CaStLeS (Compute and Storage for the Life Sciences): a collection of compute and storage resources for supporting research at the University of Birmingham

Simon J. Thompson[1], Stephanie E. M. Thompson[1] and Jean-Baptiste Cazier[2]

For correspondence; bearinfo@contacts.bham.ac.uk

Date of publication: 20th June 2019

Affiliations

1. Advanced Research Computing, University of Birmingham
2. Centre for Computational Biology & Institute of Cancer and Genomic Sciences, University of Birmingham

Abstract

Life Sciences researchers increasingly rely on advanced computing and storage facilities to support their work. Whether this be storage to retain images from high resolution microscopy for later processing, or powerful CPU compute resources to complete bioinformatics analysis, well-resourced and managed systems are essential to support leading research. The rapidly expanding interest in Data Sciences and AI is also rendering more prevalent the need to support the application and development of novel methods for data analysis. This paper outlines the strategic facilities made available by the University of Birmingham to support Life Sciences research.

Introduction

During 2015, the University of Birmingham formed the Centre for Computational Biology (CCB) to combine many of the field's components, from both the development and applications sides. The CCB focusses around the three major areas; 1) applications, 2) training and 3) development, which are disseminated through the many fields of expertise found within the Colleges of Medical and Dental Sciences, Life and Environmental Sciences, and Engineering and Physical Sciences. With the formation of the CCB, a new team of academics joined the University and whilst they appreciated the availability of high performance, central computing provision as part of the Birmingham Environment for Academic Research (BEAR, www.birmingham.ac.uk/bear), they also quickly identified the need for additional compute and storage facilities tailored for the specific requirements of the growing field of bioinformatics.

Following further discussions with senior researchers across the life sciences field at the University, it was agreed that a strategic funding paper would be submitted to the University

which resulted in the provision of £2M in capital funding to support the purchase of computing and storage facilities dedicated to these disciplines. This funding was provided to Advanced Research Computing to deliver a compute and storage facility that could be integrated into the existing BEAR facilities, providing economies of scale and concentrated expertise to manage, operate and support this advanced infrastructure and the services built on top. The resulting infrastructure was defined as Compute and Storage for Life Sciences (CaStLeS). In addition to the capital investment, a number of staff posts were also funded to strengthen Advanced Research Computing who manage the infrastructure and services as well as deliver training and software development expertise to researchers in all fields.

Resource Overview

Facilities Available

The outline of the facilities provided in this paper is a snapshot representation at the time of publication, continued funding is expected to enhance, develop and replace the facilities in the future.

The CaStLeS resources are broadly split into two key areas, the provision of large volumes of storage for research data and the provision of high-powered compute resources. Within these areas there are some further sub-divisions.

Storage Facilities

Storage is provisioned from a central set of storage systems which use Spectrum Scale (IBM Corporation 2019) to abstract the underlying storage from the data and control planes for user access and managing the system. Presently this is deployed on DDN SFA12kx and DDN SFA7700 storage arrays fronted by 4x Lenovo x3650m5 and 2x Lenovo x3650m5 NSD storage servers. The storage is attached using FDR InfiniBand with point to point connections with the NSD servers, then in turn attached to BEAR InfiniBand fabrics with multiple 100Gb EDR connections and multiple 40Gb Ethernet connections.

The storage is deployed over two data centre sites and is also all backed up. Backups are scheduled to run nightly and are placed onto tape in two physically separate sites. The backup system automatically retains deleted data for a period of time once deleted and also retains multiple copies of a file. In addition to this, to detect silent tape failures over time, the system automatically validates all tape media regularly and flags for attention any errors when data is read from the tape. Users are also able to request two types of storage, these are non-replicated (the default) and replicated. Where non-replicated is selected, this is placed in the primary data centre for research computing. Purpose-built in 2018, this innovative facility operates highly efficiently and is designed to support the needs of research, requiring only limited operational downtime. Replicated storage is synchronously copied to a second data centre and is intended only for extremely high value data or where high availability of the data is required. The high-availability storage is designed and tested to remain accessible should there be a data centre outage in either site, this does not however mean that compute capability is also resilient. (Note that making anything resilient multiplies costs and introduces performance penalties). During early deployment, researchers were confused about the

difference between replication and backup and so it is important to distinguish between the two offerings. The storage system currently consists of 2PB in the primary data centre and 1PB in the secondary data centre, which equates to 1PB usable of replicated storage and 1PB of non-replicated storage.

Compute Facilities

There are three main types of compute facility, these are; 1) HPC compute, 2) private cloud (CaStLeS VMs) and 3) accelerated computing and AI.

The HPC compute facility is integrated as part of the University's BlueBEAR HPC facility and is presented as a number of queues within the SLURM HPC scheduler (Jette et al. 2002), which are reserved for CaStLeS users. This gives them preferential access to the compute resource, although if unused, short (<10 minute) jobs from other BlueBEAR users may run on the resource. This arrangement is aimed at ensuring the system is well utilised and the University gets the best value for the investment it has made.

The HPC capacity (CaStLeS compute, sometimes referred to as qos) consists of:

- 24 Lenovo water cooled nx360m5 compute nodes, each with 2x 10 core Intel E5-2640v4 2.4GHz and 128GB RAM (480 cores total)
- 1 Lenovo x3650m5 compute node with 2x 10 core Intel E5-2640v4 2.4GHz and 1TB RAM (20 cores total)
- 1 Lenovo x3650m5 compute node with 2x 10 core Intel E5-2640v4 2.4GHz, 1TB RAM and 2x NVIDIA P100 GPUs (20 cores total)
- 18 Lenovo water cooled SD650 compute nodes, each with 2x 20 core Intel Gold 2.5GHz and 192GB RAM (720 cores total)

The private cloud capacity (CaStLeS VMs in BEAR Cloud) consists of:

- 78 Lenovo water cooled nx360m5 compute nodes, each with 2x 10 core Intel E5-2640v4 2.4GHz and 128GB RAM (1560 cores total)

The CaStLeS accelerated computing and AI capacity, part of BEAR AI, consists of:

- 8 IBM water cooled AC922 compute nodes, each with 2x 18 core IBM POWER9, 1TB RAM and 4x NVIDIA V100 GPUs with NVLINK interconnect (288 cores total, 32 V100 GPUs)

All of the compute nodes are connected to the 100Gb EDR Mellanox InfiniBand interconnect and are equipped with 10Gb Ethernet. This allows large scale parallel MPI jobs, or multi-node GPU code to be run. The InfiniBand interconnect is also used to provide access to the CaStLeS storage systems and uses RDMA technologies to allow storage access to be directly loaded into memory across the interconnect.

The design of the compute capacity is such that the hardware can easily be moved between HPC and private cloud capacity, pilot work is also being undertaken to provision HPC compute resource with near bare-metal speeds to allow unallocated cloud resource to be utilised for

HPC batch workloads. The use of water cooling, which takes water directly across CPUs, memory and GPU cards allows up to 90% of the heat generated to be directly captured and removed from the data centre, leaving only a small amount of residual heat to be expelled to traditional air handling systems. The water used to cool the systems can be brought in at 36°C and this allows for air-blast cooling systems to be used externally, rather than traditional compressive air cooling, which saves significant amounts of energy in the cooling system.

In the case of the private cloud environment, virtual machines (VMs) can be placed on these, typically running either CentOS or Ubuntu, and are provisioned to integrate with external identity sources allowing standard University usernames and passwords to be used for access. Systems which are integrated in this are also able to access the high-speed storage systems directly, thus providing a "data everywhere" environment. Where projects require graphical applications, tools are provided to accelerate remote graphical displays. VMs are also able to be used as infrastructure as a service (IaaS) and a number of research projects take advantage of this capability.

VMs are typically provisioned by "colour", with red being the largest and green the smallest. The size of each colour VM (in terms of CPU and memory allocation) is set to optimise placement of VMs onto the systems, so that memory and CPU is not over-allocated. For example, an 18 core VM will have 18 real cores allocated to it and there will be no CPU over-allocation. Care has also been taken to ensure NUMA placement is optimised to ensure memory is directly available to cores. A small number of cores on the hypervisor systems are reserved for system management functions to ensure they remain responsive even under heavy load.

Users of the CaStLeS systems are able to request scientific software installations on both compute and VM systems. To ensure consistency of software builds, and in support of reproducible science, EasyBuild (Hoste et al. 2012) is used to assist with the management and installation of software on the systems. This tool allows "recipes" to be developed which specify how the software is compiled and installed, in addition to any dependent packages it might have. Whilst there is inevitably a lead time to productive use, once these types of tools are in place it simplifies the installation and the process when building over multiple processor technologies or operating systems. Automation tools such as xCAT (IBM Corporation 2018b) and SaltStack (SaltStack 2018) are used to support deployment and configuration of systems, enabling reliable and consistent deployment of environments. In addition, the commercial tools PowerAI Enterprise (IBM Corporation 2018a) (renamed Watson Machine Learning Accelerator in April 2019) are also installed which provide workload and accelerated capability to some of the AI/Deep Learning facilities.

Resource Allocation and Access

The CaStLeS facilities are available as a easy access, rapidly deployed, pump-priming resource to any researcher working within the Life Sciences at Birmingham. However, an application process is required to make the academic and strategic case for use of this valuable resource. This is a two-stage process and researchers are expected to outline their storage and compute requirements before access is granted. The first stage is a technical review where a member

of Advanced Research Computing will undertake a short review of the request to ensure it is technically feasible and likely to be a successful use of CaStLeS. Once this has been completed, the request is passed to the CaStLeS Executive (Exec), which is formed of the three academic chairs of the CCB who assess the academic case for use of the resources. In some cases, researchers may be asked to clarify or further justify their request before approval can be given.

Once approved, the request will be passed back to Advanced Research Computing who will then provision resources. Where a virtual environment is required beyond a basic server with data access, the team will also work to deploy the full stack of software to support the research group and to ensure that the resultant service is reliable and well-engineered.

The implementation of the governance model for CaStLeS at the very early stages of development of the facilities has been key to the successful delivery of resources to researchers. Given the finite nature of resources, the review process is essential to ensure resources are used appropriately. As part of the development of these processes, it was agreed that Advanced Research Computing would only undertake technical review of the requests and that all research/science-based decisions would be made by academic peers within the University.

To support the overall scientific goals and provide strategic direction for CaStLeS, a Strategic Oversight Group also exists which includes senior academic representation from Colleges involved in Life Sciences research, the CaStLeS Exec and members of Advanced Research Computing. This Group meets termly, reviews usage allocations and discusses any intervention required to support the goals.

Resource Impact

The CaStLeS resources have been used by a wide variety of research projects, small and large, across many Schools/Institutes within the University, as well as part of national and international collaborations. The secure environment provided is particularly essential when dealing with access to data to/from multiple institutions. Therefore, beyond the resource provision, CaStLeS is enabling the establishment of interdisciplinary and multi-site initiatives.

Principal Investigators are asked to include the costs of CaStLeS resources (both computational power and storage capacity) in subsequent grant applications to add capacity for their substantive project needs while ensuring the core service remains available to support the start-up of the next wave of projects.

PIs have already purchased additional resources from grants being awarded on the basis of extending the existing facilities and taking CaStLeS services rather than buying and running their own "server/NAS in a cupboard". The investment has been well viewed by a number of external researchers who are often surprised at the level and sophistication of facilities available to support research initiatives. Equally, feedback from funders is supportive of this model of immediately available, professionally supported, high performance service for research projects.

Finally, CaStLeS has proven invaluable in the provision of training resources with the increasing need of Data Analytics in the Life Sciences; it enables an easy provision of purpose-built VMs at scale for a short period of time, with both dedicated storage and computing power.

Recognition

In order to justify and maintain these facilities, publications and other forms of media communication, including media appearances, press releases and conferences, should acknowledge the role of CaStLeS in the success of the research project.

Acknowledgements

References

Hoste, Kenneth, Jens Timmerman, Andy Georges, and Stijn De Weirdt. 2012. "Easybuild: Building Software with Ease." In *Proceedings - 2012 SC Companion: High Performance Computing, Networking Storage and Analysis, SCC 2012*, 572–82. IEEE. https://doi.org/10.1109/SC.Companion.2012.81.

IBM Corporation. 2018a. "PowerAI Enterprise." https://developer.ibm.com/linuxonpower/deep-learning-powerai/powerai-enterprise/.

———. 2018b. "XCAT." https://xcat.org/.

———. 2019. "Spectrum Scale." http://www.ibm.com/Storage/Scale.

Jette, Morris A., Morris A. Jette, Andy B. Yoo, and Mark Grondona. 2002. "SLURM: Simple Linux Utility for Resource Management." *IN LECTURE NOTES IN COMPUTER SCIENCE: PROCEEDINGS OF JOB SCHEDULING STRATEGIES FOR PARALLEL PROCESSING (JSSPP) 2003* 2862: 44--60. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.10.6834.

SaltStack. 2018. "SaltStack." https://github.com/saltstack/salt.