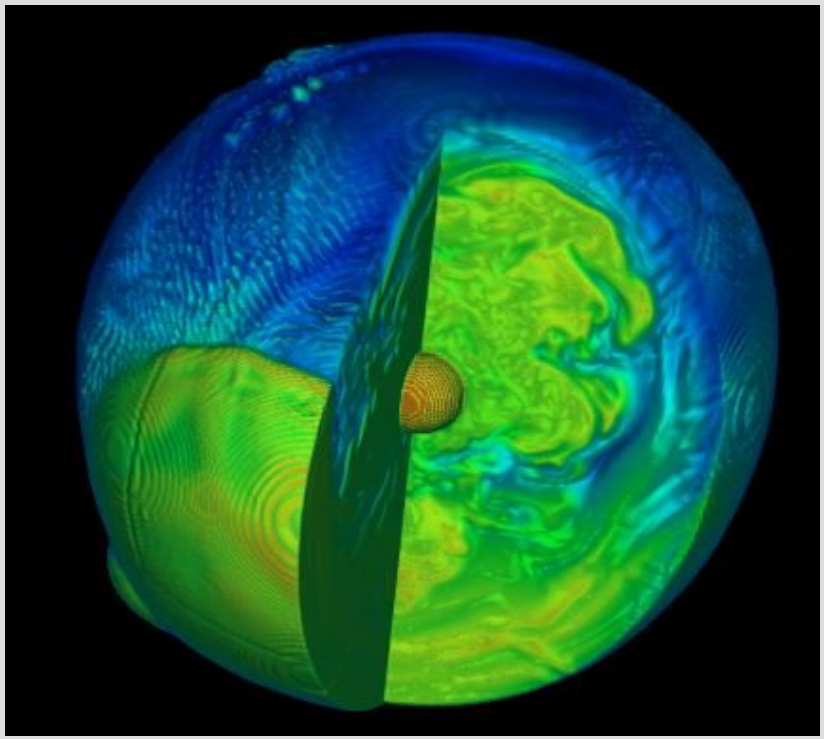
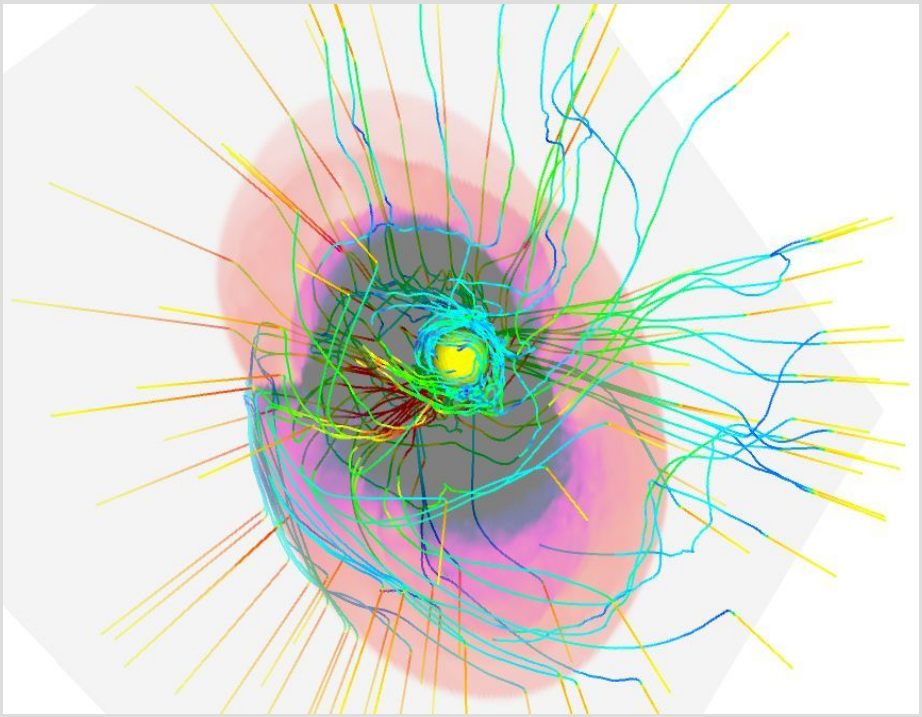
Applications of SDM in Astrophysics

Elliot Peele
Department of Physics
North Carolina Sate University
ebpeele2@ncsu.edu



Scientific Data Management - All Hands Meeting
Salt Lake City, UT
March 2nd, 2005

Neutron Star Spin-Up Discovered with 3D Simulations on Cray X-1

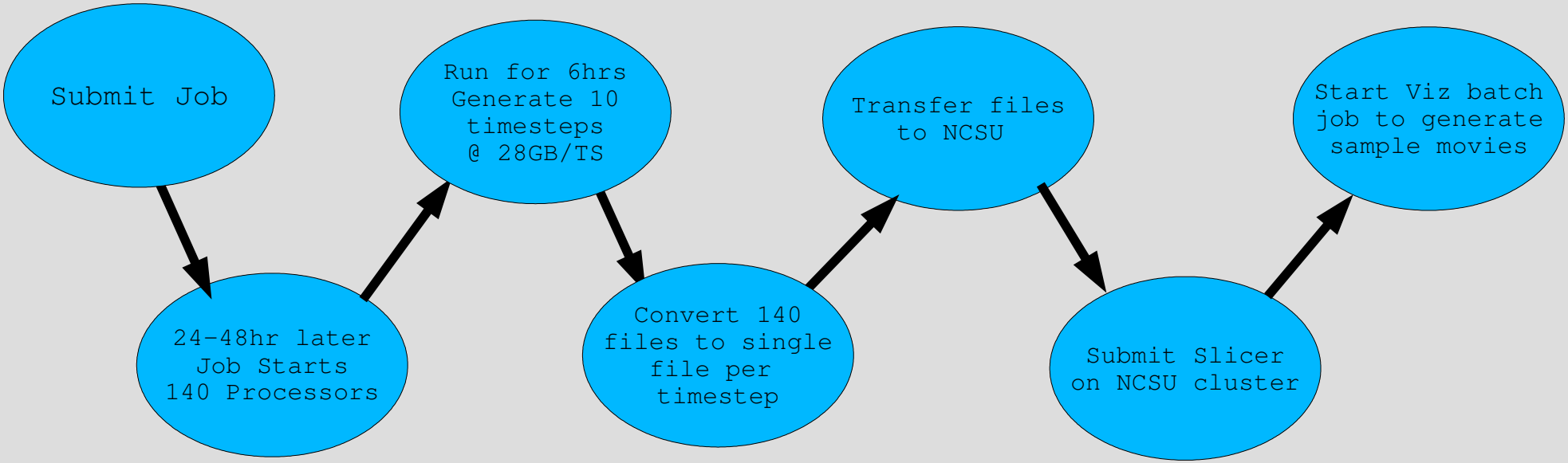


The image at left was generated from a 3D hydrodynamic simulation using 600 million zones. This simulation was computed on the Cray X-1 at the Oak Ridge National Labs. The slice through the center is shaded by pressure to show the extent of the young supernova shock. Outside this shock the stellar material is falling radially in towards the center of the collapsing star, as shown by the yellow streamlines in the outer regions. The flow is decelerated at the unstable shock (streamlines change from yellow to blue) and deflected from radial infall. These 3D simulations show that the flow develops into a strong, stable, rotational flow (streamlines wrapped around the proto-neutron star). The flow deposits enough angular momentum on the inner core to produce a young neutron star spinning with a period of only a few milliseconds.

The Simulation

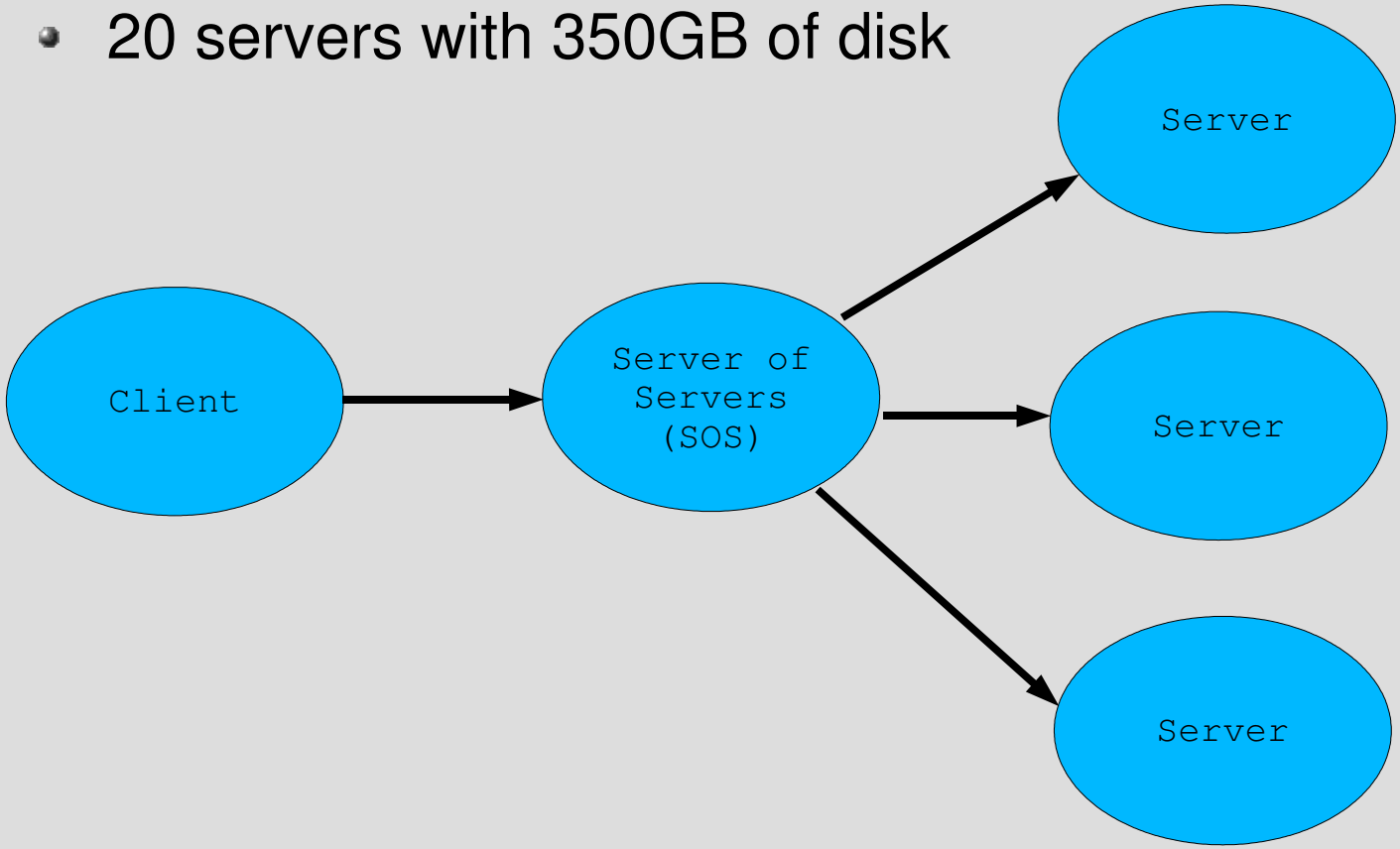
- Fortran 77/90
- 3D Super Nova Simulation
  - 1080 x 1080 x 1080 x 3 variables
- one time step = 28GB NetCDF
- one 6 hour run = 10 time steps
  - at the end of each run we write out a set of data that allows the simulation to pick up where it left off the next time it is run.
- one full simulation = 300-500 time steps = 8.5-13.7TB

Current Workflow

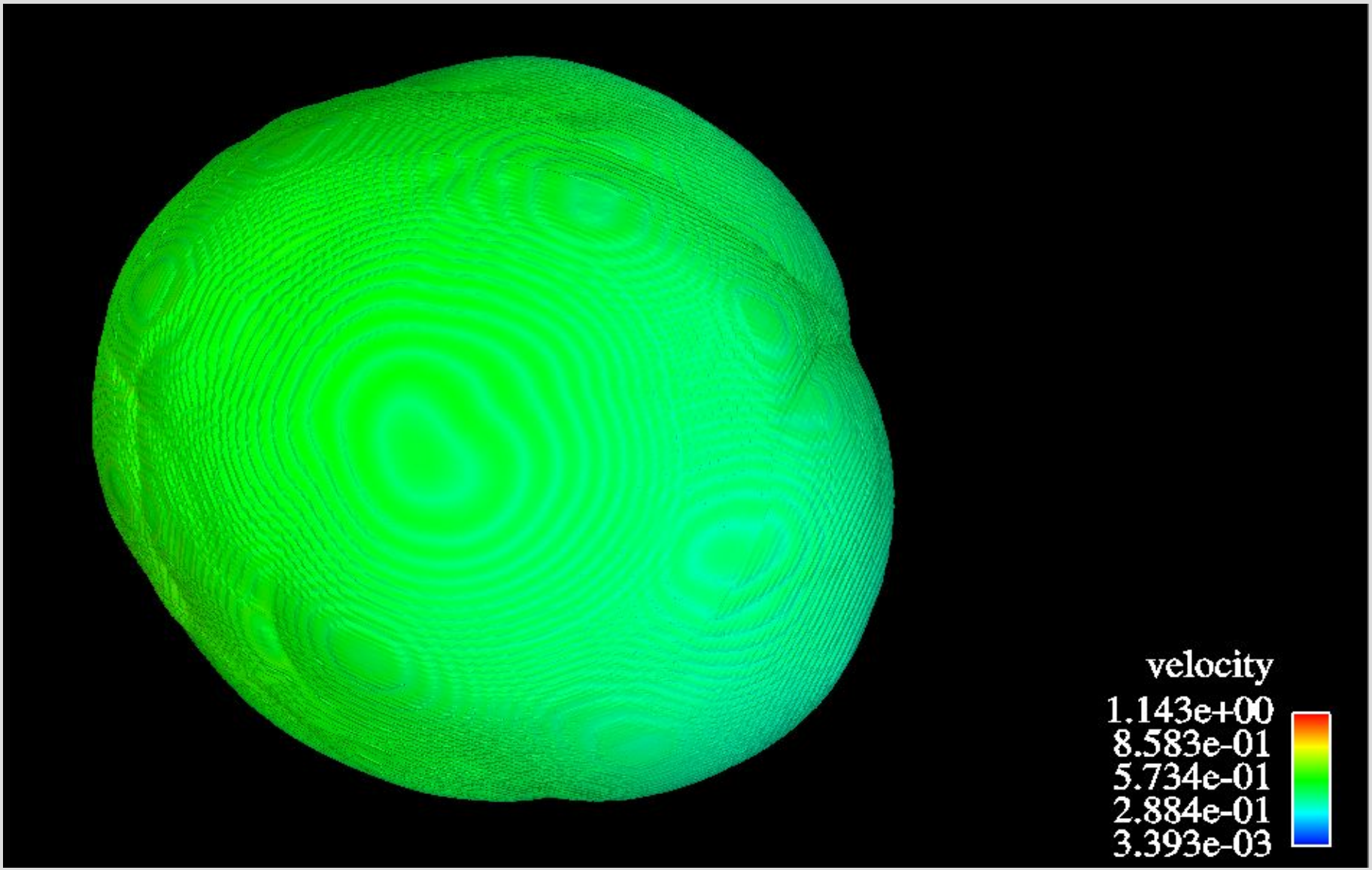


Visualization

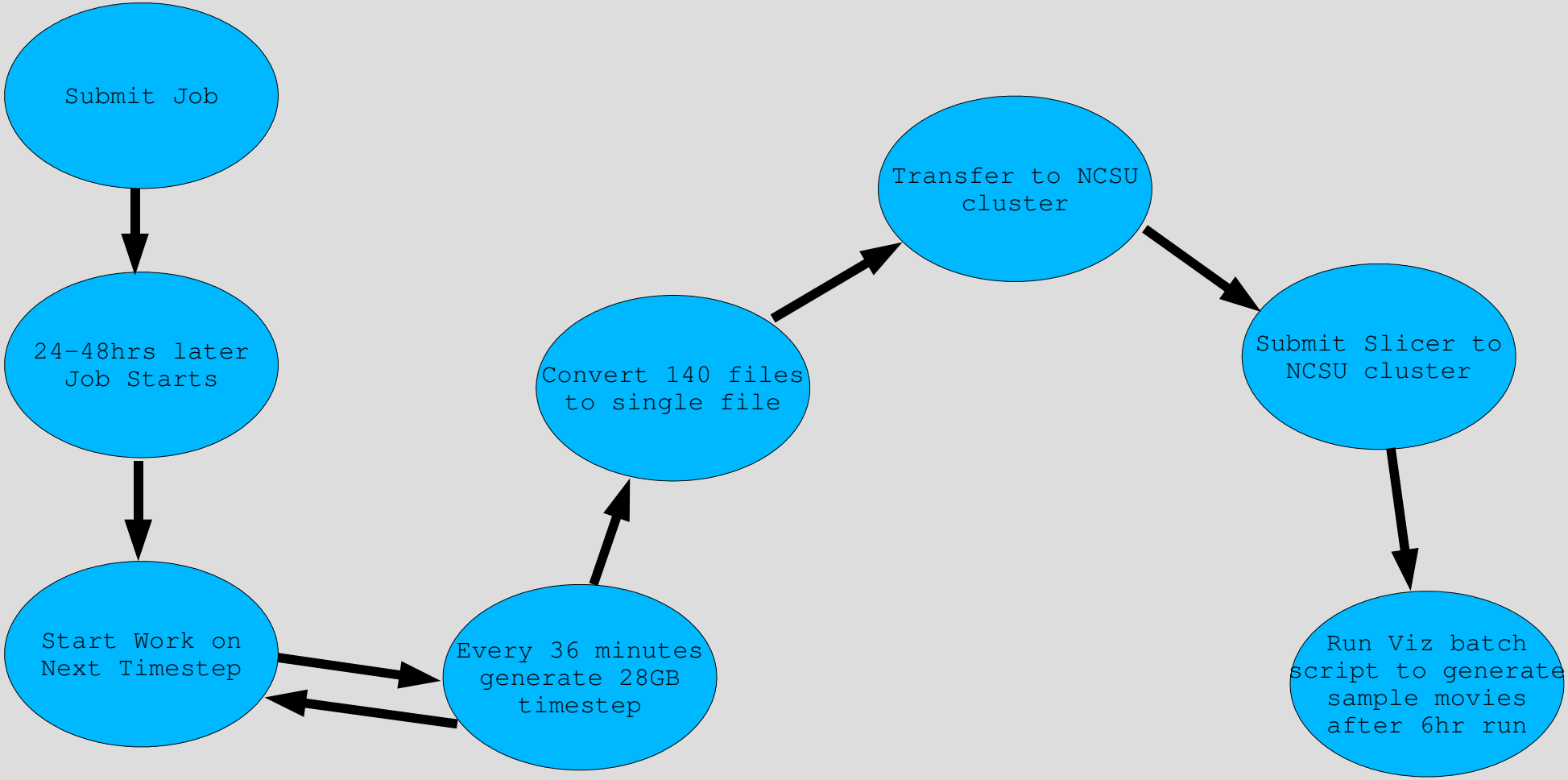
- Ensight - http://www.ensight.com
  - produced by CEI out of Apex NC
- Can be run in batch mode or interactive mode
- 20 servers with 350GB of disk



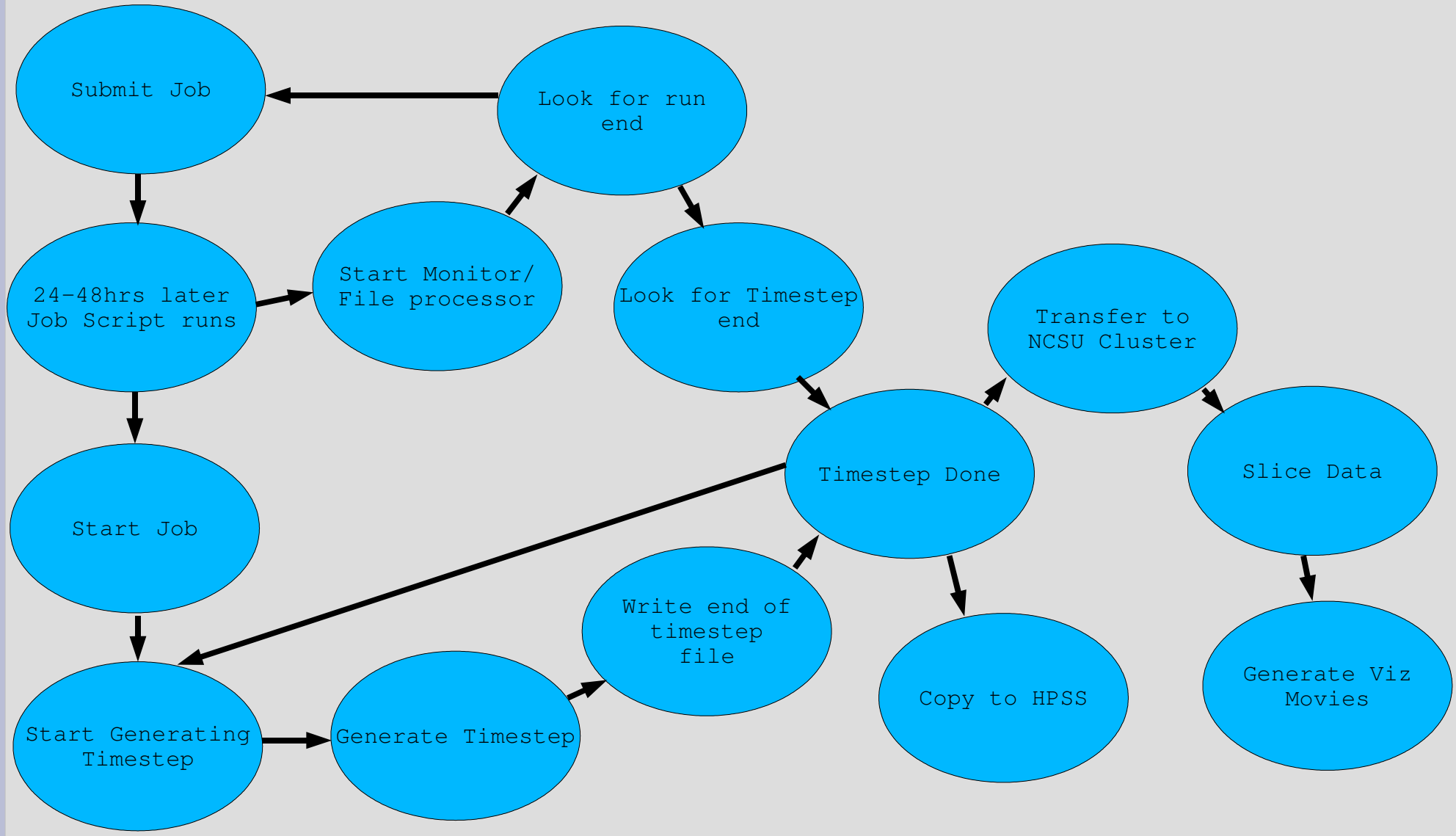
864^3 Surface of Accretion Shock



Short Term Workflow



Long Term Workflow



Currently Utilized Technologies

- BBCP – used to copy data from ORNL to NCSU
- Ensight – visualization software
- SPA – Scientific Process Automation
- NetCDF – portable file format
- LoRS Tools

What we are looking at

- pNetCDF
  - would allow us to remove several steps from the current workflow
- pNetCDF -> LoRS
  - combines many steps into one
  - doesn't really have error checking
  - not finished
- PVFS2
  - mostly needed to utilize pNetCDF
  - provides single large file system for data store

What we are looking at
(cont.)

- Parallel GridFTP
  - useful if we need to move large amounts of data from ORNL
  - optimal for taking advantage of available bandwidth

Needs

- More Parallel Storage
  - currently we can only store one full simulation (~8TB and growing)
  - HPSS is slow
- More storage on compute nodes
  - locally store nodes portion of the data for visualization
  - can only store one model on nodes
- More memory on compute nodes (currently 2GB)
  - visualization software running out of memory at full resolution

Needs (cont.)

- Highly available compute cycles
  - as more users are added to the X1 it will be harder to get the necessary compute time
  - queue time is long
- Tracking of data generation
  - where all of our data is
  - what simulation each file belongs to
  - what path did the file pass through

Bottlenecks

- Highly available mass storage
  - could be at NCSU or ORNL
  - HPSS isn't fast enough
- CPU cycles
  - queue time is long although better than some other machines
  - X1 not growing at the same rate as the user population

Questions?