Making *Grid* Computing Work for LIGO

- *grid* means distributed heterogeneous resources “*transparently* accessible” to scientific users.
- “*transparently accessible*” implies significant coordination of management of *resources*.
- *resources* include
  - hardware
  - software
  - network
  - data
  - personnel who support all this.

*…well, reasonably or almost or somewhat, transparently*
3 Separable General Issues

• Data reduction and analysis “usage model”
  – see White Paper and further map to grid

• Identify coordinated (grid) portion of “resources”
  – hardware, software, network, people
  – explicitly identify
    • what is to be transparently available to all users
    • what is to be held back for dedicated or local use
  – early drafts of MOUs address this aspect

• Management of coordinated (grid) resources
Management of Coordinated Resources

• Tautology: management = coordination
• Our most difficult and critical task is to turn an intrinsically uncoordinated bag of resources into a “transparently” functioning “grid”
• Intrinsic problem we must overcome: we are dispersed
  – Geographically dispersed
  – Institutionally dispersed
    • People are not (never will be) in a single, clear chain of command
  – Funding sources, timing, and motivations dispersed
  – System design and configuration planning dispersed
• Nonetheless we must handle this as ONE entity
  – whatever you want to call it, it is WE, not US and THEM
Support: Demand Side Resources: hardware, software, data, network

- Development (or procurement)
  - Specify configuration
  - Installation and testing
- Upgrades and version management
  - Ports to different HW or SW platforms & configurations
  - Local installs
- Repairs and bug response
- Help desk
- Data distribution
- Resource allocation
  - Access control
Hardware Usage

• Production computing
  – Simulation
  – Real time (e.g. stochastic)
  – Compute bound (e.g. cw)

• User service
  – Code development and rapid prototyping
    • Detector characterization
  – Fast turn around analysis of reduced data sets

• Storage
  – Large scale community data (Tier 1 and ??)
  – Reduced community data
  – Private data
Software categories

- OS
- “Grid” tools
- Infrastructure (LDAS)
- Science tools
- Production packages specific to WGs
  - Simulation
  - Data reduction
- Individual data analysis
Support: Supply Side Resources

• Personnel
  – LIGO Lab
  – Tier 2 Institutions
  – Contract/out source

• Data Distribution
  – Wide Area Network
    • At Tier 1
    • At Tier 2s
  – Hi Latency (FedEx)

• Archives
  • HPSS
  • Tape Copy Facility
Our task: manage resources

• Look at where we are …
  … define where we want to get
  – with a transition plan from here to there
• In terms of
  – Usage model
  – Allocation of existing supply side resources (people, net)
  – Optimizing existing systems & software configurations
    • Short term issues -- make things work, e.g. HPSS
  – Pleas (justification) for additional resources
  – Plans to allocate additional resources
    • Fall back plans to optimize with less than needed/desired resources
  – Establishment of a chain of command coordination
  – Rapid prototyping of management plan
    • Short term issues -- how to analyze E7, E8, …