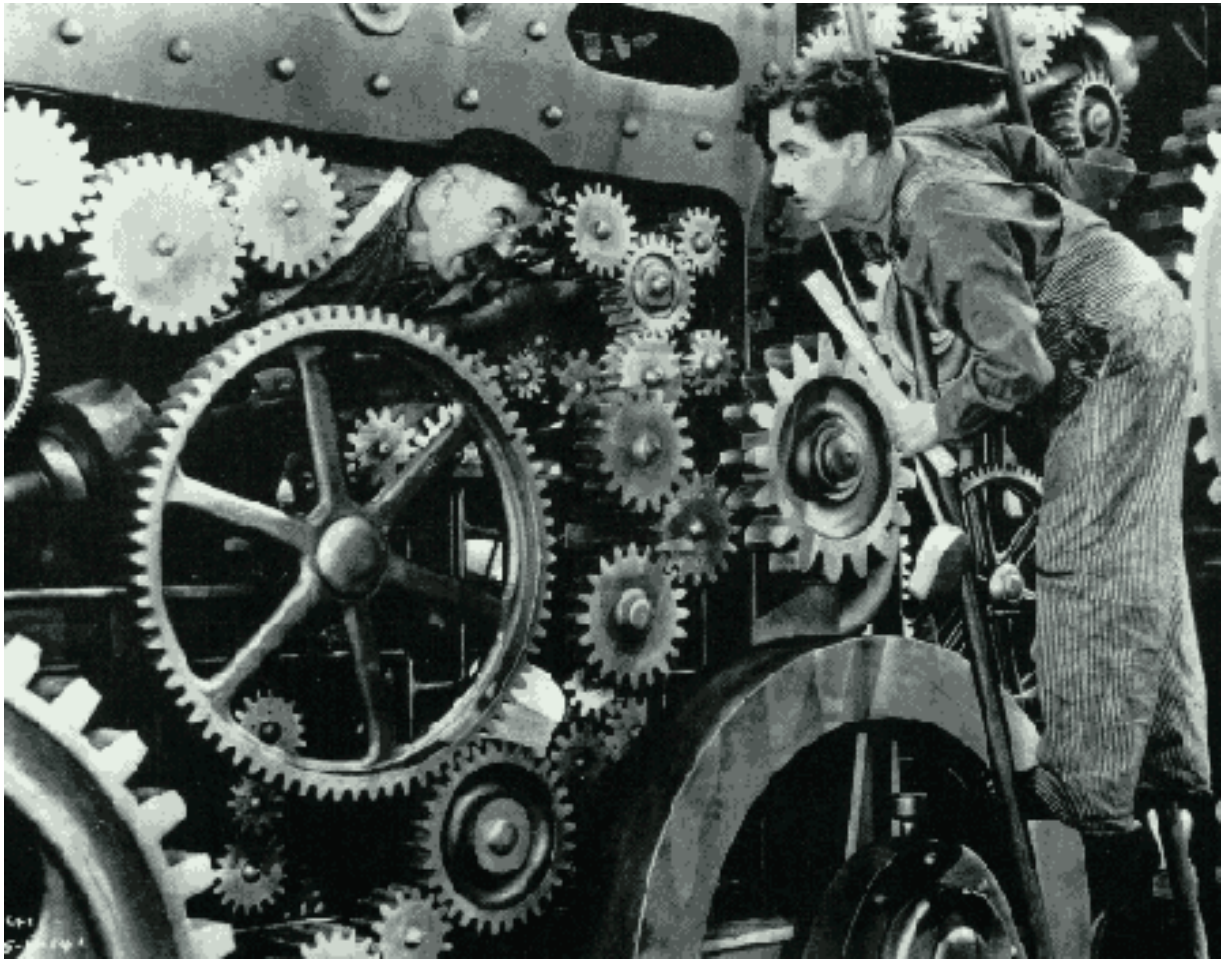
Introduction to High Performance Computing

Jeff Falgout
Janice Gordon

Shelley Knuth
Pete Ruprecht
Thomas Hauser
Monte Lunacek
Bruce Loftis

July 2014

HPC Has a Lot of Moving Parts

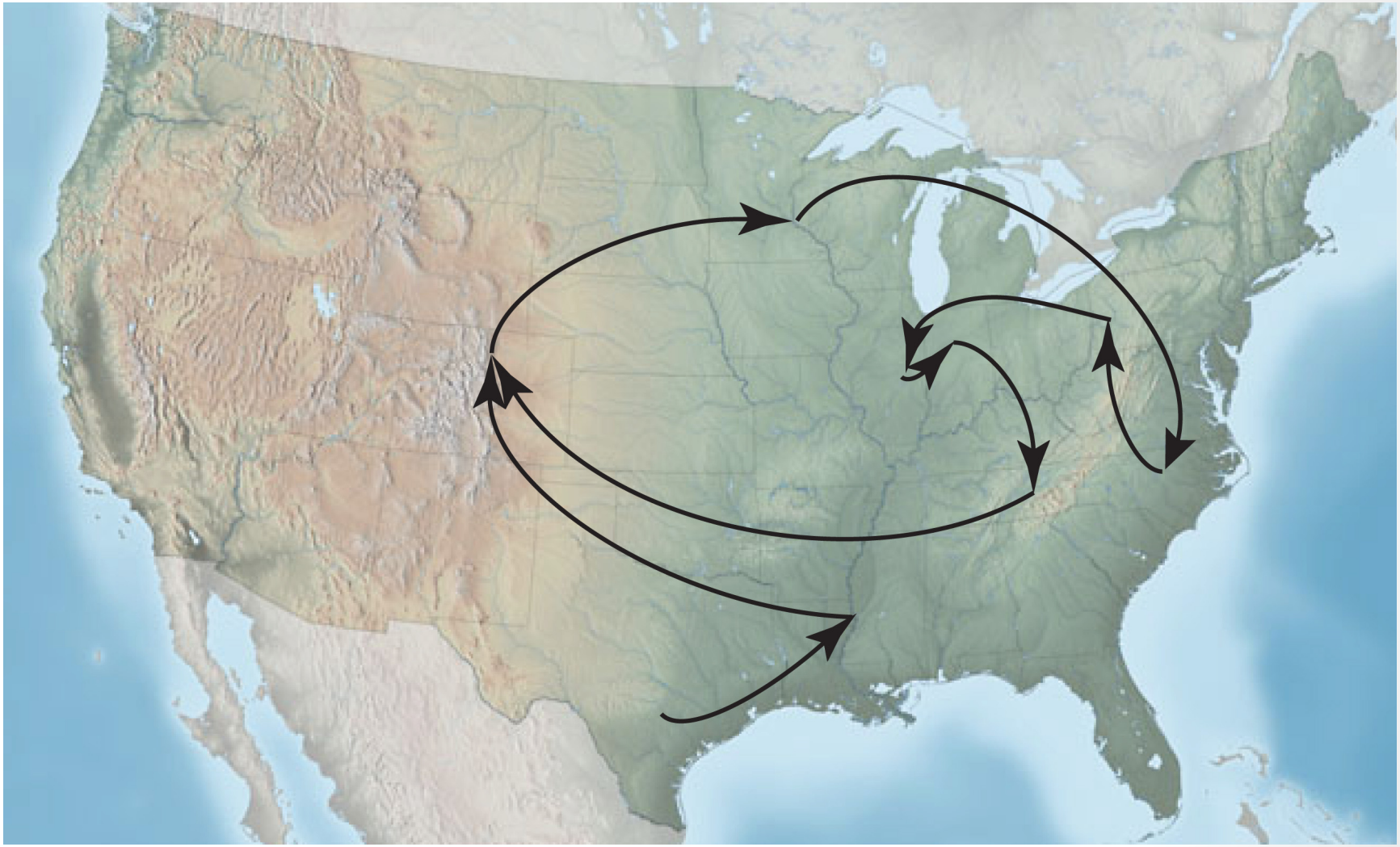


From Charlie Chaplin's movie Modern Times

Agenda

- Greetings, Goals, …
- Quick Look at HPC
- How to Use a Supercomputer
- Transferring Data
- Managing Source Code
- Python for Scientific Computing
- Visualization and Data Analysis
- Architecture, Performance, Parallel Computing
- What Else Should I Know?
- Survey
- Participant-Presenter Interactions

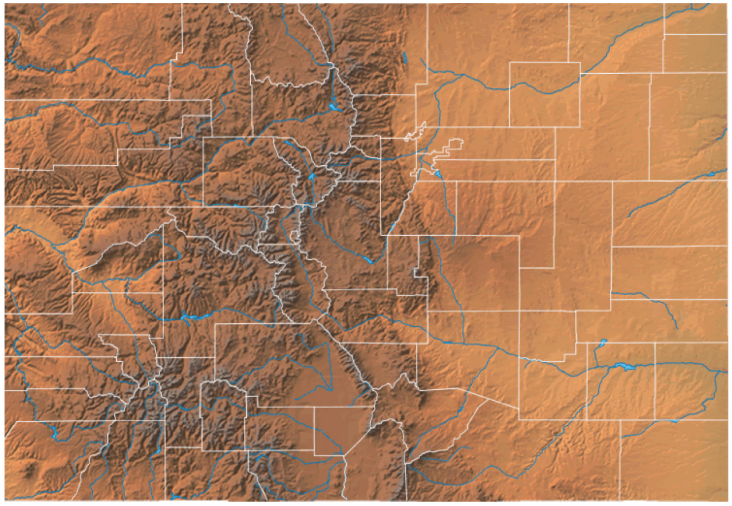
A Professional Odyssey



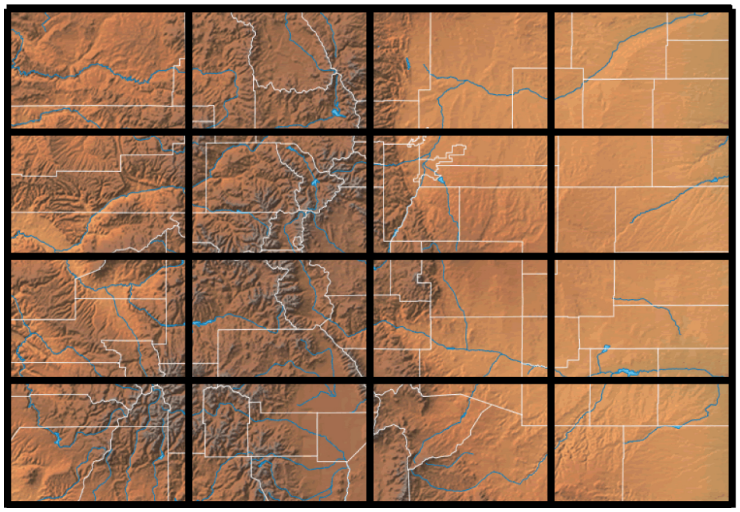
High Performance Computing

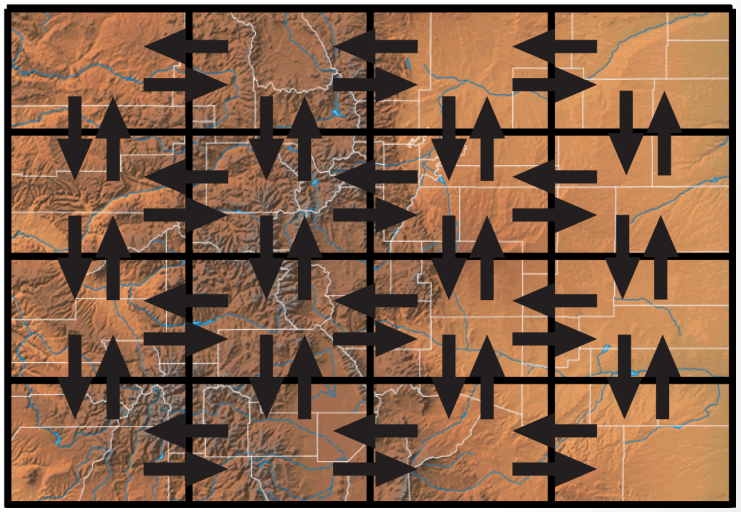
Time Was Invented To Keep Everything from Happening All at Once

— Anonymous

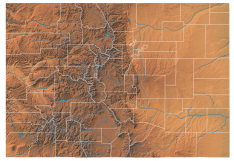


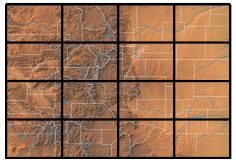
A Quick Look at Parallel Computing

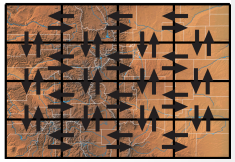




Why Do Parallel Computing?







- **Limits of single CPU computing**
  - Available memory, performance
- **Parallel computing allows**
  - Solve problems that do not fit on a single CPU
  - Solve problems that cannot be solved in a reasonable time
  - Larger problems, faster turn-around, more cases
- **Writing effective parallel application is difficult**
  - Communication can limit parallel efficiency
  - Serial time can dominate
  - Application load balance is important
  - Do the CPU requirements justify parallelization?
  - Will the code be used enough times?
- **Supercomputers are good at solving large coupled systems of PDE's**

HPC Use Models

- **Tightly-Coupled MPI**
- **Embarrassingly Parallel**
- **Work Flow**
- **Compute and Store (or Visualize) in Different Places**
- **Science Gateways**
- **Access to Databases, Instruments, …**

What Is Petascale ?

- **FLOPS – Floating Point Operations per Second**
  - Measures performance of a supercomputer
- **Megaflops** **1,000,000** **operations per second**
- **Gigaflops** **1,000,000,000**
- **Teraflops** **1,000,000,000,000**
- **Petaflops** **1,000,000,000,000,000**
- **Exaflops** **1,000,000,000,000,000,000**

- **… peta, exa, zetta, yotta,**
  xona, weka, vunda, uda, treda,
- **googol, … googolplex**

Top500 for June 2014

www.top500.org

Rank	System	Vendor	Cores	Rmax (PFlops)	Rpeak (PFlops)	Power (mW)
1	Tianhe-2 (MilkyWay-2) National University of Defense Technology- China	Intel Xeon and Xeon Phi	3,120,000	33.8	54.9	17.8
2	Titan ORNL	Cray XK7 NVIDIA K20x	560,640	17.6	27.1	8.2
3	Sequoia LLNL	IBM BlueGene/Q	1,572,864	16.3	20.1	7.9
4	K Computer Riken - Japan	Fujitsu	705,024	10.5	11.3	12.7
5	Mira Argonne National Lab	IBM BlueGene/Q	786,432	8.2	10.1	3.9
7	Stampede U of Texas	Dell Xeon Intel Xeon Phi	462,462	5.2	8.5	4.5

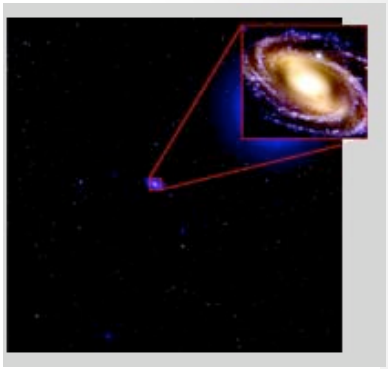
Some Research Areas that Use HPC

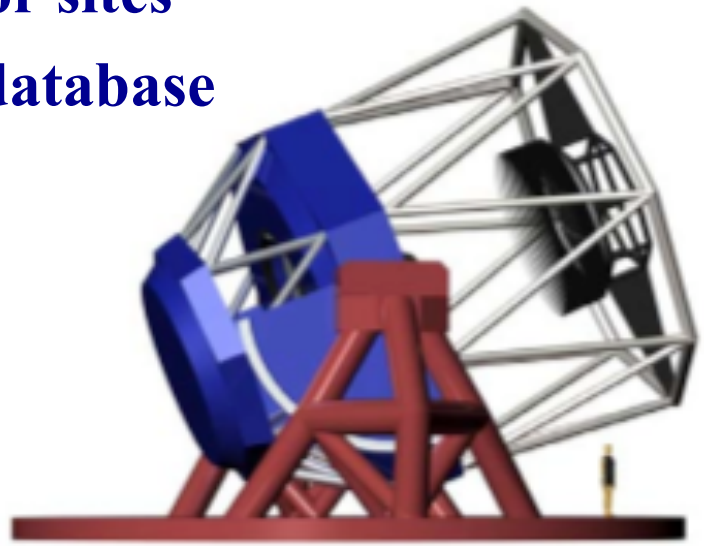
- Climate Modeling
- Weather Forecasting
- Storm Modeling
- Computational Nanotechnology
- Astrophysics
- Cosmology
- Numerical Relativity
- High Energy Physics
- Quantum Chromodynamics
- Condensed Matter Physics
- Molecular Dynamics
- Proteomics
- BioInformatics
- Ion Channel Simulations
- Virus Structure
- Biomedical Informatics
- Biomedical Engineering
- Drug Design
- Geophysics
- Seismic Modeling
- Oil Reservoir Simulations
- Earthquake Engineering
- Groundwater Modeling
- Nuclear Engineering
- Computational Fluid Dynamics
- Numerical Wind Tunnel – Aircraft Design
- Aeronautical Engineering
- Computational Chemistry
- Automotive Crash Testing

LSST – Large Synoptic Survey Telescope

www.lsst.org

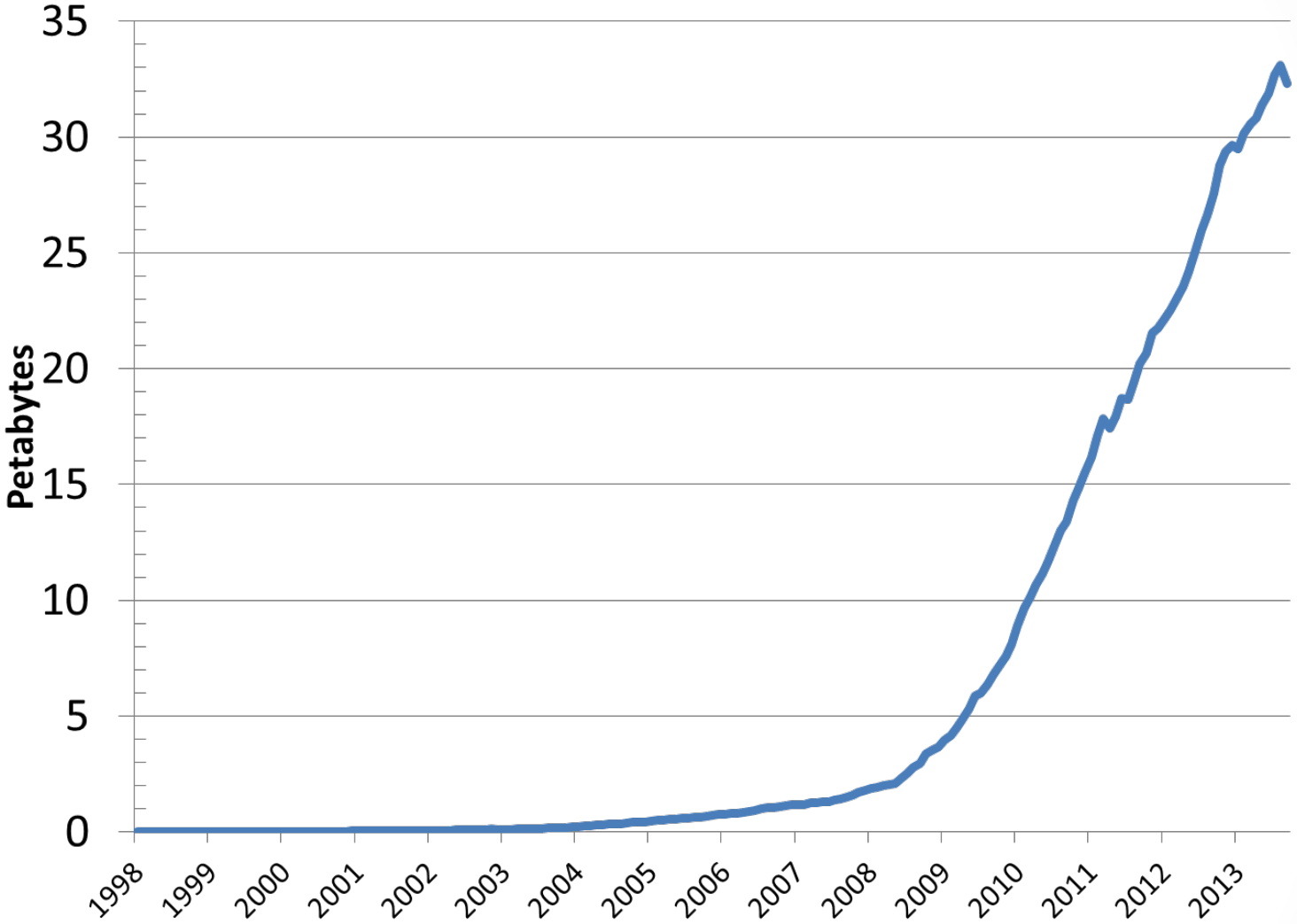
- Observe data on Chilean mountain top
- 3200 megapixel camera - 3 Tbytes per night
- Photograph entire sky every few nights
- Down the mountain to base station
- Across network to NCSA and mirror sites
- Validate data and populate public database
- Index and manage results
- Enable multiple pipelines
- LSST has a lot of moving parts





Pictures are from the LSST website

Managing Exponential Growth in Storage

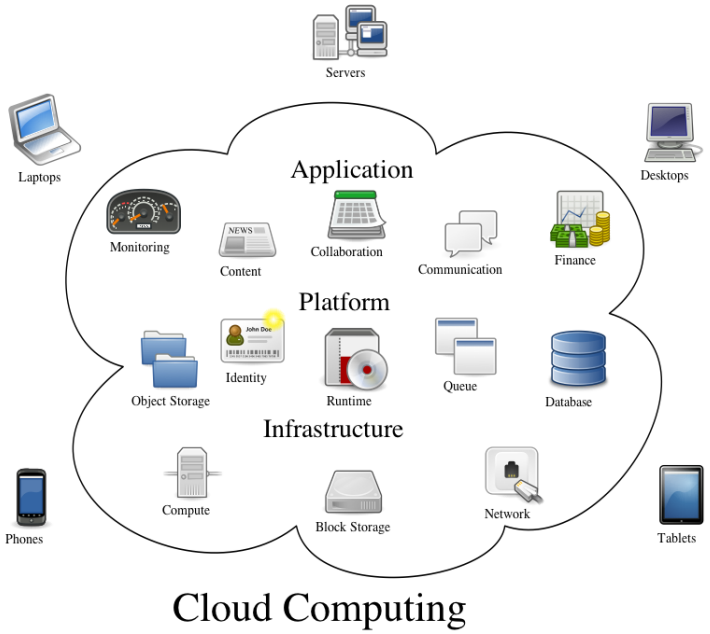


HPSS at ORNL is managing 30+ PB and growing at more than 30 TB per day.

New, Interesting, (Maybe) Important CyberInfrastructure Technologies

- **Science Gateways**
- **Distributed Computing**
  - Clouds
  - Grids
  - Condor
- **Accelerators**
  - GPGPU
  - Xeon Phi
  - FPGA

Cloud Computing



- **What Is the Cloud?**
  - Buncha computers hooked to a network
  - What's the big deal?
  - Larry Ellison Rant:
    - We've defined cloud computing to include everything we currently do
    - The computer industry is the only industry that is more fashion-driven than women's fashion
  - Computing resources (hardware and software) delivered as a service over a network – software-as-a-service
  - Users' data are stored on servers at a remote location
  - Business can reallocate IT operations costs away from hardware/software spending and personnel expenses
  - 235,000,000 Google entries for *cloud computing*
  - Does it make sense for scientific computing?

Challenges / Opportunities at Campus HPC Centers

- **Staffing — Work-force development**
- **Adapting scientific codes to use new technologies**
- **Storing, moving, analyzing, visualizing mass quantities of data**
- **Space, power, and cooling**
- **Developing research partnerships in academia, government, for-profits, non-profits**

What Is XSEDE ??

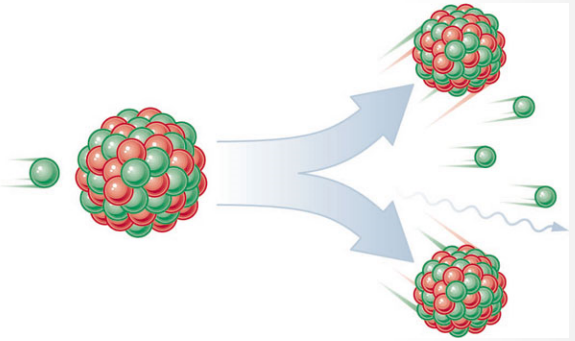
- ***www.xsede.org***
- **Extreme Science and Engineering Discovery Environment**
  - (No one *ever* remembers this name)
- **Funded by the National Science Foundation**
  - ~$120M over 5 years
  - Not for hardware
  - Renewal anticipated
- **Rah-Rah:** *Most powerful integrated advanced digital resources and services in the world*
- **Makes computational resources available to researchers in academia, government, non-profits, for-profits**
  - Includes: computing, visualization, storage, data collections, software and tools, advanced support services, …

XSEDE Computing Resources

Stampede	Texas Advanced Computing Center	Large Dell cluster with Xeon Phi – capability computing
Keeneland	NICS and GaTech	GPU's
Lonestar	TACC	Visualization and GPU programming
Gordon	San Diego Supercomputing Center	Data intensive applications
Trestles	SDSC	High-throughput, science gateways, small core counts
Blacklight	Pittsburgh Supercomputing Center	SGI – large shared memory
	Open Science Grid	High-throughput computing
Quarry	Indiana U	Support for web services
	Storage and Extended Support	
Jan 2015	Systems at TACC and SDSC	

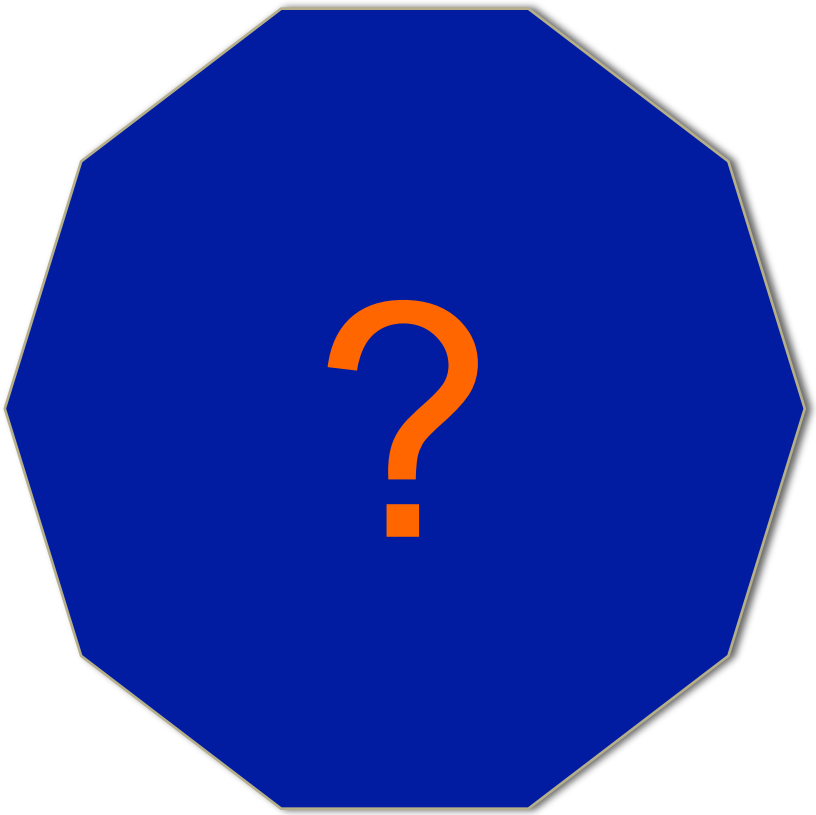
Questions?

What does a nuclear physicist eat for lunch?



Fission Chips!

From a presentation by Hai Ai Nam of Oak Ridge National Lab



bruce.loftis@colorado.edu