FERMILAB-Conf-01/255-E

CERN-IT-2001-006

Large Scale Cluster Computing Workshop *

Fermilab, IL, May 22nd to 25th, 2001

Alan Silverman/CERN and Dane Skow/Fermilab

1.0 Introduction

Recent revolutions in computer hardware and software technologies have paved the way for the large-scale deployment of clusters of commodity computers to address problems heretofore the domain of tightly coupled SMP processors. Near term projects within High Energy Physics and other computing communities will deploy clusters of scale 1000s of processors and be used by 100s to 1000s of independent users. This will expand the reach in both dimensions by an order of magnitude from the current successful production facilities.

The goals of this workshop were :-

1. To determine what tools exist which can scale up to the cluster sizes foreseen for the next generation of HENP experiments (several thousand nodes) and by implication to identify areas where some investment of money or effort is likely to be needed.
2. To compare and record experiences gained with such tools.
3. To produce a practical guide to all stages of planning, installing, building and operating a large computing cluster in HENP.
4. To identify and connect groups with similar interest within HENP and the larger clustering community.

Computing experts with responsibility and/or experience of such large clusters were invited to the workshop. The clusters of interest were those equipping centres of the sizes of LHC[1] Tier 0 (thousands of nodes) or Tier 1 (at least 200-1000 nodes) as described in the MONARC project (http://monarc.web.cern.ch/MONARC/). The 60 attendees came not only from various HENP sites worldwide but also from other branches of science, including bio-medicine and various Grid projects as well as from industry.

The attendees shared freely their experiences and ideas and proceedings are being currently edited, prepared from material collected by the convenors and offered by the attendees. In addition, the convenors, again with the help of material offered by the attendees are in the process of producing a "Guide to Building and Operating a Large Cluster". This is intended to describe all phases in the life of a cluster and the tools used or planned to be used. This guide should then be publicised (made available on the web, presented at appropriate meetings and conferences) and regularly kept up to date as more experience is gained. It is planned to hold a similar workshop in 18-24 months to update the guide.

All the material for the workshop is available at the following web site:

http://conferences.fnal.gov/lccws/

In particular, we shall publish at this site various summaries including a full conference summary with links to relevant web sites, a summary paper to be presented to the CHEP conference in September and the eventual Guide to Building and Operating a Large Cluster referred to above.

2.0 Overview of the Challenge

2.1 Fermilab Computing

Matthias Kasemann, head of the Computing Division at Fermilab described Fermilab's current and near-term Scientific Programme, including Fermilab's participation in CERN's future LHC programme, notably in the CMS experiment. He described Fermilab's current and future computing needs for their Run II experiments, pointing out where clusters, or computing farms as they are sometimes popularly known, are used already. He noted that the overwhelming importance of data in HENP's current and future generations of experiments had prompted the interest in Data Grids. He posed some questions for the workshop to consider over the coming 3 days:

[1] LHC, the Large Hadron Collidor, is a project now in construction at CERN in Geneva. Its main characteristics in computing terms are described below, as are the meanings of Tier 0 and Tier 1.

* note: tops of pages cutoff; best original available.

- How much could HENP compute models be adjusted to make most efficient use of clusters?
- Where do clusters not make sense?
- What is the real total cost of ownership of clusters?
- Could we harness the unused CPU power of desktops?
- How to use clusters for high I/O applications?
- How to design clusters for high availability?

2.2 LHC Scale Computing

Wolfgang von Rueden, head of the Physics Data Processing group in CERN's Information Technology Division, presented the LHC Computing needs. He described CERN's role in the project, displayed the relative event sizes and data rates expected from Fermilab RUN II and LHC experiments and presented a table of their main characteristics, pointing out in particular the huge increases in data expected and consequently the huge increase in computing power which must be installed and operated. The other problem posed by modern HENP experiments is their geographical scope, with collaborators spread throughout the world requiring access to data and compute power. He noted that typical HENP computing is more appropriately characterised as High Throughput Computing as opposed to High Performance Computing.

The need to exploit national resources and also to reduce the dependence on links to CERN has produced a multi-layered model of computing (MONARC). This is based on a large central site to collect and store raw data (Tier 0, at CERN) and multi-tiers (for example National Computing Centres, Tier 1[2], down to individual user's desks, Tier 4) each with data extracts and/or data copies and performing different stages of physics analysis. He showed where Grid Computing will be applied. Lastly, he ended by expressing the hope that the workshop could provide answers to a number of topical problem questions, cluster scaling, making efficient use of resources, etc and some good ideas to make progress in the domain of the management of large clusters.

3.0 Panel Summaries

The largest part of the workshop was a series of interactive panel sessions, each seeded with questions and topics to discuss and each introduced by 1, 2 or 3 short talks. Full details of these can be found in the forthcoming summary and most of the overheads presented can be consulted on the workshop web site. The following sections are summaries of the panels.

3.1 Panel A1 - Cluster Design, Configuration Management

Some key tools[3] that were identified during this panel included --

- cfengine – apparently little used in this community (HENP) but popular elsewhere Does this point to a difference in needs? or just "preferences"
- SystemImager and LUIS (Linux Unique Init System)– neither does the full range of tasks for configuration management but there are hopes that the OSCAR project will merge the best features of both
- NFS – warnings about poor scaling in heavy use
- Chiba City tools – these do appear to scale up well

Variations in local environments make for difficulties in sharing tools, especially where the support staff may have their own philosophy of doing things. However, almost every site agreed that having remote console facilities and remote control of power was virtually essential in large configurations. On the other hand, modeling of a cluster only makes sense if the target application is well defined and its characteristics well known – this is rarely the case for the participants

3.2 Panel A2 - Installation, Upgrading, Testing

Across the sites there are a variety of tools and methods in use for installing cluster nodes, some commercial, many locally-developed. At KEK, Dolly+ is a mechanism that uses a ring topology to support scalable system builds -- it is still in its testing phase and has not been tested at scale. Rocks (http://rocks.npaci.edu) is a cluster installation/update/management system. It is based on standard RedHat 7.1, has been tested on clusters up to 96 nodes (< 30 minutes to install all 96 nodes), is currently used on more than 10 clusters and is freely available now. In addition, Compaq references Rocks as their preferred cluster integration for customers wanting a 100% freeware

---

[2] Examples of Tier 1 sites are Fermilab for the US part of CMS and BNL for the US part of ATLAS.

[3] This summary, indeed this workshop, did not attempt to describe in detail such tools. The reader is referred to the web references for a description of the tools and to conferences such as those sponsored by Usenix for where and how they are used.

Redhat tools and basic shell scripts. They must have something working by September of this year.

Many sites purchase hardware installation services but virtually all perform their own software installation. Some sites purchase pre-configured systems, others prefer to specify down to the chip and motherboard level.

Three sites gave examples of burn-in of new systems: FNAL performs these on site and BNL and NERSC requires the vendor to perform tests at the factory before shipment. VA Linux has a tool, CTCS, which tests CPU functions under heavy load; it is available on Sourceforge. On the other hand, cluster benchmarking after major upgrades is uncommon – the resources are never made available for non-production work!

3.3 Panel A3 – Monitoring

Of the three sites that gave presentations, BNL use a mixture of commercial and home-written tools; they monitor the health of systems but also the NFS, AFS and LSF services. FNAL performed a survey of existing tools and decided that they required to develop their own, focussing initially at least on alarms. The first prototype, which monitors for alarms and performs some recovery actions, has been deployed and is giving good results.

The CERN team decided from the outset to concentrate on the service level as opposed to monitoring objects but the tool should also measure performance. Here also a first prototype is running but not yet at the service level.

All sites represented built tools in addition to those, which come "free" with a particular package such as LSF. And aside from the plans of both Fermilab (NGOP) and CERN (PEM) ultimately to monitor services, everyone today monitors objects – file system full, daemon missing, etc.

In choosing whether to use commercial tools or develop one's own it should be noted that so-called "enterprise packages" are typically priced for commercial sites where downtime is expensive and has quantifiable cost. They usually have considerable initial installation and integration costs. But one must not forget the often-high ongoing costs for home-built tools as well as vulnerability to personnel loss/reallocation.

A couple of other tools were mentioned – netsaint (public domain software used at NERSC) and SiteAssure from Platform.

3.4 Panel A4 – Grid Computing

A number of Grid projects were presented in which major HENP labs and experiments play a prominent part. In Europe there is the European DataGrid project, a 3 year collaboration by 21 partners. It is split into discrete work packages of which one (Work Package 4, Fabric Management) is charged with establishing and operating large computer fabrics.

In the US there are two projects in this space – PPDG and GriPhyN. The first is a follow-on to a project that had concentrated on networking aspects; the new 3 year project aims at full end-to-end solutions for the six participating experiments. The second, GriPhyN incorporates more computer-science research and education and includes the goal of developing a virtual data toolkit – is it more efficient to replicate processed data or recalculate from raw data. It is noted that neither of these Grid projects has a direct equivalent to the European DataGrid Work Package 4. Is this a serious omission? Is this an appropriate area where this workshop can provide a seed to future US collaboration in this area?

All these projects are in the project and architecture definition stage and there are many choices to be made, not an easy matter from among such wide-ranging groups of computer scientists and end-users. How and when will these projects deliver ubiquitous production services? And what are the boundaries and overlaps between the European and US projects?

3.5 Panel B1 - Data Access, Data Movement

A number of sites described how they access data. Within an individual experiment, a number of collaborations have worldwide "psuedo-grids" operational today. These readily point toward issues of reliability, allocation, scalability and optimization for the more general Grid. Selected tools must be available free for collaborators in order to achieve general acceptance. For the distribution of data, multicast has been used but difficulties with error rates increasing with data size have halted wider use.

The Condor philosophy for the Grid is to hide data access errors as much as possible from reaching the jobs.

Nikhef are concerned about network throughput and they work hard to identify each successive bottleneck (just as likely to be at the main data center as the remote site). Despite this however, they consider network transfers much

collaborator can generate up to 20 TB of data per year. Much of this stored data can be recreated, so the challenge was made: "Why store it, just re-calculate it instead."

The Genomics group at the University of Minnesota reported that they had been forced to use so-called "opportunistic cycles" on desktops via Condor because of the rapid development of their science. The analysis and computation needs expected for 2008 had already arrived because of the very successful Genome projects.

Turning to storage itself as opposed to storage access, one question is how best to use the capacity of the local disc, often 40GB or more, which is delivered with the current generation of PCs. Or, with Grid coming (but see the previous section), will we need any local storage?

3.6 Panel B2 - CPU and Resource Allocation

This panel started with a familiar discussion - whether to use a commercial tool, this time for batch scheduling, or develop one's own. Issues such as vendor licence and ongoing maintenance costs must be weighed against development and ongoing support costs. A homegrown scheme possibly has more flexibility but more ongoing maintenance.

A poll of the sites represented showed that some 30% use LSF (commercial but it works well), 30% use PBS (free, public domain), 20% use Condor (free and good support) and 2 sites (FNAL and IN2P3) had developed their own tool. Both IN2P3 (BQS) and FNAL (FBSng) cited historical reasons and cost sensitivity. CERN and the CDF experiment at FNAL are looking at MOSIX[4] but initial studies seem to indicate a lack of control at the node level.

3.7 Panel B3 – Security

Large sites such as BNL, CERN and FNAL have formal network incident response teams. Both of these have looked at using Kerberos to improve security and reduce the passage of clear text passwords. BNL and Fermilab have carried this through to a pilot scheme. On the other hand, although password security is taken very seriously, data security is less of an issue; largely of course because of the need to support worldwide physics collaborations.

Other security measures in place at various sites include --

- Disabling as many server functions as possible
- Firewalls in most sites, but with various degrees of tightness applied according the local environment and the date of the most recent serious attack!
- Increasing use of smartcards and certificates instead of clear-text passwords
- Crack for password checking, used in some 30% of sites

Many sites have a security policy; others have a usage policy, which often incorporates some security rules. The Panel came to the conclusion that clusters do not in themselves change the issues around security and that great care must be taken when deciding which vendor patches should be applied or ignored. It was noted that a cluster is "a very good error amplifier" and access controls help limit "innocent errors" as well as malicious mischief.

3.8 Panel B4 - Application Environment, Load Balancing

When it comes to accessing system and application tools and libraries, should one use remote file sharing or should the target client node re-synchronise with some declared master node. One suggestion was to use a pre-compiler to hide these differences. Put differently, the local environment could access remote files via a global file system or issue puts and gets on demand; or it could access local files, which are shipped with the job or created by resynchronisation before job execution.

The question "dynamic or statically-linked libraries" was summarised as – system administrators prefer static, users prefer dynamic ones. Arguments against dynamic environments are increased system configuration sensitivity (portability) and security implications. However, some third-party applications only exist in dynamic format by vendor design.

As mentioned in another panel, DNS name lookup is quite often used for simple load balancing. Algorithms used to determine the current translation of a generic cluster name into a physical node range from simple round-robin to quite complicated metrics covering the number of active jobs on a node, its current free memory and so on.

---

[4] MOSIX is a software package that enhance the Linux kernel with cluster computing capabilities. See the Web site at http://www.mosix.cs.huji.ac.il/txt_main.html.

mixes make this node the least appropriate choice for execution. These schemes worked well (often surprisingly well) at spreading the load.

Where possible, job and queue management are delegated to user representatives and this extends as far as tuning the job mix via the use of priorities and also host affiliation and applying peer pressure to people abusing the queues. Job dispatching in a busy environment is not easy – what does "pending" mean to a user, how to forecast future needs and availability.

4.0 Conclusion

The workshop ended with the delegates agreeing that it had been useful and should be repeated in approximately 18 months. No summary was made, the primary goal being to share experiences but returning to the questions posed at the start of the workshop by Matthias Kasemann, it is clear that clusters have replaced mainframes in virtually all of the HENP world at least but that, in particular, administration of them is far from simple and poses increasing problems as cluster sizes scale. In-house support costs must be balanced against bought-in solutions, not only for hardware and software but also for operations and management. And finally that there are several solutions, and a number of practical examples, of the use of desktops to increase overall computing power available.