

GPCF* Update

- Present status as a series of questions / answers related to decisions made / yet to be made

* General Physics Computing Facility (GPCF) is not a memorable name. Suggestions for a better name and TLA are welcome!

What needs are we addressing?

- Common solution for a varied community
 - Intensity and Cosmic Frontier experiments
 - Some of the old fnalu functions
- Shared resources
 - To optimize utilization
- Focus on long term management and operation
 - Reduce the burden on the experiments / users
- Reduction of “one off” solutions and orphans
 - Reduce the burden on the CD

What are we not addressing (yet)?

- Data management schemes
 - And implications on processing and data access patterns
- Performance
 - Learn from experience
 - Build in flexibility

Thinking started, but a “plan” needed

Guiding principles

- Use virtualization
- Training ground and gateway to the Grid
- No undue complexity – user and admin friendly
- Model after the CMS LPC where sensible
- Expect to support / partition the GPCF for multiple user groups

Basic architecture

- Interactive facility
 - VMs dedicated to user groups
 - Access to common, group, and private storage
- Local batch facility
 - VMs dedicated to user groups
 - Logins possible
 - Otherwise close to or same as grid environment
- Server / Service Nodes
 - VM homes for group-specific or system services
- Storage
 - BlueArc, dCache, or otherwise (Lustre, HDFS?)
- Network infrastructure
 - Work with LAN to make sure adequate resource

VMs

- Q: Which VMs are allowed?

A: Supported (baselined) SLF versions. Customized for user groups. Patches will be applied to VM store and active VMs.

- Q: Resources per VM?

A: 2 GB memory per core
x GB local disk storage
n guaranteed / n shared processors
x guaranteed / x shared network bandwidth
Where oversubscription is allowed.

VMs (#2)

- Q: Which hypervisor?

A: Xen (for now)

- Q: How are VMs provisioned and deployed?

A: Will be guided by FermiCloud work, but currently use manual provisioning of static VMs

- Q: How are the VMs stored?

A: Will be guided by FermiCloud work, but currently envision BlueArc

→ These choices do not impact user environment

Storage Systems

- Q: Which storage / file systems will be used?
A: This is the principal remaining question for the hardware architecture. We expect to start with use of BlueArc and public dCache, operated in a manner largely unchanged.
Storage system capacity is reasonably well specified, but performance as a function of usage is not.

Storage systems (cont'd)

- Q: What about Hadoop or Lustre or ...?
A: It's too early to think about these for production systems in a "new" facility. We want to study these within the FermiCloud facility, and perhaps introduce limited capacity within the GPCF facility.
- Q: What are the implications of delaying a decision on storage?
A: This affects specifics of hardware purchase. Distributed storage systems might want many nodes with associated disks, possibly with dedicated (FC or Infiniband) network. For now we will assume separated storage systems.

Security

- Q: Are there special security needs?

A: All of GPCF will be within the General Computing Enclave (GCE), meaning they are treated like any other local cluster.

- Only Fermilab Kerberos credentials
- No grid cert access
 - Except maybe Fermi KCA certs???

Network Topology

- Q: How are VMs named / addressed?

A: Current plan is:

- Fixed IPs for interactive VMs
- Dynamic IPs for batch VMs
- Fixed IPs for server VMs
- Fixed IPs for network storage

Resource Provisioning

- Q: How many VMs/nodes/servers/...?
A: Using NuComp / Lee's numbers for IF needs.
Budget request is for 2x – though may not see this
- Q: How are resources to be distributed among groups?
A: TBD. To some level, based on contributions to purchases.

User Accounts

- Q: How are groups “segregated”?
A: NIS domain per group. Any VM associated with one NIS domain. Privileged access restricted to admins.

VMs (#3)

- Q: What “fancy features” are envisioned?
A: None for now... Possibilities for the future are:
 - High availability (HA) for services
 - VM failover / relocation
 - VM suspension / restart

Physical Location

- Q: Where are the physical nodes?

A: There are building power constraints. FCC is the “high availability” center, but “no room at the inn”. May consider only storage in FCC, nodes in GCC.

FY10 Budget request

- Overlap with BlueArc, dCache requests to be resolved

Qty	Description	Unit Cost	Extended Cost	Fund Type
16	Interactive Nodes	\$3,300	\$52,800	EQ
32	Local Batch Nodes	\$3,100	\$99,200	EQ
4	Application Servers	\$3,900	\$15,600	EQ
3	Disk Storage	\$22,000	\$66,000	EQ
1	Storage Network	\$10,000	\$10,000	EQ
1	Network Infrastructure	\$40,000	\$40,000	EQ
1	Racks, PDUs, etc	\$3,000	\$3,000	EQ

Schedule

- 2 phases:
 - ASAP: put out requisitions for:
 - BlueArc disk
 - Additional dCache disk
 - $\sim 1/4$ total number of nodes
 - Spring, or as needed:
 - Remaining number of nodes