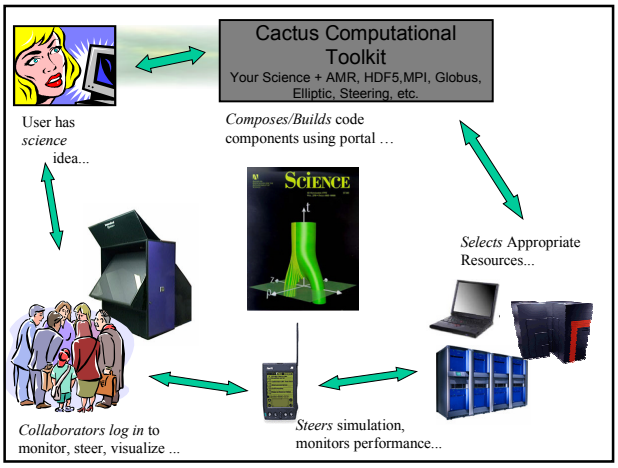
Cactus Project & Collaborative Working

Gabrielle Allen

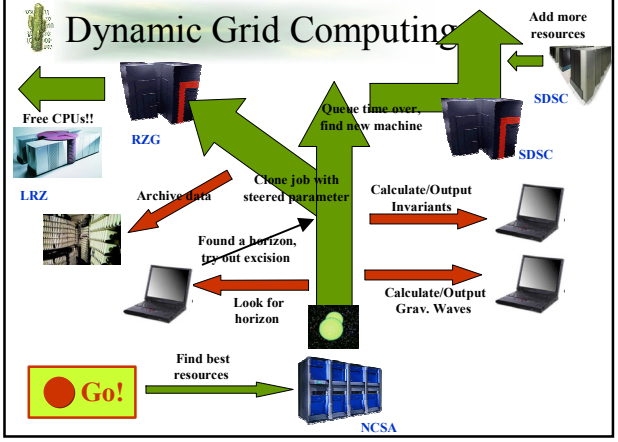
Max Planck Institute for Gravitational Physics,
(Albert Einstein Institute)

Cactus

- Modular, collaborative, parallel, portable framework for large scale scientific computing: http://www.cactuscode.org
- Application/Infrastructure Thorns (modules in Fortran/C/C++ etc) plug into core Flesh (glue in ANSI C)
- Driven by needs of Numerical Relativity community (very large scale simulations, really need (right now) resources/flexibility promised by grid computing), large distributed collaborations, lots and lots of data, visualization crucial for physics understanding, …
- Can do now: simulation monitoring, steering (web server thorn), streaming data for visualization (HDF5), readers for many different viz clients, developing simulation portal (ASC), distributed simulations, focus on making it transparent for users (VizLauncher: automatic visualization from web browser interface).



Dynamic Grid Computing



Collaborative Scenarios

- Collaboration consists of geographically distributed researchers with access to different sets of resources (supercomputers, Idesks, Caves, Visualization software, highspeed networks, …).
- Limited number of high resolution simulations taking place at unknown times and places (queuing systems).
- Important to use this run-time effectively … need to monitor that everything is running properly (steer output dir, switch of expensive analysis, …) … and learn as much physics as possible (steer analysis, output variables, physics parameters, …).
- Everyone in collaboration needs to be able to interact (in different ways) with the simulation.
- Any everyone also needs to be able to interact together to see what other people are seeing (Access Grid, Collaborative Visualization).

Requirements I

- Remote Visualization
  - Visualize data from simulation in realtime using best available tools on local machine
  - Data streamed directly from simulation across network, or accessed from various local filesystems
  - Downsampling, zooming, …
  - Shouldn't slow down simulation (separate data server needed)
  - Each user should be able to customize what they are seeing (variables, downsampling parameters etc) …
  - … or see exactly what someone else is seeing!
  - Need viz on all platforms (laptop, PDA, phone, Windows, Mac, …) and integrated in same way.

Requirements II

- Event Description and Transportation
  - E.g. Remote Steering requests
  - Protocols/APIs
  - Focused on grid aware visualization systems, including distributed collaborative environments. (CaveLib getting close, but not flexible enough)
- Data Description
  - Remote, distributed in different ways across different machines/file systems/archives, scientific data (multidimensional-arrays, different geometrical objects)
  - With visualization in mind (clients should be able to extract enough info to recognize what the data is and what to do with it).
  - Along the lines of OpenDX data model (general RDF/XML not enough)

Requirements III

- Security and Security Policies
  - Who can interact in which way with the simulation (monitor, access data, steer [physics and/or analysis,output data]).
  - How to implement this? Hierarchies, associations …
- Information
  - What simulations are running now, which ones are already queued (want to be able to check/amend parameters for these), what is estimate for when queued jobs will be running.
- Notification
  - Jobs running/finished, significant events (disk space nearly full, event horizon formed).
  - Email, SMS, Fax, …

Requirements IV

- Be able to reproduce simulations
  - Need detailed log of all events to know what happened.
  - Scripting language for Cactus to be able to do it again.
- Collaboration
  - Project portal for shared simulation configurations (thorn lists, machine configuration files, parameter files)
  - Access to data/configurations for past runs
  - Interactions across Access Grid etc.
- User Interfaces
  - Need to be intuitive, easy-to-use, but also flexible.
  - Suitable for users, developers, physicists, mathematicians, computer scientists, …

Requirements V

- System
  - At the moment, everyone in the collaboration has personal access to each different resource (disk space, inodes, number of jobs allowed on queues, everyone has to get local environment in sync …).
  - More useful if these resources were accessed on a group/collaborative basis.

GridLab Project (EU in Negotiation)

AEI, Poznan, ZIB, Lecce, Cardiff, VU, Brno, Sztaki, Sun, Compaq, ISI, Argonne, Wisconsin, Athens