IDC Australia Level 20, 8-20 Napier St, North Sydney NSW 2060 Australia P.61.2.9925.2298 F.61.2.9957.2330 http://www.anz

IDC CUSTOMER SPOTLIGHT

Research Cloud at Monash (R@CMon) – A New Chapter For Australia's Global Leadership in Scientific Research

December 2013

By Raj Mudaliar

Sponsored by Dell and AMD

In this era of digital connectivity, NeCTAR (National eResearch Collaboration Tools and Resources) is combining an exciting mix of ideas and technologies to create new ways for Australian researchers to connect, work and collaborate on a national and international scale. NeCTAR, a $47 million Australian Government Super Science project, has joined hands with the Australian research sector, which has committed $54 million for coinvestment, to create a new world-class infrastructure for Australian researchers. This will include: new Virtual Laboratories; a national Research Cloud; new eResearch Tools; and a secure and robust national hosting service for researchers. From physicists, archaeologists, software engineers and arts disciplines, this is a bold and pioneering attempt to encourage researchers to connect, colloborate, create new ways of working, and create new knowledge of ever increasing complexity. All of which are key ingredients for new innovations, scientific discoveries and important breakthroughs. In June 2013, Monash University became the second organisation to join the NeCTAR Research Cloud adding 2,300 new cores. Dell and AMD are the technology partners chosen to create the NeCTAR Research Cloud at Monash University (R@CMon).

Introduction

Since the start of the industrial revolution more than 250 years ago, the global economy has been on a steep growth trajectory propelled by a series of advances in technology. From steam engines that replaced water mills to electricity, telephones, automobiles, airplanes, transistors, computers, and the Internet, each new wave of technology has brought about surges in productivity and economic growth, enabling efficient new methods for performing existing tasks. The Internet is an excellent example. It introduced new ways of communicating and using information that enabled major innovations, imposing new rules from outside on all sorts of industries, rearranging value chains and enabling new forms of competition.

MONASH University

nectar

Solution Snapshot

Organisation: Monash University

Operational Challenge: Design and deploy a high-performance computing node site at Monash as part of NeCTAR, a nationwide federated research cloud.

Solution: Monash University selected the Dell PowerEdge C6145 using AMD Opteron™ 6200 Series processors to deploy their HPC cloud.

Project Duration: 5 years

Project Cost: AU$4.5m

Benefits:

- Offer innovative services for researchers to be rapidly deployed.

- Deliver high quality research outcomes.

- Foster a culture of innovation that drives scientific breakthroughs.

- Attract and retain top global talent.

- Facilitate collaboration on a national and international scale.

#AU63323W

Advances in science and technology contribute significantly to economic development, innovation, social well-being, environmental sustainability and the nation's prosperity. With this in mind, the Australian research community was invited by the Australian Government in September 2011 and May 2012 to submit project proposals that fulfill an aspect of the NeCTAR vision. An expert panel of Australian research leaders, nominated by the Australian research community, recommended 16 e-Research tool projects, 13 virtual laboratory projects and eight research cloud node projects for funding. The Government and Australian research sector together have committed to inject AUD$101 million to Australia's research infrastructure.

As part of this initiative, the NeCTAR project is partnering with Australian universities and research institutions to build for the first time, a national cloud for Australian researchers. Owned and located at eight institutes around Australia, yet operating as one cloud system, the NeCTAR research cloud is a federation believed to be a world-first. In 2012 Monash University was selected to be one of the eight NeCTAR node sites.

Monash University has longstanding commitment to world-class infrastructure that fosters research innovation – with countless examples in the fields of biomedicine, science and engineering. It has an established network of technology platforms that underpin much of its research efforts. With a commitment of AUD$4.5 million, Monash will develop and operate the R@CMon infrastructure and services in accordance with the NeCTAR lead node and governance framework. The facility will include approximately 4,000 mixed commodity and high-performance computational nodes, plus an additional critical mass of general purpose computing on graphics processing units (GPGPUs) and high-I/O data-throughput processing capabilities that will compliment other proposed NeCTAR activities and add value to existing infrastructure.

The responsibility of building the R@CMon was assigned to teams across the Monash eResearch Centre (MeRC), eSolutions (Monash's central ICT provider) and the NeCTAR lead node at University of Melbourne. The facility is being built in two phases, allowing the configuration of heterogeneity in the second phase to match the demands of the diverse research disciplines across Australia.

Steve Quenette, Deputy Director and Strategic Initiatives Manager of the Monash eResearch Centre, provided the guiding vision and agenda for R@CMon and Monash's research computing platforms at large. Colin Blythe, servers and storage manager of eSolutions at Monash, has the mandate to ensure that the computing and application infrastructure is designed and operated to meet defined criteria, in terms of performance, availability, recoverability, manageability, security and cost of delivery. Blair Bethwaite, Senior HPC Consultant at the Monash eResearch Centre, was the lead architect for the solution.

Key Challenges

Research is about discovering something new. With 30yrs of commodity computing and ever increasing digitisation and sensing means society has vast amounts of data about the things we're interested in. For example, it is common now for data-intensive applications to process terabyte- to petabyte-sized datasets. This data commonly comes in several different formats and is often distributed across multiple locations. Processing requirements typically scale near-linearly with data size and are often amenable to straightforward parallelisation. Key research issues involve data management, filtering and fusion techniques and efficient querying and distribution. Conversely, data/compute-intensive problems tend to rely less on the data throughput but stresses the processor and memory capabilities of a computer. Application requirements may also place time bounds on producing useful results. Some of the key challenges to build the R@CMon were:

- High performance computing platforms to provide uniform high speed memory access to multiterabyte data structures.

- Specialised hybrid interconnect architectures to process and filter multigigabyte data streams coming from high-speed networks, scientific instruments and simulations.
- Software integration capabilities that facilitate the plug and play integration of software and hardware components running on diverse computing platforms to quickly form analytical pipelines.
- New metadata management technologies that can scale to handle complex, heterogeneous, and distributed data sources.

R@CMon Goes Live

Cloud services today are no longer just about IT cloud services, they are about delivering new value to the business. Cloud services as alternatives to IT infrastructure investment is still a common tactical move by CIOs, but already the early cloud adopters have moved onto the next stage of generating value from the cloud. Key to this is identifying the real benefits that the new cloud-enabled business services deliver. IDC observes that in 2012-2013, the decision to use cloud services or not is now secondary to how well or not the business service being sought suits the business requirements.

NeCTAR's vision is to establish a federation of cloud nodes with dedicated resource management and a consistent user feel by mandating the use of the OpenStack architecture. Users dial in which resources they need through the NeCTAR OpenStack dashboard and access rights are determined by merit allocation or commercial means.

Monash worked with Dell and AMD HPC specialists to design a cloud solution that would realise research impact – the primary business driver to both NeCTAR and Monash University. Given the diverse usage – from web applications to databases to HPC, there was recognition that application demands will be unpredictable. For effective management of the resources and quality of service levels, considerable effort went into establishing consistent and robust throughput on integer compute, floating point compute, memory latency and file latency. The requirement was to understand and maximise the scalability of each measure. To achieve this high-core density, high memory density, expanded I/O capabilities and power efficiency were all considered.

The NeCTAR HPC research cloud at Monash used Dell's PowerEdge C6145 since it fulfilled the functionality requirements of its current and expected future workloads. The PowerEdge C6145 not only delivers performance with a two 4-socket AMD Opteron™ 6200 Series processor-based servers in a 2U but also uses less floor space, cabling and racks compared to individual servers. However, high cores and memory aren't enough without the I/O bandwidth to quickly compile data and results, and be able to connect to other resources. The PowerEdge C6145 server ramps up the I/O with 40 Gb/second throughput per x16 slot. Six PCIe Gen 2 x16 slots, two x8 mezzanine and two dedicated x16 host interface card slots (HICs), offers a total of 10 slots for much flexibility.

The twelve-core AMD Opteron 6234 processors in the Dell PowerEdge C6145 allows users to scale up to 96 cores and up to 1 terabyte of memory to cope with sudden peaks in demand, while at the same time remaining power efficient. The processors are specially designed to perform well in heavily threaded and virtualised server environments, but they also have years of power management expertise integrated into the design, such as TDP power gating and AMD Cool Core technology. The AMD Opteron 6200 Series processor design also optimises compute efficiency by sharing certain features among the cores, such as L2 cache and a 256-bit floating point unit per every 2 cores, while dedicating resources such as integer pipelines and L1 cache to each individual core. This enables optimal power efficiency without compromising performance. After successful hardware and software testing, Monash was satisfied with Dell and AMD's HPC solution that was fully aligned with R@CMon's core focus on delivering high quality research outcomes. Secondly, the solution was

offered with solid commercials that were highly cost effective compared to competition whilst at the same time ticking the box on all specified technical requirements. The proposed HPC research cloud node site at Monash was successfully approved by NeCTAR.

In June 2013 the first phase of the Monash node of the NeCTAR Research Cloud went live to public. This implementation delivers 2,300 cores of aggregate compute capability and is managed using the open source 'OpenStack' HPC management tools. The monthly utilisation average is about 75 to 80% and nearly 8,000 virtual machines are created every month. At any time, on average 90% of raw capacity (nodes) is available whilst the remaining 10% is under maintenance. Interestingly, the majority of the users are from outside Victoria, notably South Australia, reflecting preference of researchers to use the R@CMon because of its high performance, availability and reliability. Phase 2 will double the capacity to approximately 4,000 cores and is expected to be completed by Q1 of CY 2014.

Benefits and Outcomes Delivered

Monash takes prides in being one of Australia's leading Universities, aspiring to be a world-class research institute. Monash has long been an incubator for groundbreaking research. Things that we take for granted now, such as seat belts in cars and in-vitro fertilisation (IVF) treatment owe a debt of gratitude to the initial work of Monash researchers.

The R@CMon is the beginning of a new chapter for Australia's global leadership in scientific research, innovation and discoveries that will enable scientists to lead and collaborate with teams around the world, across all levels of industry and government. As an Australian secure platform to share access to research applications, the R@CMon supports the increasingly collaborative nature of Australian research. Some of the key benefits delivered are:

- Facilitate high quality research outcomes that may lead to new scientific discoveries. The self-service environment provided by the cloud allows researchers to quickly and effortlessly gain access to high-performance computing at a scale that dramatically increases scientific computing throughput. In ten days of July 2013, one tool running on the cloud investigated over 200,000 power-grid configurations (equivalent to over ten years of serial computing) driven by an optimisation searching for tradeoffs between cost, grid stability and emissions – gaining accurate insights into such complex systems will help inform the shift to sustainable energy grids and systems.
- Attract and retain the world's top talent of researchers and scientists.
- Empower researchers with new self-service abilities to publish research data, share knowledge, rapidly deploy and access scalable computing resources and software applications without the burden of operating their own computer servers. An example of this success is integration of instrument data from the Australian Synchrotron, next-door to Monash Clayton Campus, with computing and storage from the research cloud. The MyTARDIS data management portal, running on the cloud, is used to ingest and process terabytes of data from the Synchrotron every week, pushing it to stable long-term storage and providing index, discovery and publication services to collaborators around the world.
- NeCTAR's eResearch tools and virtual laboratories will deploy their applications in the national research cloud, providing access to national research communities and foster innovation in research software services. Computational results can be easily shared with national and international collaboration partners.

Lessons Learnt

Monash University's research goals can be summarised in three words: excellence, relevance and impact. To achieve its research goals, Monash recognises the importance to connect with industry, government and the world's best research organisations. R@CMon is a step forward to create an environment for researchers to share information, connect and collaborate and create new ways of working, ultimately leading to new innovations and breakthroughs. Some of the key lessons learnt were:

- **Build and test rigour to be maintained:** The compute building blocks need to be HPC stress tested despite the facility being named a 'cloud' as a large cohort of research usage is compute intensive in nature. Performing these tests early in the build life cycle saves time and money.
- **Being ahead of the learning curve:** Since the research cloud uses OpenStack cloud management tools, it is important to remain ahead of the learning curve for new releases. Also it helps to identify and get the right people from the team to get involved in the OpenStack community to make sure there is more collaborative learning.
- **Better monitoring and increased automation:** From an operational perspective, a more concerted effort is required to better monitor and detects performance issues in the node by using appropriate automation tools. Adopting a process centric approach that allows smooth flow of information between server, storage and software teams was found to be very helpful.
- **Understanding the research domain:** Cloud and HPC requires a better understanding of the research process vis-a-vis enterprise computing. Broader engagement across IT and research teams can enable the delivery of appropriate technology solutions that caters to specific user requirements. This also helps to build the right workflows and next generation tool sets that supports high quality research outcomes.

Methodology

The information for this IDC Customer Spotlight was obtained from multiple resources including information supplied by Monash University and questions posed by IDC's senior analyst directly to Colin Blythe, the servers and storage eSolutions manager at Monash University in September 2013.

IDC Australia P/L: Level 20, 8-20 Napier Street, North Sydney, NSW 2060 www.idc.com