Build a Grid-Aware, Cost effective, 400 Core Tier 3 Center – A “How To”

2012 J. Phys.: Conf. Ser. 396 042031

(http://iopscience.iop.org/1742-6596/396/4/042031)

Build a Grid-Aware, Cost effective, 400 Core Tier 3 Center – A "How To"

Kurtis F. Johnson

Department of Physics
Florida State University
Tallahassee, FL, USA

E-mail: kurtis [at] hep.fsu.edu

Abstract. The availability of commodity digital hardware: fast, cheap and standardized, makes building a high performance computing center possible for even small groups. Delays and overruns - some foreseeable, some not, will occur even when anticipated but still more if not. In the following we detail explicit procedures, tradeoffs and advice to create a several hundred core, Grid-aware Tier 3 facility such that a small research group can enjoy maximum utility of its research resources.

1. Introduction: Small Group HPC

The advantages to a research group of having its own high performance computing facility are clear: a group may store large datasets locally making access fast and efficient, tailor the amount of CPU power vs. storage and networking, and set policy to favor the group's computing needs. Control over policy allows a group to dedicate resources towards development, data analysis or throughput, whatever is most critical at the moment.

The major ownership cost will be that someone has to keep the facility running. This will be the largest resource sink associated with a facility. By far most of the effort will not be maintaining the hardware, but maintaining the system and application software. If the facility must also be Grid-aware, expect to expend a large amount of effort debugging the Grid-to-facility communication interface and updating the Grid software [1].

We recommend items needing attention and pitfalls to avoid for a group wishing to have local high performance computing. Much pain, delay and expense can be avoided by comprehensive preparation and anticipation of problems.

Following is a case study of a tier 3 center built from commodity parts.

2. Case Study – FSU Tier 3 Center

The Florida State University High Energy Physics group received ~K$70 for a tier 3 facility in support of its activities at an ongoing Large Hadron Collider experiment [2]. It desired to obtain as much performance per dollar as possible by assembling the machines locally. By "performance" is meant primarily computing cycles and storage. At the same time, keeping the sysadmin load, and attendant expense, as low as practical was an important design goal. These were the major design factors.

Although the hardware selected then would not now be optimal, not to say obsolete, the components are listed for completeness. The Intel Core i7 9xx series CPU [3], ASUS motherboard (MOBO), 12GB G.Skill RAM and 2TB 5400RPM Samsung HD were selected. The cost in mid-2010 totaled just over $850 (MOBO: $220, CPU: $285, RAM: $200, HD: $90); cables; "T"'s (see Figure 2) and power supplies make up the remainder. One year later the cost had dropped by over $100 while speed had increased and today would be much cheaper. The time to assemble a machine is about 1/2hr per "T" and can be done with unskilled labor. In this manner, a cluster of ~400 execution cores and ~100TB of RAID6 storage was built within the funding constraints.

2.1. Jobmix

The major determinant of the structure of a facility is the expected jobmix. This tier 3 was not, for example, intended to service http requests, but to be able to process several hundred analysis jobs in parallel. The users expected the majority of jobs to be compute intensive, in which a machine ponders a few megabytes of input for several minutes, rather than I/O intensive in which many megabytes/sec of input is skimmed for a few relevant data elements. Thus, the emphasis was on installing a large number (~50) of fast (3.6GHz) worker nodes (WNs) rather than maximizing the data transfer rate between WNs and storage, which in any case was capped by the 1Gb wire speed of the CAT6 cable and LAN cards.

The most straightforward organization of machines for this task is to have many WN machines, one or more storage servers to handle software and datasets, and a job submission machine normally designated the “head node” through which users can submit their job requests to the WN’s, all connected within a common, private subnet. The head node connects both to the subnet and to an exterior net from which users may log into it. Placing the WN’s into an exclusive subnet with non-routable IP addresses, rather than connecting them to a higher level net allows for easier maintenance, and they do not have a visible interface to the internet, a plus for security.

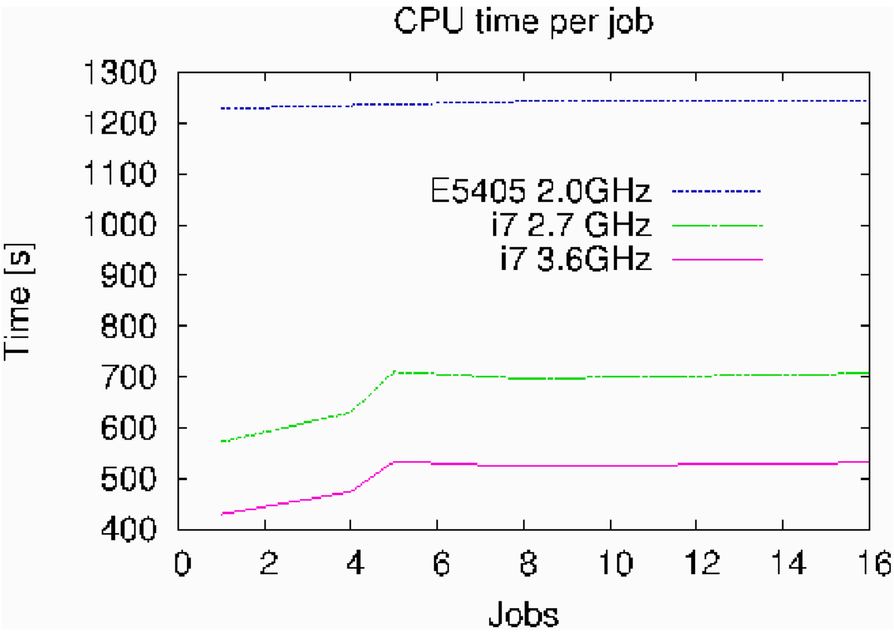

Figure 1: Example of a compute intensive benchmark of a Core i7 prototype design vs. a commercially available computer. Note suppressed zero.

2.2. CPU

As mentioned, speed of execution rather than throughput was to be emphasized. In deciding upon a CPU, code benchmarks on different machines were carried out. Figure 1 is an example of one such series of tests. The figure shows the execution time per job in seconds as a function of the number of simultaneous jobs submitted. The line at the top of the plot is indicates the execution time for a CPU intensive benchmark running on an Intel XEON based machine offered by a large commercial vendor sold at about twice the price of a locally-assembled Intel Core i7 based machine. Both machines have 4 physical core processors, but the Core i7 has 8 "hyperthreaded" cores. The assembled machines were configured so that the clock rate can be increased at will, shown by the two lower lines, as can be seen in figure 1. For these machines, the performance/cost ratio for CPU intensive jobs is greater than 4 for the Core i7 920 at its preset clock of 2.7GHz, and larger still when the clock is increased.

2.3. RAM Footprint

The expected jobmix also suggests how large the RAM footprint should be; the FSU T3 utilizes 12GB RAM/machine (3GB per physical core), as the largest jobs are expected to use a maximum of 4GB and the smallest just a few MB. It is also expected that a single machine would rarely have 8 large jobs. If this had proven in practice to be inadequate, an additional 12GB could be added to the MOBO's without incurring intermittent stability problems. It should be mentioned here that some CPU/MOBO/RAM combinations which are stable at smaller RAM loads become unstable when RAM is added. It is imperative to check this using a prototype before committing to a design.

2.4. Local Storage, UPS, LAN Switch

Other choices to be made include: the amount of WN local storage, whether to use UPS backups and what kind of LAN switch/router to install.

For local storage on each WN, 2TB hard drives were chosen. Each WN has a 500GB partition of its 2TB available as temporary storage for jobs in execution; the rest is configured as a separate partition for a planned future Hadoop expansion. Typical file sizes used in the CMS experiment are smaller than 5GB, thus 500GB was deemed to be adequate (and there have been no job failures due to low /tmp space). An alternative which was considered was a "solid state disk" (SSD), ~$200 for 120GB at the time, which when new, have latencies 1000 times smaller than a magnetic disk. However, reducing disk latencies even down to zero does not noticeably impact job completion times for a typical analysis job where data transfer to and from the job is a small fraction of the execution time, whereas using the newest technology would introduce the maintenance uncertainties due to a smaller experience base. In line with the guideline of keeping human sysadmin effort to a minimum, disks were selected.

UPS backups are expensive, introduce new failure modes into a system and have a limited, typically ~3 year lifetime. Nevertheless, if a facility does not have reliable power, UPS may be imperative. Power failures do not always occur as a clean drop to 0 Volts, but often appear as multiple cycles of voltage slumps lasting several hundred milliseconds followed by reactive spikes as large power grid switches unlatch, or voltage surges when re-routing is successful. Multiple slumps, when descending below about 90V, the typical low-end of compliance of a PC power supply, result in rotating media losing synchronization and data being incorrectly written, bias voltages on motherboards to drift and BIOS's to fail, in short: many man hours of work until recovery. One should obtain data on power service reliability and act accordingly.

Selecting a switch/hub/router for a subnet goes too far afield to treat in detail here. Note that switches come in a large range of prices and with an enormous range of capabilities, most of which a small facility will never need. One must decide whether, for example, Quality-of-Service capability is really needed and be cognizant that someone will have to keep it running. The primary purpose of a switch is to transfer data. Measure the data transfer rates needed by running the target application software. Understand that, without a lot of effort, the actually achieved data transfer rate will seldom exceed 30% of wirespeed.

Journal of Physics: Conference Series **396** (2012) 042031 doi:10.1088/1742-6596/396/4/042031

2.5. The Caseless Computer
The effort to extract as much performance per dollar as possible required jettisoning of any unnecessary parts of a computer, starting with the case and continuing through half the power supply. The purpose of a computer case is to protect the machine while being shipped or moved and, in the case of rack mounts, to provide an easy way to organize machines in vertical racks. It is not actually required for the machine to function. The FSU minimalist design can be seen in figure 2.

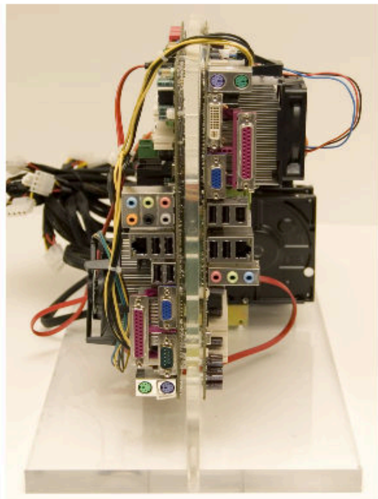

Figure 2: Prototype of a basic WN machine. The spine of the inverted "T" supports two computers which share a single power supply. There is no case. Of course, a WN has no need for a keyboard, monitor or mouse. The volume of each "T" comprising two computers is approximately that of a standard 2U by 27" format drawer.

2.6. Storage
For storage of large datasets, JABODs (Just A Bunch Of Disks) are not adequately reliable – even modern disks have a non-negligible probability of failure resulting in loss of the entire disk, so RAID units or something similar are obligatory for data storage. The cost was a little less than $100/TB for a 44TB unit. The assembly procedure is described in the next section.

2.7. Advantages of local assembly
Assembling machines can be much less expensive than purchasing from a vendor, though occasionally major vendors offer excellent value. But going the "do-it-yourself" route allows a group to select reliable components and a specific CPU/RAM/HD/MOBO combination which can be tailored to a desired jobmix and which can result in high reliability, while the package offered by a major vendor may not meet the desired jobmix or storage requirements. In addition, vendor supplied hardware will normally be declared out-of-warranty if, for example, the designers tinker with the machine clock to increase performance.

The FSU T3 RAID6 storage servers were built along the same pattern as the WN's – but with just one machine per "T", with 24 2TB disks affixed to the baseplate and a commercial RAID PCI card (~$1100) as the controller. The cost averaged to ~$100/TB.

2.8. Purchasing

Purchasing should be completely routine in this modern age, but two points must be made. The first is well known: one should purchase as late as possible. Prices of components can drop so rapidly that a few months' procrastination in procurement can buy 10% or more additional hardware. The second point is less obvious – large organizations often have "purchasing departments" (PD), which control how funds are spent. A PD may have strict rules about spending, such as lists of allowed and forbidden vendors. It can happen, and has happened to the author, that required hardware components may not be purchased because the PD-approved vendors do not offer them. One obstinate bureaucrat can incur months of bitter delay. Designers should be sure to communicate with the PD often and in detail before hardware selection is finalized.

2.9. OS/Cluster management installation

The FSU T3 uses Scientific Linux 5. First a prototype WN machine was loaded with a bootable SL5. After it was found to be satisfactory copies for all the WN's were made using a commercial disk copy program [4]. In whatever manner the duplicate disks are created, be sure that the duplicates are tested for functionality.

The FSU T3 uses the SGE cluster management and job submission software, which is freely available and of high quality. There are several others to choose from [5]. Local requirements will determine which to use.

Facility security is a constant source of uncertainty. If a facility is compromised, rather expensive manpower may have to be retained to "clean up" the intrusion and re-certify the system. Exploits abound, merely keeping track of them is a major effort. The FSU T3 attempts to reduce its target profile by exposing only those machines which must be open to the internet, running only required applications and running an active security oversight system. This is yet another manpower drain (typically for the sysadmin).

Finally, Grid software updates are released frequently [6]. Sometimes an update will break some aspect of an otherwise working system, requiring human intervention and increasing the sysadmin load. A facility should estimate carefully the sysadmin requirement and then plan for much more than estimated.

3. Conclusions

Inexpensive digital hardware has altered the computing landscape for small scientific groups. What a few years ago would have been called a high performance computing center is now within the reach of even small group budgets. The preceding case study of a low cost but powerful tier 3 facility is a proof-in-practice. We summarize this study of a tier 3 facility built from commodity parts by a series of condensed conclusions, crisply formulated so as to strip away all but the essential content.

1) **Check and benchmark everything on prototypes**:
   execution speed, data transfer speed, stability, possible future variants.
2) **Don't go for cutting edge**:
   much more expensive for small increase in system performance.
   incompatibility/failure rates higher at cutting edge means much effort
   expended on diagnosing/fixing driver incompatibilities.
3) **Buy late**.
4) **Plan for more sysadmin effort than any reasonable expectation**.
5) "**Keep your sysadmin close, keep your purchasing department closer**."
   (Sun Tszu, slightly modified.)

4. Acknowledgements
The author would like to thank Maurizio Bertoldi, Jeff MacDonald, Blake Sharin and Joe Ryan (all FSU) and Douglas Johnson (U. Colorado, Boulder) for much useful aid, support and advice.

5. References
[1] https://www.opensciencegrid.org
[2] LHC, see: http://public.web.cern.ch/public/en/lhc/lhc-en.html
[3] Core i7, see: http://ark.intel.com/products/37150/Intel-Core-i7-950-Processor-%288M-Cache-3_06-GHz-4_80-GTs-Intel-QPI%29
[4] Acronis, see: http://www.acronis.com
[5] Available cluster management systems include:
Condor: http://research.cs.wisc.edu/condor/
PBS: http://www.pbsworks.com/
SGE: http://www.oracle.com/technetwork/oem/grid-engine-166852.html
LSF: http://www.platform.com/workload-management/high-performance-computing
[6] For example, between April 25 and May 22, OSG released versions 3.1.0, 3.1.1, 3.1.2 and 3.1.3 of its OSG software package. This does not include other package updates.
See: http://osggoc.blogspot.com/