HPCCC – Future Plans in Supercomputing

Philip Tannenbaum

Manager HPCCC

Background

The HPCCC is a collaboration between the Bureau and CSIRO to provide high performance computing resources for the diverse needs of both organisations.

Generalising, the Bureau has mission critical requirements for real-time computing, and CSIRO for capacity computing, and neither organisation uses true supercomputing capability where applications take over an entire system during execution; additionally, the most usable resources across the broadest range of requirements must be selected because budgets are finite.

In response to these diverse requirements, the HPCCC has proposed a vision for future systems that departs from considering "the supercomputer" as a monolithic resource, but rather as an integrated collection of high performance computing resources having varied architectures and characteristics.

Further, integration of the supercomputer into Bureau and CSIRO infrastructures will necessitate that they have similar software interfaces to the respective storage systems as mid-range and cluster systems, to provide a data-centric system environment.

Discussion on Computing

Today, the HPCCC supercomputer, Bureau mid-range systems, and CSIRO cluster systems are separated by management, network, and function, although they could and probably should be more tightly integrated in the future. Such integration would enable a simplification of applications and thus more robust operational systems, improve data locality and provide faster and more reliable data transmission between hosts, and would enhance overall usability by providing a simplified and more consistent computing environment.

Bureau operations and Bureau and CSIRO research programs have differing and diverse needs, which could be well served through a collection of high performance systems having varied architectures such as parallel vector, scalable parallel, and cluster systems.

Performance characteristics for Bureau applications reinforce that some run very well on parallel vector systems, others equally well on scalable parallel systems, and in a few cases adequately even on modest servers. In most cases, the success of Bureau applications is dependent on end-to-end run time, not simply the performance of a modelling step. Researchers are also sensitive to turn-around time more than absolute performance, because it affects workdays and project schedules.

End-to-end application time includes data set-up, pre and post-processing, and data distribution steps, so the concept of applications performance needs to be defined in a broader context than sustained computational GFLOPS. As an example, an applications system providing superior integrated data management, superior file system performance, and

moderate computational performance, could often provide a shorter end-to-end application time as compared to a poorly integrated system requiring multiple network transfers of data, moderate I/O performance, but having vastly superior computational performance.

A diverse range of computing platforms helps to enable efficient "applications systems" designed to minimise end-to-end time rather than focusing on single steps such as an NWP, climate, or ocean model in isolation from its ancillary components. The most appropriate application platform(s) could then be selected from those available with regard to capacity, capability, and data management, rather than only on computational performance as has been done in the past.
BMRC researchers have specifically requested linux based systems for research use, and Bureau operational applications are now either being developed on linux platforms, or there are plans to migrate them to linux platforms in the future. From the CSIRO perspective, experience with the HPCCC has underscored that their users have a broad range of needs beyond atmospheric and ocean sciences needing high performance computing cycles, a diverse range of applications, and ongoing integration with their data archival system.

linux operating systems are pervasive on scalable parallel and cluster systems, resulting in a significant base of free, shared, and licensed applications for those systems. In contrast, NEC vector parallel systems use SUPER-UX, and while IBM pSeries scalable systems offer linux, IBM recommends AIX for its superior reliability on larger configurations.

There are very few commercial applications available for the NEC system, and porting applications to it varies from simple to impractical. Applications availability for AIX is still quite good although it is progressively losing support in favour of linux.

The vision is therefore a core collection of high performance scalable systems that are integrated with a parallel vector system that is used as a back-end. The former systems provide capacity and generality for diverse applications, and the latter provides capability for specific applications. These systems would collectively be "the supercomputer."

Finally, as new applications are developed, the vision promotes ongoing comparative assessments of different computing products and architectures, which in turn promotes portability as a developmental goal. Because of often dramatic changes in the HPC industry over short times, it is suggested that from an organisational view, an application that runs reasonably well on many different platforms can be superior to one that achieves high efficiency on only one.

Discussion on Integration

Usability and end-to-end application time includes considerations such as system access to data storage systems, the functionality of data management software including hierarchical storage managers, the locality of data to the host requiring that data, and the flexibility of capacity provisioning.

In response the HPCCC has proposed and the Bureau has approved the Large Scale Data Storage System (LSDSS) project. The LSDSS aims to consolidate data storage into a single section that will provide data services to the preponderance of head office systems. The key feature of the proposal is migrating the Bureau central computing architecture from a system centric approach to a data centric approach.

Supercomputers have traditionally had significantly lesser support for ancillary systems, but virtually all vendors now offer solutions for such devices as global file systems, NAS-head devices, and true shared or sharable storage systems.

The LSDSS it is intended, among other aspects, to introduce a storage architecture that can enable capacity and file sharing among disparate operational and research systems. Such integration can reduce end-to-end time through eliminating or reducing data transfers between systems, and simplifies the user experience by presenting a global view of file systems and files – essentially enabling users to focus more on the data rather than the system hosting their data.

Beyond 2010 it is expected to become a requirement that HPCCC supercomputers will be capable of fully integrating into the LSDSS, rather than the LSDSS being adapted to interface to the supercomputer.

The LSDSS will be a flexible and extensible system that could host additional requirements should it prove to be attractive to CSIRO, or CSIRO could also continue with an independent large scale storage system.

Summary

The HPCCC's mission is to serve the broad needs of the Bureau, which includes mission critical real time operations and environmental research, and of the CSIRO, which includes a diverse application base requiring high capacity more than absolute performance.

The HPCCC has proposed a holistic view of high performance computing that departs from that of a traditional monolithic supercomputer. It includes multiple systems architectures, data management, and integration as being equally important.

A collection of various scalable systems running linux operating systems can support Bureau operations, and BMRC and CSIRO research. A wide range of applications software required is widely supported across virtually all linux systems, and many Bureau applications require, or will require, linux hosts for their next generation.

A parallel vector system will be minimally necessary to sustain operations during the transition from the Bureau's suite to the MetOffice Unified Model, and the balance between the various types of systems will be periodically re-evaluated to ensure systems continuously meet performance and service requirements.

Complementing the supercomputer will be the LSDSS that will provide the foundation for future integration of the supercomputer with the greater data storage systems used by a wide range of systems.

The goal is to provide an integrated infrastructure rather than to focus on optimising individual parts, in relative isolation.