

# Scalability Challenges Facing Visualization R&D



James Kohl

Computer Science and Mathematics  
Oak Ridge National Laboratory

SCaLeS Workshop ~ Visualization Breakout Session  
June 24-25, 2003



# Barriers to Scalable Visualization

(like déjà vu all over again... ☺)

---

- I/O and Data Movement
- Dedicated Resources for Visualization
- Interactivity vs. Batch Mode
- Data Storage ~ Size and Bandwidth
- Degree of Parallelism and Algorithm Scalability



# I/O and Data Movement

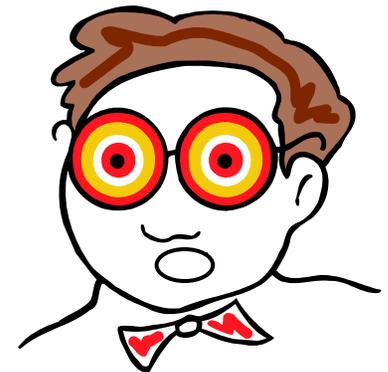
---

- I/O ~ Low Priority in Large System Design
  - Limited number of I/O nodes, network links
  - Data movement degrades overall system throughput
    - Idle time waiting for data to clear I/O channels...
  - Simply hard to get data out of “BAMs”
- Analyze and Render Data *In Situ*?
  - Big resources still not available...
  - Scientists want to archive “original” data anyway...

# Dedicated Visualization Resources

---

- Discrepancy between “big iron” and viz
  - Need “matching” / comparable resources
  - “Expensive stereo... cheap paper cone speaker.”
- BAM “allure” distracts from practicality
  - Reduces usefulness / benefit to actual science
- Sociological Issue:
  - Funding model...
  - Perspective of domain scientists...



# Interactivity vs. Batch Mode

---



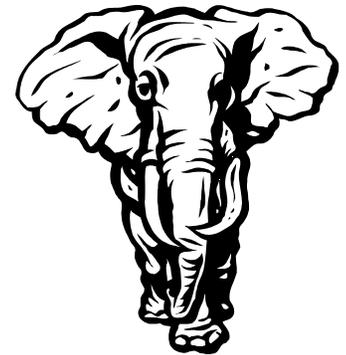
- Visual Exploration is Naturally Interactive
  - Statistical data mining only goes so far
    - Heuristically based, assumes target is known...
- “Batch” mode common to big iron systems
  - Minimal “interactive” nodes... (*throughput*)
- Need powerful back-end to enable a dynamic data exploration path for front-end scientist
  - Fundamental policy\* change at computer centers

# Data Storage ~ Size & Bandwidth

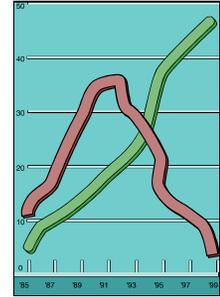
---

- Obvious Issue

- Applications organize data for efficiency in algorithmic access patterns
- Data analysis and visualization are different...
  - Streaming vs. monolithic data movement
  - Must specially cater network latency and bandwidth
  - Re-organize application data for viz / analysis...
    - Optimize search & lookup, storage methodologies



# Degree of Parallelism and Algorithm Scalability



- Often overlooked problem...
  - No “real” scalability to most algorithms!
- Data Analysis and Reduction
  - Almost strictly serial in nature, few parallel solns
- Parallel Rendering
  - Many algorithms, few scale beyond 10s of CPUs?
- More hardware resources not (nearly) enough!