HPC Analytics

Dan Stanzione
Fulton High Performance Computing

dstanzi@asu.edu
2/20/05

ASU Arizona State University
Ira A. Fulton School of Engineering
High Performance Computing

Theme

- Gap between **data** and **knowledge** (as has been discussed here before)
- High Performance Computing continues to exponentially increase our ability to generate data
  - This can be an enabler of new science...
  - ...but also a huge obstacle
  - ...or an excuse not to think

Outline

- How much data is “large”.
- Evolution of system design to deal with large data
- What to do with it all - Analytics

How Much Data?

- What is a “large” dataset nowadays?
- My current machine:
  - 2+ Tflops
  - Network bisection bandwidth ~1Tb/s
  - I/O subsystem writes ~500MB/s
    - (30 GB/minute)

How Much Data?

- Mars project: ~60TB
- One ASU faculty member has contacted me about a ~2 Petabyte dataset.
- A Chilean observatory can produce more than 1TB an hour (12 hours data must be processed before next pass starts...)
- A potential Australian array telescope would produce multiple EXABYTES per year by 2010.
- Not unique to astronomy...

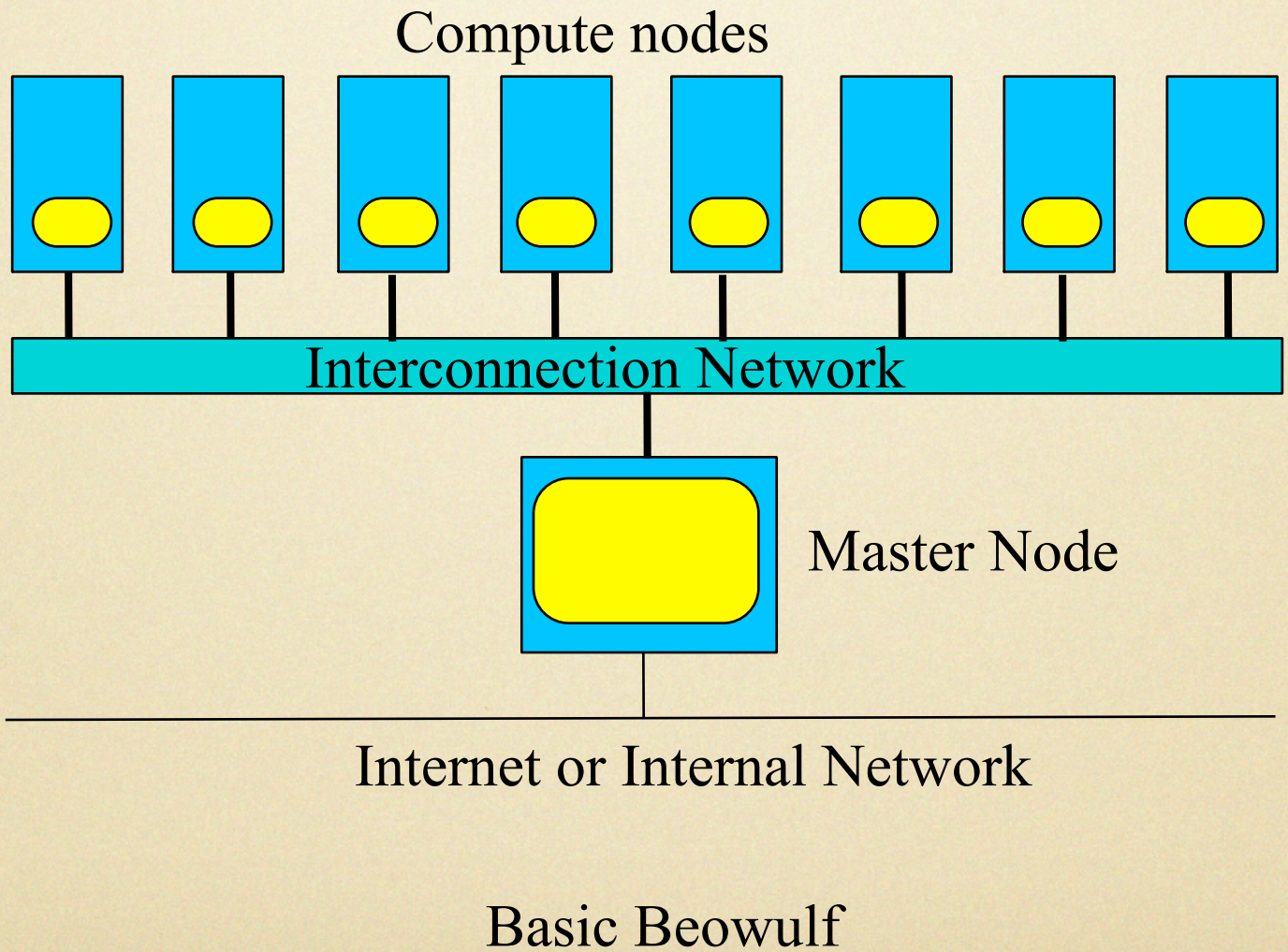
How Much Data?

- Machines will be constructed in *next 12 months* with several tens of thousands of processors (hundreds of TF)
- Network bandwidth >10TB/sec
  - 1PB/2 minutes
  - 1 Exabyte per 30 hours
  - 1 Zettabyte during machine 3 yr. lifetime (yottabytes are next, if anyone's counting...)
- Google has much more computation, much less network/flop

Evolution of Storage Systems

- Evolution at all levels:
  - RAW/Text Files -> Hierarchical Formats -> Schemas -> Database
  - Filesystems -> LVM -> Parallel Filesystems -> Global Name Space/Storage Request Brokers
  - Single disk volumes -> RAID1-5 -> RAID 10 -> Storage Hierarchies

HPC Storage Hierarchy

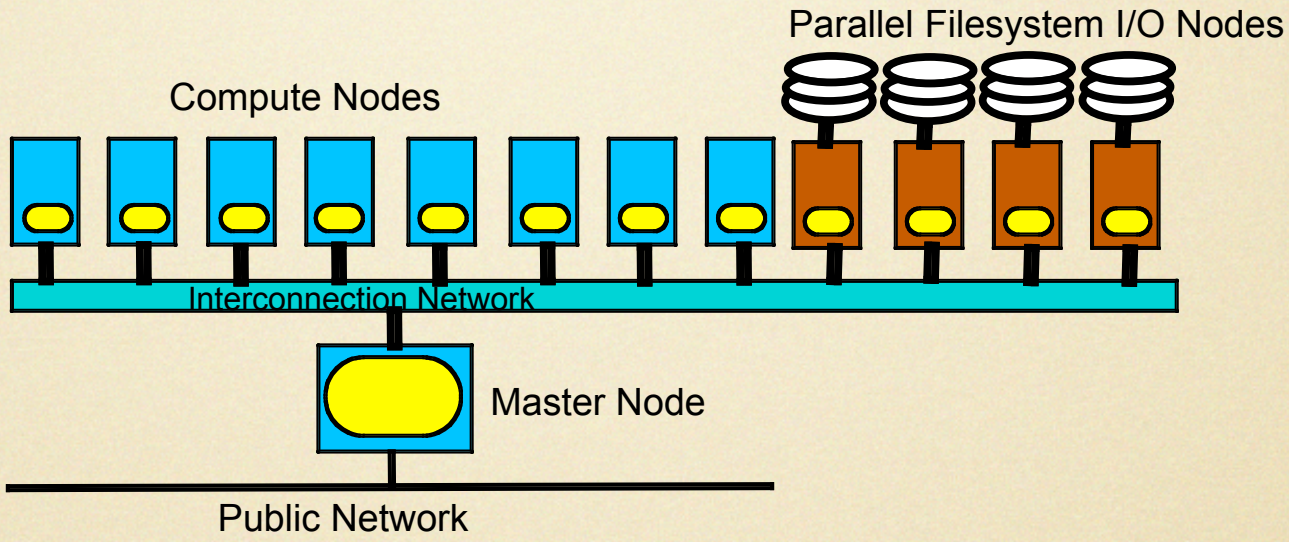


Basic Beowulf

Tier 1 Storage

In Cluster High Speed Scratch

Beowulf Cluster



Parallel Filesystems support this: PVFS, Panasas, Lustre, IBRIX -- MPI I/O is the interface

ARIZONA STATE UNIVERSITY

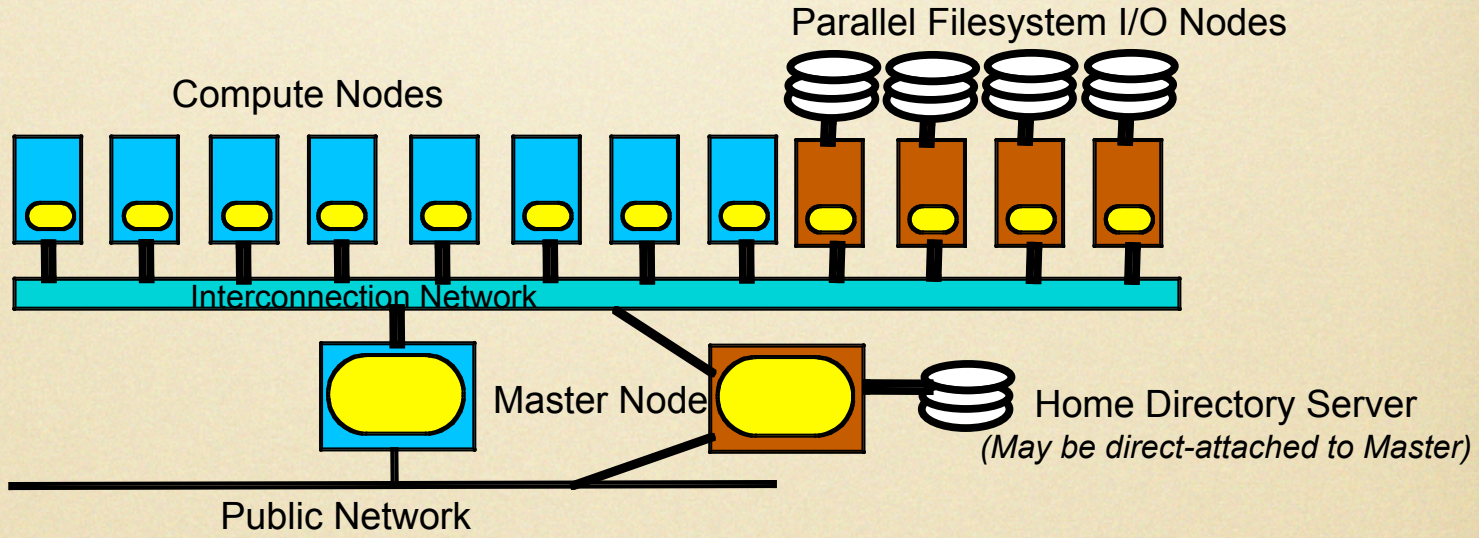
IRA A. FULTON SCHOOL OF ENGINEERING

HIGH PERFORMANCE COMPUTING

Tier 2 Storage

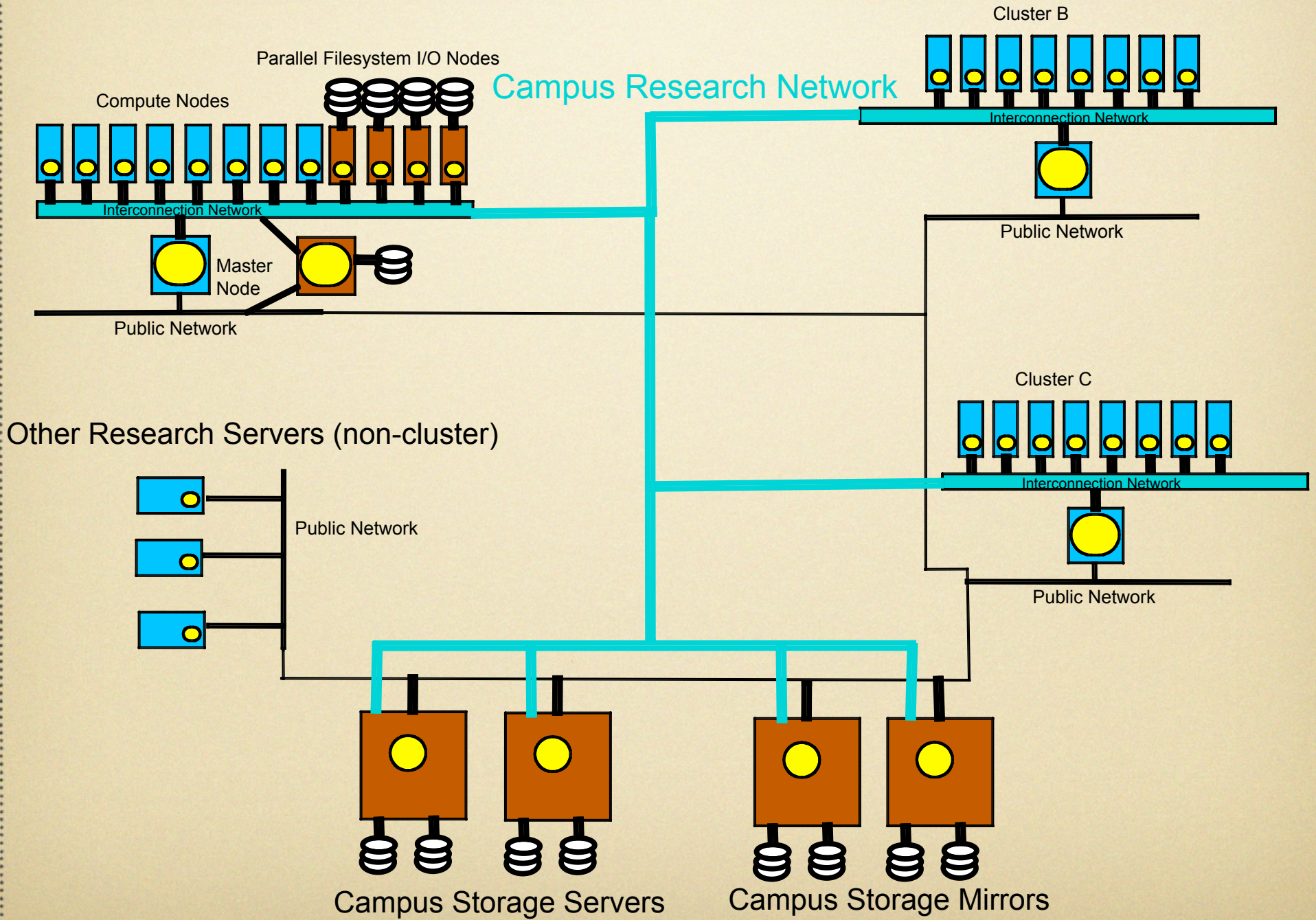
Shared Home Directories

Beowulf Cluster



Tier 3 Storage

Campus-Wide Research Storage



We can build Multi-PB Storage Systems - Now What?

- Applications spit out lots of this data (or sensors/sequencers/instruments wrapped in applications).
- Status Quo:
  - Applications codes generate FORTRAN unformatted or ASCII text data to a (multitude of) files
  - Some domain exception (.pdb, gridgen)

Three problems:

- Too many files (my worst offender has 750,000 - find anything useful in that).
- Files too big (one student generated 700GB in 18 hours)
- Too many formats (can't connect weather and ocean, application and visualization).

Things are happening

- Broad Domain Frameworks taking hold:
  - e.g. ESMF (Earth System Modeling Framework)
    - Connect WRF (climate) to ADCIRC (Ocean)
  - Hierarchical, standard, descriptive data formats
- *Broader introduction of **metadata** is the key...*
- This is the right trend, but has costs...

Costs of Frameworks

- Application complexity goes way up
- Converse -> value of applications written "outside"community goes way down.
- XML is not the most efficient format in the world...

XML:

```
<particle> <coordinates> 10 0 10 </coordinates>
        <velocity>
                <x> 12  </x>
                <y> 9 </y>
                <z> 8 </z>
        </velocity>
</particle>
```

(~100 BYTES)

FORTRAN Raw:

0a000a0c0908

(6 bytes)

Costs of Frameworks

- Enterprise-class backed-up storage:
  - ~$10,000/Terabyte
- Cost of 10-1 inefficiency on one PB of raw data:
  - $100,000,000.00
- In fairness, compressed XML mitigates a fair amount of this... but an app-specific binary format will always win

HPC Analytics

- We can build systems, we can make filesystems, we can create well-ordered files
- This can roughly be called "Data Management"
- Well-ordered data is a foundation, but still not knowledge.
- The next phase is the emerging field of **Analytics**

Analytics

- SC05 - "HPC Analytics Challenge" 11/05
  - "...showcase innovative techniques of rigorous data analysis..."
- Dept. of Energy - Visual Analytics Center solicitation 10/05.
  - PNNL NVAC (Nat'l Visualization and Analytics Center)
    - Recommended Reading:
      - "Illuminating the Path" - National R&D Agenda in Visual Analytics http://nvac.pnl.gov/agenda.stm

Analytics, acoording to the "Path":

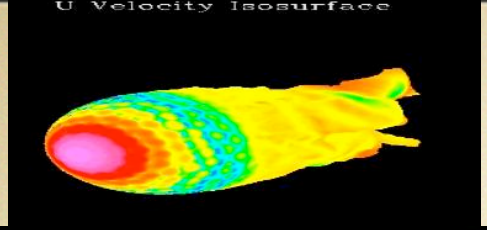
- The science of analytical reasoning
- Visual representations and interaction techniques
- Data representations and transformations
- Production, presentation, and dissemination.

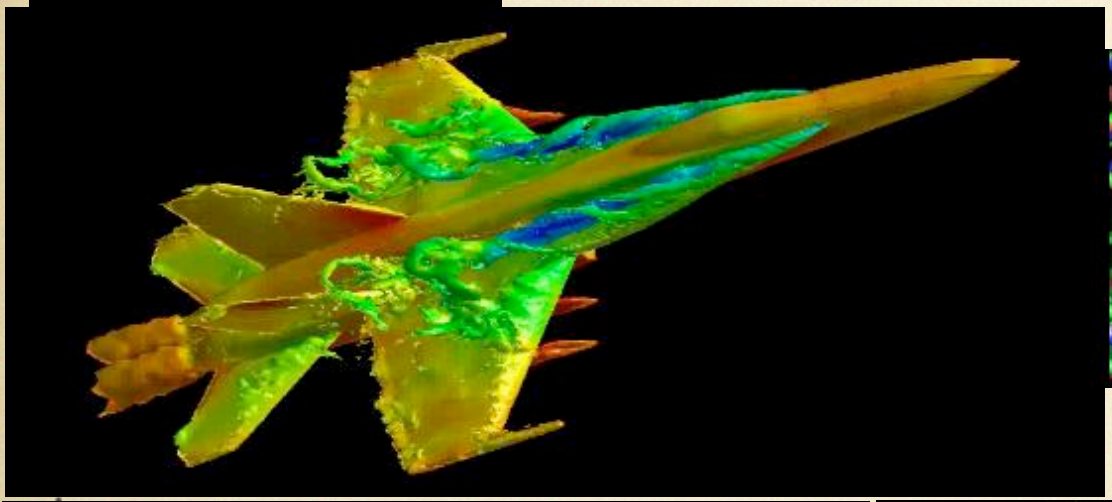
State of Analytics

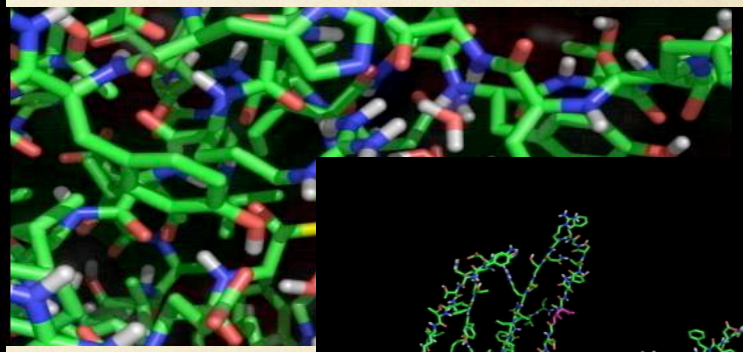
- At SC05, all five finalists did Visualization
  - Not an expansive view of analytics...
  - One used data mining to produce visualizations
- While much, much quality work has been done in visualization techniques,
- ...visualizations are still used as much for fundraising as science

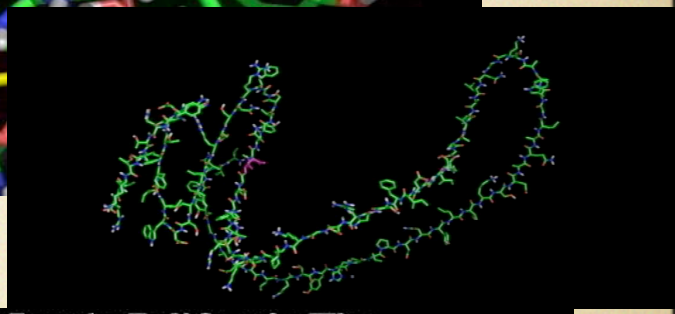
Of course, *I* wouldn't use visulizations for this...

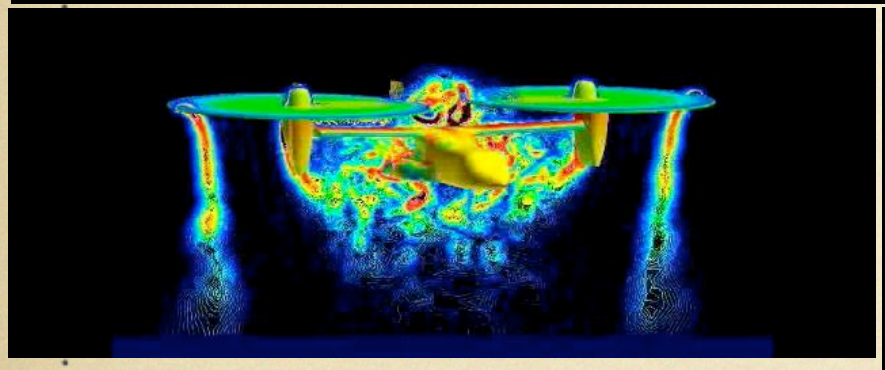
HPC Applications

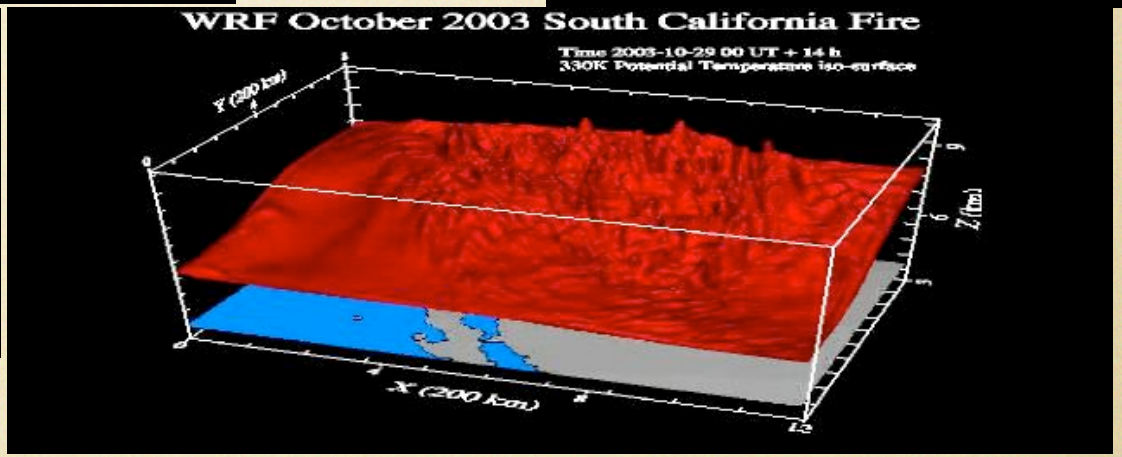


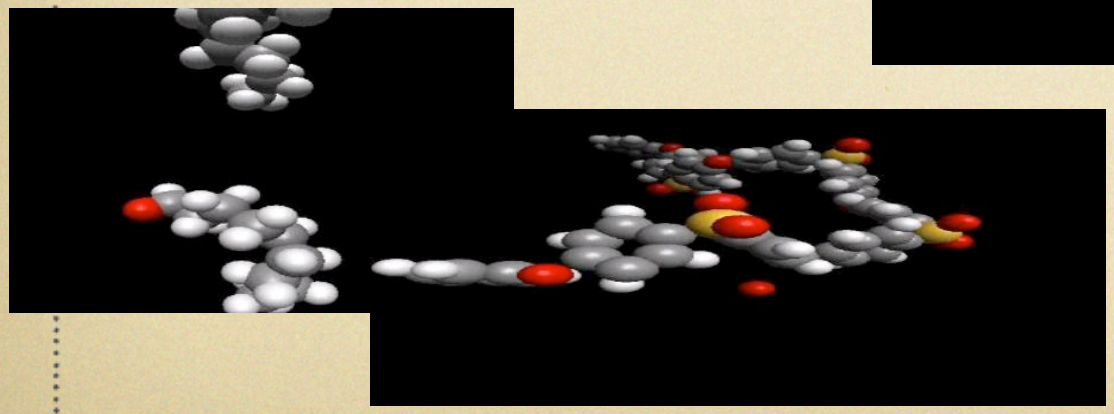


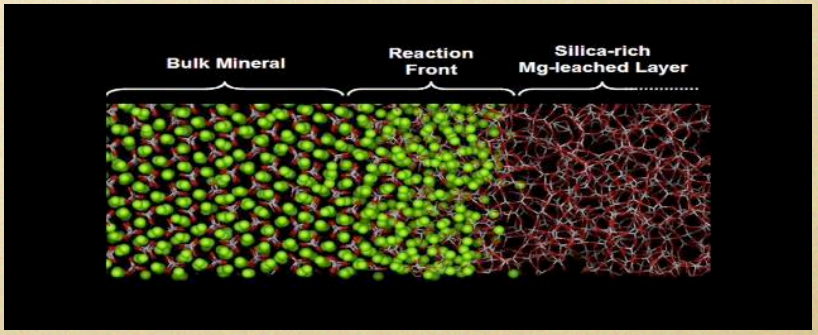


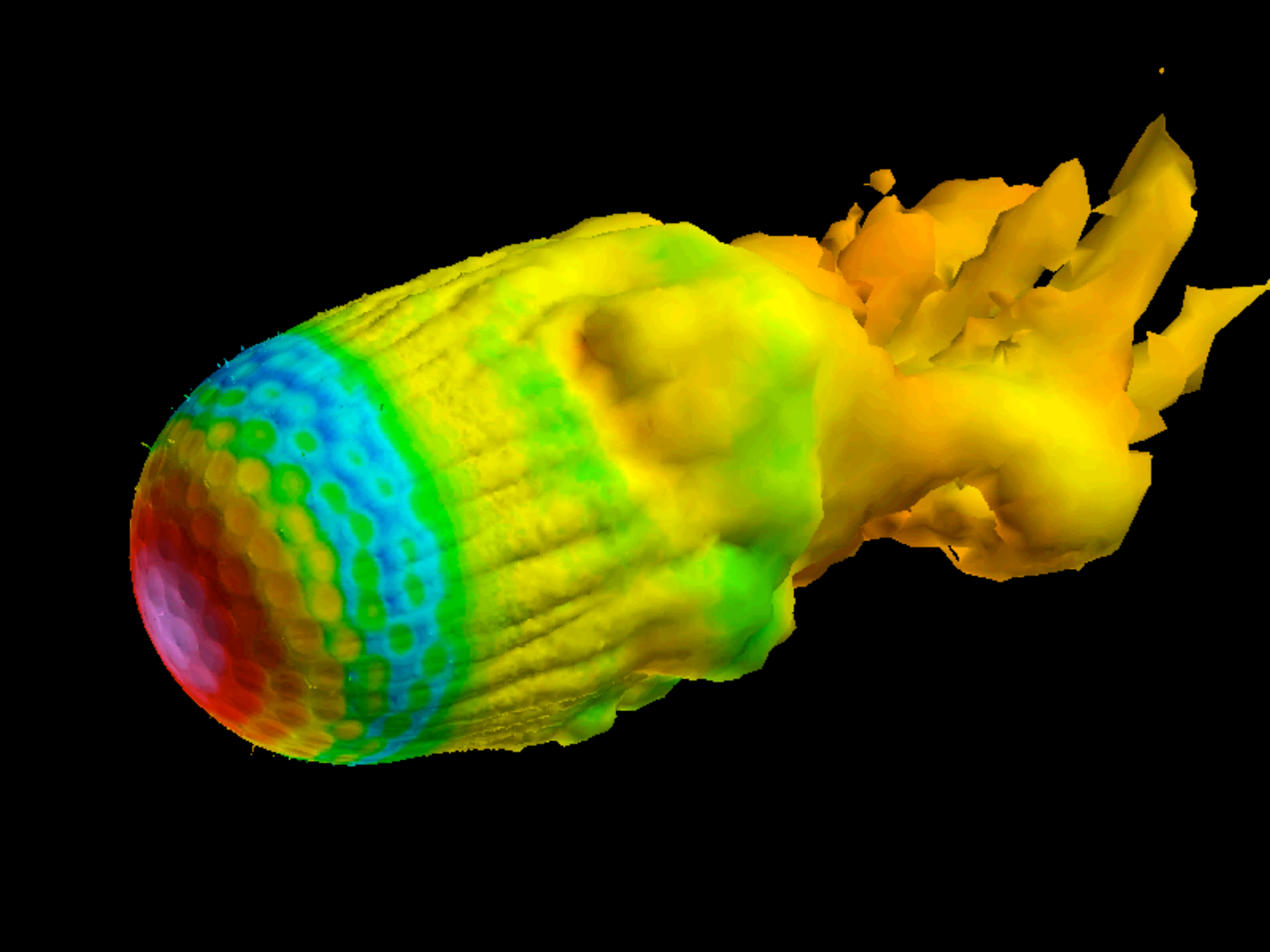












Visualization Advancing

- 3D visualization does add something
  - Decision Theater
- Formats are a key to making this routine.
- More tools beyond Excel, Matlab
- Need to accelerate to real-time, “what-if” scenario
- Hierarchy matters here - don’t render whole earth at 30cm resolution

Analytics Beyond Visualization

- Databases are a key to the HPC future - See Dr. Chen's earlier talk for an excellent introduction
  - Large databases of small records well understood
  - Large databases of large, sparse records of ill-conforming data not understood.
- Experimental Management tools increasing in value
  - Frameworks for parameter study, goal-directed search

Analytics Beyond Visualization

- Two more technologies must be imported from other fields:
  - Data Mining (database-enabled)
    - In large datasets, the *trends* are the knowledge
    - Acxiom is a good model (and, gets them out of the junk mail business).
  - Search
    - One Word: Google
    - Pre-(multi)-indexing, divided search space
    - search multi-PB space in 0.01 seconds... by using a massive cluster to do most work ahead of time.

Takeaways:

- Intelligent I/O
- Standard Formats
- Hierarchy - multiple views of data
- Database/Data Mining/ Search
- Visualization
- All of the above require more sophisticated application codes, more use of tools:
  - *Computational Science Literacy*