Computational Shared Facility: Summary of the User Group Meeting, Friday 21st October 2011

Author: Neil Burton & Pen Richardson

Date: 27th January 2012

Dr Hood presented the current status of the CSF (presentation slides will be available on the CSF site); Dr Hood presented the data and proposals for the subsequent discussions.

Dr Pinning presented a brief overview of the Service Level Definition (SLD) and comments for the subsequent discussions based on the argument of Sustainability.

Discussions:

Revolving Green Fund

Two old (eScience) HPC clusters, owned by the University and administered by ITS, have been 'traded in' for new CSF hardware as part of the University's Revolving Green Fund. Two new C6100s (4 Intel nodes - 96 Cores) and four Nvidia 2050 GPUs will be ordered shortly. Dr Hood's presentation gave six ways in which these could be used and recommended that four of them were practical and worthwhile. It was generally agreed that free-and-open access to all University members could not be supported in the current model, but that some degree of controlled access to non-contributors would be appropriate: Hence the User Group approved options 3, 4, 5 & 6 to allow access to the 'RGF share' of the CSF for evaluation, pump-priming, contributor's soft-landing and training (the exact nature of the latter TBC but will include ITS for Research courses).

FLS expressed interest in the next round of RGF. The fund for this academic year has already been fully allocated.

Contribution – small/School contributions

At present the standard minimum buy-in contribution involves a cost of approximately £15-16k to enable the purchase of one C6100 unit (48 cores), or equivalent. There is general feeling that enabling some contributions below this minimum value would be advantageous; however there is

no official mechanism to pool smaller contributions. One administrative barrier to pooling small contributions is the current University finance model which does not provide a usable mechanism for carrying funds over to the next financial year.

It was generally agreed that although possible collation of funds at the individual School or Faculty level may be an option for some, formal mechanisms at this level would not be appropriate.

In the absence of any formal financial mechanism, workable suggestions included:

- ITS for Research to act as a broker to bring contributors with small amounts of money together;
- the User Group to coordinate the collation of unspent funds from potential contributors, perhaps via a special meeting in advance of end of year.

No satisfactory approach to managing contributions below the minimum unit of buy-in was reached.

The need to potentially increase storage in the future was raised. The new University Storage Area Network (SAN) currently being procured by ITS may help provide extra home filestore at a reasonable cost.

Future Service Sustainability

Hardware maintenance

The draft SLD from ITS states that the *service user* will "fund hardware maintenance beyond the 3 year support period up to 5 year hardware end of life". It was agreed that this was not in the spirit of the original CSF agreement.

It was further agreed by the User Group that the SLD should be updated to acknowledge the following:

- Hardware maintenance of contributed equipment beyond the initial three years maintenance will not be the explicit responsibility of either the University (ITS) or the contributors.

- The intention is that the hardware will remain in use until it is deemed to have reached end-of-service, providing that it does not break.
- If compute equipment which is no longer on maintenance fails, the default option will be to use available spare parts (if available from other defunct nodes) to keep as much of the system in operation as possible.
- The end-of-service for individual nodes or compute equipment is anticipated to be 5 years, as originally agreed by the user group.
- The end-of-service was agreed to be determined on the grounds of sustainability.
- The lifetime of hardware is not directly linked to the availability of the CSF service to a contributor; access to the CSF will be based on a share model.

Action: ITS/Dr Pinning to update the SLD to reflect the above points.

Shares and depreciation of a share

The first meeting of the User Group in March agreed to effectively pool all resources contributed and use fair-share scheduling to handle access to the resources, rather than allocating hardware to specific contributors and/or queues. At the June meeting the User Group decided that, for the purposes of the scheduling system, a trial 'share' system would be implemented based on the financial contribution made by a contributor at '1 share = £1000 worth of current compute power/resource'. This had been deemed fairer than a model based purely on the number of compute cores as some groups have contributed GPUs and Infiniband in addition to compute cores. It was acknowledged at the June meeting that as new contributions are made using the same share definition, but with up-to-date technology, old contributions would effectively have depreciated in relative terms, and as such the fair-share algorithm would need to take into account the depreciation of a contributor's 'share'. This concept would also help to define the 'soft landing' after the end-of-service defined for the hardware originally contributed.

It was agreed at the June meeting that a share depreciation model based upon the University's financial model would be investigated. It is now believed that the University depreciates all equipment over 5 years to zero. Thus for the purposes of the CSF it was felt that a different model of depreciation would be preferred since after the end of year five a contributor's share and thus access would drop to zero and would not take into account the previously agreed soft-landing. Some alterative depreciation examples were highlighted in Dr Hood's presentation, including models based on Moore's law to illustrate the effective long-term depreciation anticipated from a user's relative share value in a CSF system which continues to attract investors. It was stressed that any implemented share-depreciation based model would be normalised such that although a

contributor's relative share may change, the contributor's absolute share over time would continue to be approximately equivalent to the amount (hardware) of the original contribution.

Since the concept of depreciation of shares generated some initial confusion, alternate models of re-defining the value of an initial share, which would then remain constant, were also proposed and considered. It was noted that the advantage of such an approach over depreciation models would be that the impression that a contributor's initial contribution would reduce over time may be avoided.

Dr Pinning described an alternate model based on the number of cores contributed at a given time, with the share defined accordingly. When the agreed hardware contribution to the facility is stepped due to new technology, a scaling factor could be calculated using an agreed benchmark. The shares would be re-calculated based on this rather than a pre-defined rate of depreciation.

It was noted that re-scaling shares based on the technology refresh would be fairer than a pre-defined rate of depreciation and was, in principle, similar to a bespoke share-depreciation option presented earlier.

Discussions revolved around clarifying the financial model and the feasibility of benchmarking to determine a scale factor (using user codes or those widely used by the HPC industry) and the frequency of hardware upgrade points used to define the value of a share. The concept of basing a share on the number of cores was not popular since it did not account for a full range of hardware contribution options, particularly relating to interconnects, and that it was unlikely that a compute-based benchmark would be able to reflect the heterogeneity of possible core architectures and the variation in applications of all CSF contributors. Since vendor price at purchase reflects the contribution directly, it was generally agreed to be a more acceptable measure for the allocation of shares.

It was agreed that the final method used to regulate the SGE fair-share system should be simple, fair and clear to both current and new contributors. Not all contributors had been able to attend the meeting and this was deemed an important issue which needed further consideration. It was agreed that a share model based upon financial contribution was still the preferred way forward and that a depreciation model influenced by specific benchmark data at defined upgrade points could be devised.

Action: Service Owner and Chair to develop a proposal for a share allocation system to be included in the Service Level Definition (SLD) which would bring together the various options and considerations discussed.

Service Level Definition

Time did not permit a full discussion beyond the points raised as part of the other agenda items but it was agreed that the project needs to move into service and the SLD was agreed in principle. The Chair was interested to hear other views on the document.

Operational issues - scratch 'clean up' policy

Dr Hood's presentation proposed that files of one month or three months old be automatically deleted from scratch, with warning emails prior to deletion, to prevent scratch filling up. This mechanism has worked successfully in the past. Quotas could be applied, but this would slow down the Lustre filesystem and partially defeat the object of having a high performance filesystem. There was some discussion about the 'legality' of users using 'touch' and it was re-iterated that scratch is intended for running jobs only and temporary storage of files. Long term reliance on scratch is a bad idea because it is not backed up and users have in the past lost valuable work.

It was suggested by Dr Pinning that users could consider purchasing more home area to effectively store their results.

The User Group agreed that the initial policy should be set to three months and allow individual applications of 'touch'. This would be reviewed at a later meeting and usage of 'touch' monitored.

Action: ITS/Dr Hood to implement this policy.

Software – contribution policy

This was raised briefly as something the group may wish to consider in the future for the purchase of common applications and tools, such as the recent example of PGI compilers.

Next Meeting – to be confirmed.