"Technology" as seen by a "Modeler"

Bruce Allen

Max Planck Institute for Gravitational Physics

(Albert Einstein Institute, or AEI)

Hannover, Germany

GW Models

MATO/Virtualization + Clouds

- Commercial clouds still not so cost effective

- Can't manage IO and large data volumes

- Virtualization: now efficient and reliable enough to use in the real world

- A possible 'standard method' for grid jobs. A job is a virtual image CONTAINING the data (might work for GW analysis, other fields?). A grid resource is nothing other than a hypervisor box, waiting for a virtual image

- Great way to simplify the nightmare of software environments and distribution

- BUT (quiet private remark from a friend) "does it REALLY work or is this just the usual pretty picture from the technology advocates". Counterargument: there appear to be lots of working examples of this.

Grancher/Database Technologies

- Databases already approaching the 100 TB level

- Will there eventually be a convergence of Database and Filesystem technologies? Both are ways of storing information and "metadata". Both need to be distributed to meet current needs. Both have 'transaction' and 'atomiticity' and 'synchronization' demands.

- Incorporating SSD's into Databases can result in significant performance boosts. (NB: also filesystems!)

- Note to self: look into NoSQL and more "lightweight" systems.

- CERN is now in 'exponential data mode'

- No way to avoid disk/tape hierarchy. (NB: SSD/disk/tape)

- We need decent file systems (just look at the complexity of some of these storage diagrams, aaarrgggghhhh!!)

- At some point one will need to tear down the legacy structure and start over (Swiss-army knife!)

- Could we do away with Posix (locking in particular). What the hell IS Posix anyway??

- Storage systems need much better instrumentation and monitoring. (Currently Strace/Ptrace, Nagios, etc)

- A decent file system need to include real authentication and authorization infrastructure. UGO+-RWX ain't enough anymore, Dorothy.

- At some point, a ground up redesign of large scale protocols is probably needed for large data use.

- Looking at issues of scalability, namespace management, performance I can't get the filesystem versus storage system versus database convergence question out of my mind.

- HSM systems can't replace "user-level" knowledge for file set management. Better tools and protocols needed here too.

Hardware??

- This ain't technology, it's "software"!!

- What about GPUs? SSDs? Integrated GPU/CPUs? OpenCL?

- How about the evolution of the Linux kernel? Lot of low-cost processors? Are 3.5" disk drives dead?

- How about new networking technologies? Hands off management tricks? Latest ways of building clusters?

- How about power generation and cooling? Green data centers?

- How about volunteer distributed computing, for example as a distributed storage pool.

Technology from Model Perspective