Seeing the Big Picture: A Digital Desktop for Researchers

Bernard Meade[1,2], Christopher Fluke[1], Richard Sinnott[2], Steven Manos[2], Neil Killeen[2], Paul Mignone[2], and Michael Wang[3]
[1]Swinburne University of Technology; [2] The University of Melbourne; [3] NVIDIA Corporation

The rapid increase in size of experimental and simulation data requires researchers to rethink the way they interact with data to discover new knowledge. One of the many challenges of big data is how to support visual inspection of very large datasets. With sophisticated software, extremely large datasets can be reduced to more understandable graphical summaries. However, these data reduction methods can make it difficult to observe unexpected phenomena at the limit of detectability. In the case of very-high resolution images or image collections, it is beneficial to include a manual inspection stage to support and verify automatic detection algorithms. Tiled Display Walls (TDW) provide a valuable aid for such a process, but because of costs and physical size, have been overlooked by many researchers as a viable option. The recent availability of commodity UltraHD screens offers a cost-effective alternative. For desktop-based activities that draw data from several sources, having a display that allows all these items to be displayed simultaneously improves cognitive performance (Ball & North 2005).

A second consideration is how to use TDWs or UltraHD screens effectively for remote collaboration. While networks have become increasingly robust and reliable, the bandwidth is not expanding at the same rate as data collection technologies. Local storage often represents a potential single point of failure and traditional local backup methods are no longer as cost effective as online options. Also, many datasets used by modern researchers exceed the storage capacity of local systems.

Our particular application area of interest is astronomy: where high-resolution images, vastly exceeding the resolution of standard displays, are generated at a rapid pace from new observational facilities. In this paper we discuss the research underpinning the use of TDWs in astronomical research. We consider UltraHD displays as intermediate options between standard desktop displays and TDWs, and discuss the practicalities of using such displays to enhance the typical desktop environment. Finally we test the capabilities of the Australian Academic Research Network (AARNet) in terms of very large file transfers. Transfer tests have shown that for files from one gigabyte to one terabyte, the network scales up approximately linearly, particularly for some parts of the country, such as Canberra to Melbourne, but less so for other places, such as Western Australia to Melbourne. This allows us to put limits on the image size, and interaction speed, for remote collaborative inspection of high-resolution images.

Introduction

Technological advances in High Performance Computing infrastructures, such as distributed architectures, graphics processing units and cloud computing, have led to a dramatic increase in the volume of data available for scientific purposes. The challenge now is to determine how best to meaningfully interpret these enormous datasets to enhance and advance knowledge discovery.

Visualization is often the key ingredient to understanding data. Presenting information in graphical summaries can help reduce an overwhelming volume to its essence, and provide improved insight into the phenomena being studied. One of the most challenging aspects of this reduction process is to ensure that no salient information is discarded or compressed beyond detection. The reality facing the modern, data-rich researcher is that the number of data points to display vastly exceeds the number of screen pixels available.

While parallel computing has given rise to parallel visualization [e.g. ParaView; (Ahrens et al. 2005)] display technologies have not necessarily kept pace with computer power. As Table 1 shows, typical desktop or portable devices are only able to display images ranging from 2-5 Megapixels in size. Rather than being restricted to a single display, a reasonable alternative is to spread the data across multiple displays. We refer to such a solution as a tiled display wall (TDW).

Table 1
Screen resolutions of standard displays, UltraHD displays and tiled display walls.

Display	**Resolution**	**Image Size (Megapixels)**
Standard desktop display	1680 x 1050	1.7
iPad (with retina display)	2048 x 1536	3.1
Dell UltraSharp desktop display	2560 x 1600	4.1
Macbook Pro	2880 x 1800	5.2
4K UltraHD display	3840 x 2160	8.3
OzIPortal	15360 x 6400	98.3

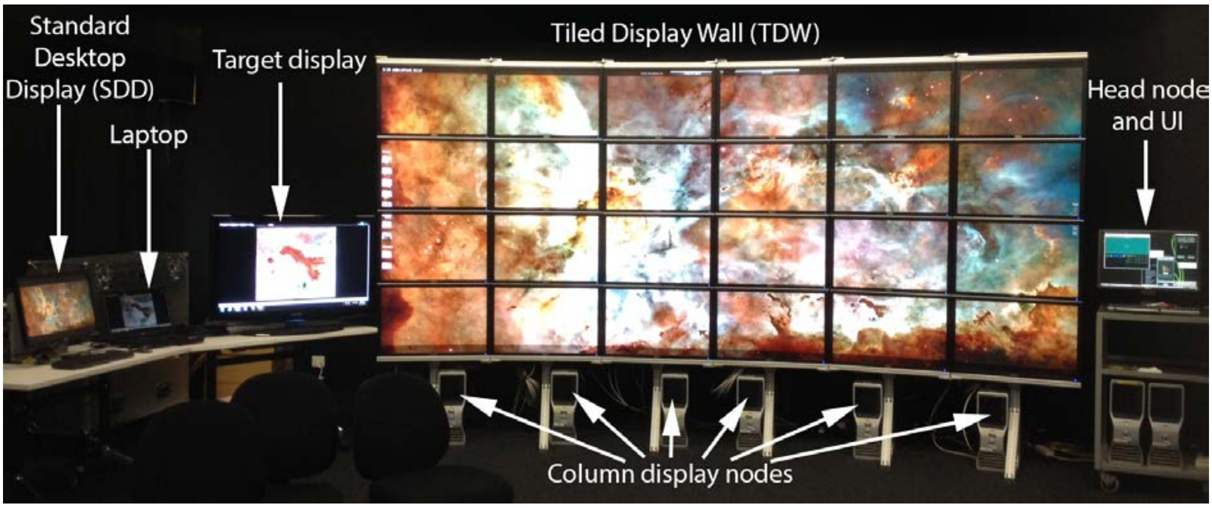

Figure 1. Configuration of the OzIPortal tiled display wall used by Meade et al. (2014). The six columns and four rows of Dell UltraSharp displays (2560 x 1600 pixels each) combine to produce a 98.3 Megapixel image (15360 x 6400 pixels). Note the presence of bezels between pairs of displays, which can distract from the visual content. The displayed image is the Carina nebula (image source*:* http://hubblesite.org/gallery/album/nebula/pr2007016a/hires/true/).

A typical TDW (see Figure 1) comprises a matrix of commodity displays. User interaction is via a head node, which coordinates communication with the individual compute nodes that generate pixel data for a column of displays. First appearing more than a decade ago through initiatives such as the OptIPortal project (DeFanti et al. 2009a), TDWs were expected to become the display of choice for high-resolution data sets. Unfortunately, the uptake of TDW at research institutions has been limited in practice due to their perceived complexity, cost and space requirements.

In the last few years, a new option has appeared: the low-cost, consumer UltraHD display (typical resolution of 3840 x 2160 pixels). While it is possible to use UltraHD displays in a TDW configuration, more value is likely to be obtained by simply making use of the display as a stand-alone, high-resolution desktop display. Costing a few per cent of the price of a TDW, it becomes a far more attractive option for researchers who need more screen real estate. The UltraHD display has the benefit of replacing the multi-display desktop that is increasingly common among researchers, allowing simultaneous heterogeneous content display. Several studies (Czerwinski et al. 2003; Robertson et al. 2005; Ni et al. 2006) have shown performance improvements with typical office-like applications when display windows can be spread out, much the way a traditional desk with paper worked.

TDWs and UltraHD displays (collectively *large-format displays)* can also enhance collaboration between remote colleagues. Many applications now exist to provide effective real-time collaboration, due largely to improved bandwidth and reliability of underlying networks (e.g. Skype, Google Hangouts, Zoom, EVO and others). Using larger displays allows visual presence of a remote colleague via videoconference, as well as shared document workspace and other communication technologies, such as shared desktops and digital whiteboards. This approach is teaching researchers to engage with content that is not directly attached to their own local computer.

As data volumes become increasingly difficult to transfer, the separation of researchers from their data is causing new challenges. One solution is to transfer *some* of data in order for it to be displayed and inspected. A recent development is the Virtual Display Infrastructure (VDI) concept: the delivery of a user's computing desktop or windowed application from a remote service. In this case, a virtual desktop is created on a virtual machine, often hosted in a cloud. Pixels that would ordinarily be directed to an attached display are instead streamed, in real-time, to the remote user's computer/display. With suitable bandwidth and graphics processing capabilities of the VDI host, the experience for the user can be smooth enough that no distinction can be made between a local and remote interface.

In this work, we road-test the UltraHD display. We compare its suitability for large-format image inspection with the investigation of TDWs in Meade et al. (2014). We describe and demonstrate how VDI can be used to deliver content to an UltraHD display as a requirement for remote collaboration. Ultimately, the ability to make use of remote collaboration and VDI depends on the underlying bandwidth. We examine whether the existing national research infrastructure in Australia – the Australian Academic Research Network (AARNet) and the National e-Research Collaboration, Tools and Resources (NeCTAR) Research Cloud – has the capabilities required for remote display of UltraHD video streams in real-time.

We find that the research network is robust between well-established centres such as National Computational Infrastructure in Canberra and the University of Melbourne, but is considerably less stable between the Pawsey Supercomputing Centre in Western Australia and the University of Melbourne. We consider the implications for researchers in a bandwidth-limited environment.

Background

Visual representations have played an important role in helping researchers engage with their data, recognizing important elements and trends that lead to new knowledge (Fluke et al. 2006). Astronomy is a scientific discipline where this is particularly true. Astronomy has traditionally been a visual science, both in terms of the way that it involves the collection of images, and in the role that visual inspection of data has played in identifying anomalies, image-based artefacts and for knowledge discovery.

Existing and next generation cameras and detectors will take astronomy deeper into the realm of "big data" (see examples in Table 2). Working at the exascale (Quinn et al. 2015), the Square Kilometer Array will produce approximately 1 exabyte of data per day, about 10 times the global internet traffic (SKAFactsWeb). As the quantity, resolution and rate of astronomical images grows, astronomers will increasingly rely on fully automated calibration and analysis pipelines. It is expected that data mining techniques will play a significant role in many new astronomical discoveries (Ball & Brunner 2010).

While a data-mining algorithm can make a discovery, it still requires an astronomer to explain it. Invariably, this requires an understanding of an individual object of interest and its environment. For example, the evolutionary history of an individual galaxy is strongly affected by the density of material that surrounds it – galaxies living in isolation experience very different lives to those at the centre of a gravitationally bound cluster of many hundreds of galaxies [e.g. (Peng et al. 2010)]. As such, manual inspection of extremely high-resolution images or image sets remains an important step for verification of the automated processes (i.e. quality control) and for discovering unexpected phenomena at the limits of instrument sensitivity.

Table 2
Typical image sizes for existing and proposed astronomical cameras.

Facility	**Image Size (Megapixels)**	**Reference**
HST Advanced Camera for Surveys	16	(ACSWeb 2005)
Dark Energy Camera	520	(Mohr et al. 2012)
Subaru Hyper Suprime-Cam	870	(HyperSuprimeCamWeb 2011)

As Tables 1 and 2 demonstrate [see also Table 1 of Meade et al. (2014)], there is a clear mismatch between the resolution of the data and that of the typical desktop-based display resources used by the majority of astronomers. There is considerable support for the value of multiple displays, and more and more computer users are working in a multiple display environment. However, there remains an opportunity to make more use of non-standard display environments that could greatly enhance visual inspection, collaboration, communication and training (Fluke et al. 2006). Of particular relevance are tiled display walls and UltraHD displays.

Tiled Display Walls

TDW comprise a collection of individual, standard desktop displays linked to compute nodes that are coordinated by a head node. The displays are arranged in a tiled configuration, either as a flat or nearly flat wall, or in an arrangement that surrounds the user, such as the CAVE2 (Febretti et al. 2013) or the StarCAVE (DeFanti et al. 2009b). The combined resolution of the individual displays can produce display environments of more than 300 megapixels (OptipresenceWeb 2009), although 50-100 Megapixels is more common.

Management software such as Scalable Adaptive Graphics Environment (SAGE; (SAGEWeb 2012) provides an efficient, easy-to-use method for displaying a variety of multimedia content. SAGE also makes connecting remote walls possible, enabling improved collaboration and control of remote displays. TDWs need not be the same physical arrangement or scale to be connected. Moreover, a user at a TDW with many screens can easily and effectively collaborate with a colleague using a single screen.

While the use of commodity displays and compute nodes drives the dollar per pixel cost down, large TDWs can cost well over AU$100,000, and require considerable cluster management expertise to run the underlying infrastructure. Additionally, they are not suited for installation in a typical office area, requiring dedicated spaces to house them. Both limited physical location options and the high upfront capital costs have resulted in a relatively small number of TDWs being rolled out to Australian universities. Another challenge for TDWs is to get researchers to incorporate them into their research activity, which is often impractical if access is limited.

UltraHD displays

A cost effective intermediate step is to use UltraHD screens. These displays can be purchased off-the-shelf for less than AU$1000, though better quality devices are closer to AU$3000. With a resolution of 3840 x 2160 pixels, or four times the resolution of Full HD (1920 x 1080), a 55” display provides approximately the same pixel density as a high quality standard display, and the visual display space of four 24” desktop monitors. This also eliminates the distraction of bezels when multiple monitors are placed in a tiled configuration.

As a single display running on a single machine, no additional software is required to coordinate across machines. However, a suitable graphics card, such as a modern professional card (e.g. NVIDIA Quadro K2200), is required to drive the Ultra HD display. This often means the built-in graphics capabilities of mini, all-in-one and laptop computers is insufficient. Fortunately, any recent desktop computer capable of housing a standard PCIe graphics card should be sufficient to provide excellent performance (i.e. up to UltraHD resolution @ 60 fps).

Large-format displays in astronomy

Meade et al. (2014) undertook the first detailed study on the use of large-format displays in astronomy through a series of "visual source-finding" experiments. In these experiments, participants were presented with images created to match the resolution of the University of Melbourne's 98 Megapixel OzIPortal TDW: 15360x6400 pixels (a matrix of 24 screens with 2560x1600 pixels per screen – see Figure 1). These images included extremely high-resolution astronomical images of galaxy clusters and gaseous nebulae taken from the Hubble Space Telescope, as well as a word field comprising well-known English words.

The purpose of these experiments was to discover if participants could find objects of varying size (larger being easier) within a set time period. Here, performance referred to a participant's ability to find more and/or smaller images. Comparisons were made between performance on the TDW and a standard desktop display (with 1680 x 1050 pixel resolution). The 57 participants included both astronomer and non-astronomer groups. Overall, both groups showed better performance when using the TDW as opposed to the standard desktop display, with the astronomer group generally performing better than the non-astronomers. Of interest is the result that when working in pairs, non-astronomers performed as well as individual astronomers. The aim was to test the notion that TDWs provide better understanding for extremely large astronomy images or image sets.

The Meade et al. (2014) results were consistent with earlier studies (Ball & North 2005; Ball et al. 2007; Bi & Balakrishnan 2009; Bezerianos & Isenberg 2012; Andrews et al. 2010) suggesting physical navigation of a large image (where the participants had to move their head and/or whole body) was more effective than virtual navigation (where a mouse was used to pan and zoom the image). When searching for small objects, more effective search strategies were produced when the context of the overall image could be maintained. This was especially true when visual inspection was conducted as a collaborative exercise – but it is not always possible to get two astronomers in the same room in front of the same TDW.

Based on the approach taken in Meade et al. (2014), but without the benefit of the original study participants, we repeated the image search experiment using a LG UXD7000 UltraHD display, connected to a Dell Precision T3400 with a NVIDIA Quadro K2200 graphics card. The number of participants (n = 4 and included authors of this current work) was too small to produce a statistically significant comparison, but the results were consistent with the expectation that the UltraHD display would function better than the standard desktop display but not as well as the TDW. Observing the manner of interaction of the participants with the UltraHD display was also consistent with the Meade et al. (2014) observations that virtual navigation was only used when physical navigation techniques had been exhausted. Therefore, the increased display space while maintaining suitable pixel density improved the performance of the participants, at least within the range of display sizes tested (viz. standard 24inch display, UltraHD display, 98 megapixel TDW).

Virtual Desktop Infrastructure

An increase in resolution at the desktop for an UltraHD display requires a corresponding improvement in the graphics capabilities of the host computer. But an over-powered desktop machine may not be the solution if the full display capabilities (viz. pixels and frame rates) are not required all of the time. An emerging option is the Virtual Desktop Infrastructure (VDI) (Miller & Pegah 2007). VDI formally denotes a completely isolated virtual desktop environment, i.e. one desktop per user. The VDI desktop referred to in this paper is actually delivered using a Remote Desktop Services (RDS) model, through Citrix XenApp. The reason for this is that, at the time of writing, only Citrix XenApp was able to deliver GPU-accelerated applications and desktops capable of being scaled to the native UltraHD resolution on the client display. In order to most closely replicate the 1:1 nature of true VDI, only a single user session existed on the host VM during all remote desktop testing.

For applications that support it, the addition of a recent release graphics processing unit (GPU) card to the host server of a virtual machine provides the necessary power to generate a real-time compressed video stream at sufficient frame rate and resolution to provide an excellent desktop experience. Indeed, when combined with additional CPU cores, the power of the virtual machine can be considerably greater than a typical desktop configuration. For example, we were able to stream a Windows desktop at FullHD to a remote client consuming between ~5 and ~50 Mbps of bandwidth using a server with the following configuration:

- Dell PowerEdge R720 server, with:
    - 2x Intel(R) Xeon(R) CPU E5-2670 @ 2.60GHz (total 16 physical cores, 32 logical);
    - 192GB ECC RAM (32GB allocated to the host VM);
    - 1 TB VMFS5 storage (80 GB dedicated to the host VM); and
    - 2x NVIDIA GRID K2 with 4GB GDDR5 RAM (1 GPU dedicated to the host VM)

On gigabit Ethernet, up to 60fps with imperceptible compression is possible. Applications and desktops are streamed using XenApp 7.6, installed in a Windows Server 2008 R2 host VM. The hypervisor is VMWare vSphere 5.5. Mouse and keyboard events are transmitted to and from the virtual desktop, providing a seamless desktop experience. Furthermore, when streamed to a machine capable of driving an UltraHD display, the resolution is automatically increased accordingly and continues to run at acceptable frame rates, though some frame delay is apparent. Performance of the machine decoding the video stream also affects the display frame rate. Using VDI in this way requires a robust network to maintain the connection, but the payoff is impressive.

By combining the above server configuration with a suitably capable receiving client, we suggest that an acceptable user experience with either a FullHD or UltraHD desktop is possible over a 1Gbps network. There are several benefits of this approach:

1. Most universities opt for a regular life cycle for desktop computer infrastructure, usually around 3 to 5 years. In this model, computers are often purchased that exceed requirements in the first year, are ideal for the second year, and noticeably underpowered in the last years of the life cycle.
2. Additional processing power can be allocated as required. As e-mail, web browsing and office applications typically need relatively little processing power and memory on modern processors, fewer resources need to be consumed. Additional cores and memory can be made available when higher-than-normal processing capabilities are needed.
3. High-performance systems, such as multi-core, large memory and GPU-accelerated virtual machines can be created on demand, and shared between users.
4. Virtual machines can be left in a "powered on" state, so that they can continue processing even while the user is no longer connected.
5. Connections can be made from anywhere a suitable network is available. In many cases, a researcher's personal home network is considerably slower than that provided by their research institution. As the virtual machine would be hosted in a highly connected data centre, the network performance does not diminish, even when operated from home. This is because the user is only receiving a display stream via their home network.

The ability to use VDI effectively for remote collaboration with large-format displays depends on the existence of a suitable remote processing facility to host and serve virtual machines, and sufficient bandwidth between the local and remote facilities. In the next section, we describe the current state of both of these capabilities in Australia.

Infrastructure for Remote Collaboration

Australia's national research computing infrastructure comprises (amongst other things):

- High performance computing facilities, including the National Computational Infrastructure (NCI) in the Australian Capital Territory and the Pawsey Supercomputing Centre (Pawsey) in Western Australia;
- The internet backbone provided through the Australian Academic Research Network (AARNet); and
- A growing research-focused Cloud computing capability, most notably offered through the National e-Research Collaboration, Tools and Resources (NeCTAR) program.

For the majority of Australian Higher Education researchers transferring data beyond their home institution, bandwidth is provided by the AARNet. AARNet currently provides a 40 Gbps backbone to most Australian universities, and up to 100 Gbps in some places. Locally, many universities support connectivity at 10 Gbps throughout their campuses, although it usually falls to 1 Gbps to the desktop. At most institutions, this wired network is also supported by 802.11n wifi networks, providing up to 308 Mbps for wireless devices. The reliability of network infrastructure at a university is typically around 95% or better, inclusive of both planned and unplanned outages. Such an environment has allowed research to become a largely "online" activity, with much of the data and reference materials sourced remotely.

The NeCTAR program was established in 2011 as part of the Federal Government's Super Science Initiative. The NeCTAR Research Cloud first came online in February 2012, with the lead node established at the University of Melbourne. Since then, additional nodes have been added at Monash University, University of Tasmania, Australian National University, Queensland University of Technology, e-Research South Australia, Intersect and at Pawsey. The principal capability of the Research Cloud nodes is to provide the Australian research community with free and easy access to computational resources in the form of virtual machines (VM) offering Infrastructure-as-a-Service (IaaS) capability. VMs can be created and terminated with ease, and can be used for a multitude of research purposes.

Individual researchers working at Australian universities can be authenticated via the Australian Access Federation (AAF) and gain access to the Research Cloud dashboard. Trial resources for 30 days consisting of 2 compute cores, 8 GB of RAM and 70 GB of disk space are automatically allocated, however users can request additional merit-based, long-term resources for research projects through NeCTAR. Larger coordinated research efforts have also obtained funding and resources to establish Virtual Laboratories (VL). These typically provide the discipline specific research community with appropriate tools for the given domains. The Research Cloud VMs contribute to the underlying infrastructure for these applications. Some examples of highly successful VLs include the Genomics Virtual Laboratory (GVLWeb), Characterisation Virtual Laboratory (CVL) and the All-Sky Virtual Observatory (ASVOWeb).

Currently the Research Cloud has over 20,000 cores, with more than 17,000 cores in use. Nearly 5,000 users have registered with the Research Cloud, launching over 6,000 VM instances. The Research Cloud uses OpenStack to provide the service and the National Endpoint Status (NES) reports an availability of 99.854% for the component supporting the VMs (called Nova) over the most recent 6-month period (22nd of August, 2014 – 22nd of February, 2015). The overall performance of all components, including monitoring, storage and access security is 99.751% for the same period.

Testing the Nation's Research Bandwidth

Having established the suitability of TDWs and UltraHD displays as advanced display infrastructures for knowledge discovery, we now turn our attention to the bandwidth required to transfer high-resolution images at reasonable frame-rates. To achieve this, we performed a simple, yet instructive experiment: measuring the time taken to transfer files of known size.

The purpose of the tests was to look at how the network bandwidth and stability would affect a typical researcher trying to retrieve large datasets from remote repositories for visualization on a local high-resolution display. In an ideal situation with a direct connection between sites, a 1 GB file would travel at the maximum speed of 1 Gbps link to the desktop (since it is limited by the 1Gbs connection to the desktop machine). Adding in hops between routers and switches, and allowing for other network traffic, introduces delays and instabilities. The longer the transfer time, the greater impact of both systemic and transient effects. Therefore, transferring a single gigabyte might show acceptable transfer times, yet not provide a reliable indication of scalability or consistency of the network. As the datasets get bigger, the network stability plays a much larger role than bandwidth.

With an emphasis on high-resolution astronomical image collections, two major facilities hosting petabyte-scale astronomy data were chosen to connect to the University of Melbourne:

- NCI hosts the the Skymapper (Keller et al. 2007) dataset through the All-Sky Virtual Observatory project. This facility also houses the ANU node of the NeCTAR Research Cloud.
- Pawsey is the repository for the Australian Square Kilometer Array Pathfinder project. Compute time was provided on the Galaxy supercomputer.

These two sites also provided an opportunity to compare the well-established East coast AARNet connection with the relatively new high-speed link to Western Australia.

As we were interested in exploring the instantaneous bandwidth for transfers between NCI, Pawsey and Melbourne, we generated a set of 10 fiducial files as strings of random numbers using the Linux *dd* command:

dd if=/dev/urandom of=file-1GB.txt bs=1048576 count=1024

Here, the input parameter *bs* sets the block-size; equal to 1048576 bytes in this example. The minimum file size was 1 GB, and the maximum file size was 1000 GB, as each site had only 2 TB of storage available.

For simplicity, we performed the transfers with the *scp* protocol. While faster options such as GridFTP and Aspera do exist, they are not typically available to researchers. A short bash script initiated the transfers and the timing of the transfer results subsequently logged.

Transfer speeds

Figure 2 shows the transfer rates of files from 1 GB to 1000GB using the *scp* transfer protocol between a VM hosted on the NCI node of the NeCTAR Research Cloud and a Linux desktop computer on a 1 Gbps wired connection, connected to a 10 Gbps building switch and the 40 Gbps AARNet border router. The data centre hosting the VM at NCI in Canberra is connected via 40 Gbps border router and 10 Gbps switch to the server running the VM. Each fiducial file was transferred three times (consecutively), and the median transfer time was recorded. The transfer experiment was run continuously over a period of one week, to account for daily variations in the network. As the figure shows, the network between the sites is very stable for all transfers. However, variations are more obvious on the large file transfers as the extended duration makes them subject to greater instabilities. Overall, it appears that the NCI to University of Melbourne network is stable for any file size up to one terabyte.

Running the same experiment from Pawsey to the same desktop at the University of Melbourne showed the connecting networks are much less stable – see Figure 3. It was necessary to restart the experiment several times due to unexpected outages that caused the automatic transfers to fail. The results shown are for a single run rather than a full week, due to several transfer failures during the experiment. Even ignoring outlying data points, it is clear that the network between the Pawsey Centre and the University of Melbourne is not currently as stable as between NCI and the University of Melbourne.

There are several possible reasons to account for the increased instability between Pawsey and the University of Melbourne:

1. The network is relatively new and may not be optimally configured yet;
2. The Galaxy server, located in the Pawsey Centre, has experienced several system errors in recent months, which have contributed to the failed transfer attempts, and may also have contributed to delays even for the successful transfers;
3. The timing of the experiment may have occurred during an atypical period of network instability.

Figure 4 shows the results of transfers of a 1GB file with a five-minute sleep between transfers over the course of a week, from NCI to University of Melbourne and Pawsey Centre to University of Melbourne.

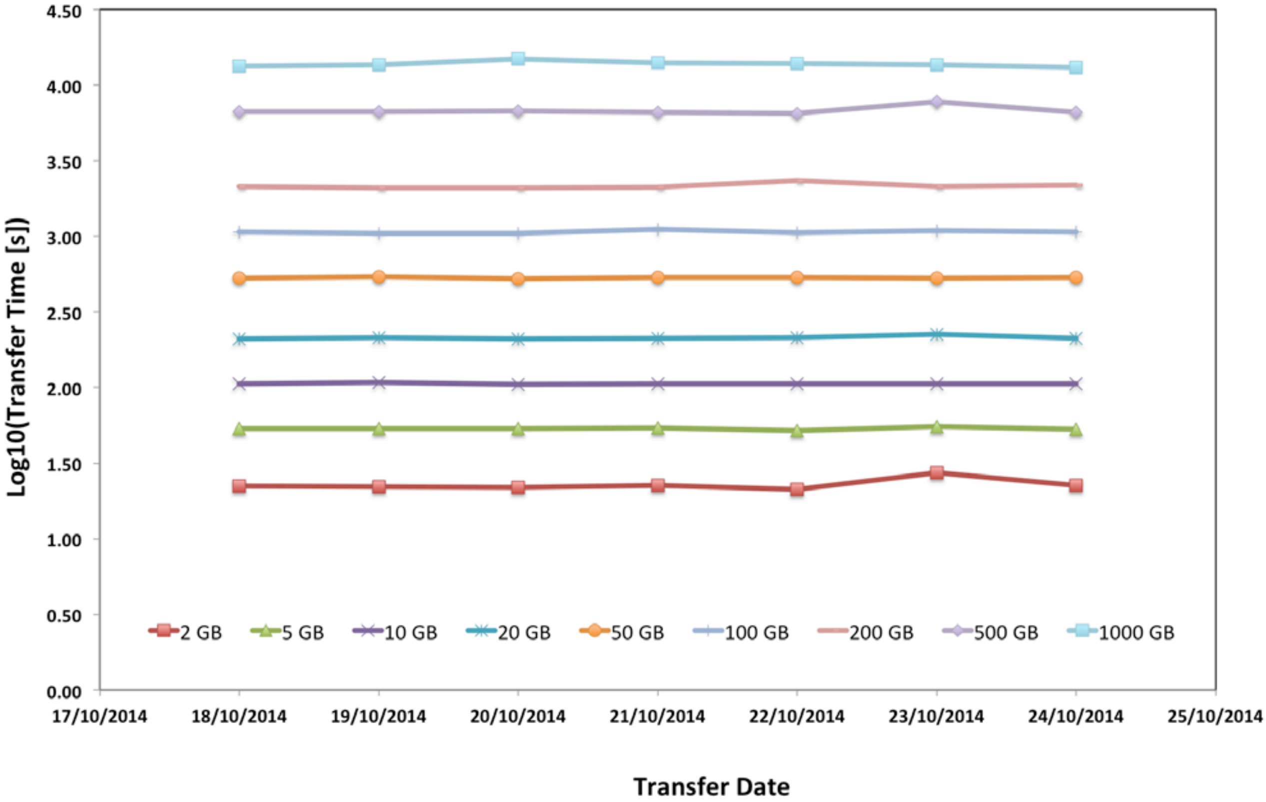

Figure 2. Daily data transfer speeds between the NCI and the University of Melbourne. There is little variability in the transfer times during the week for each of the file sizes.

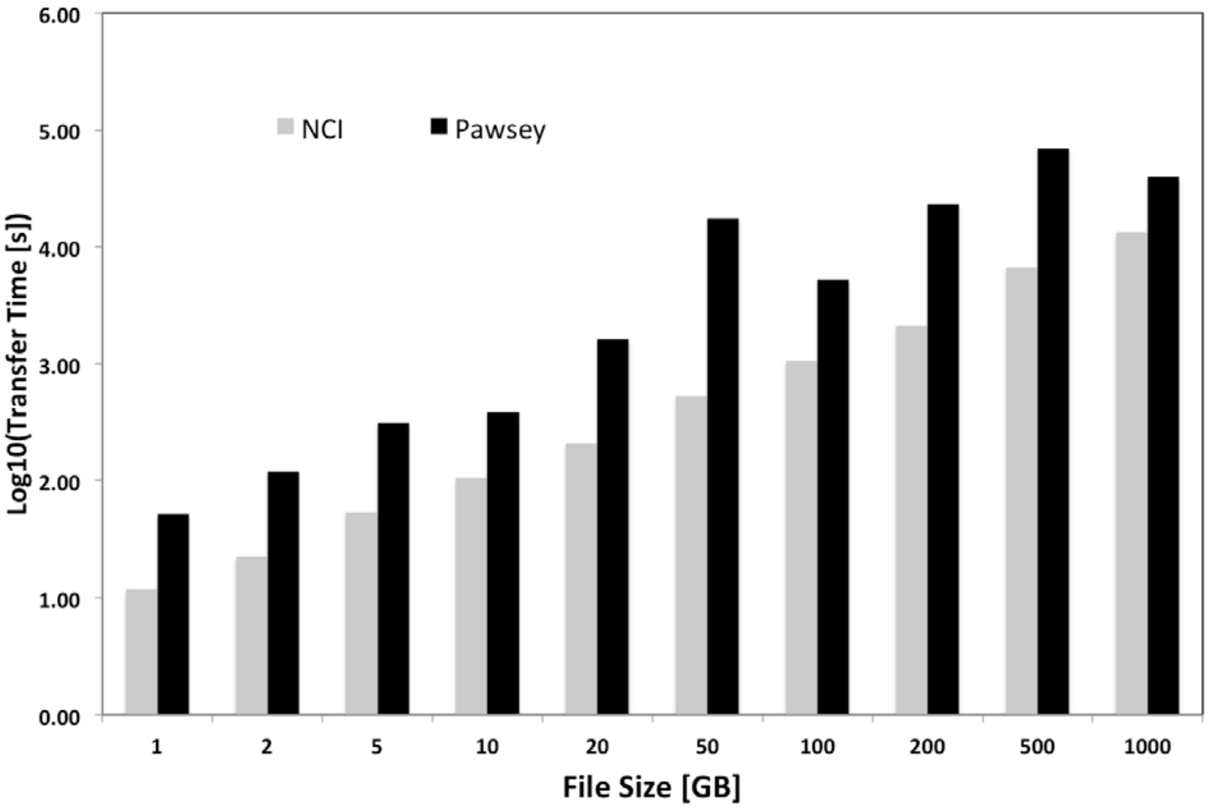

Figure 3. Data transfer times to the University of Melbourne from NCI (grey) and Pawsey (black). In all cases, transfer time from Pawsey is longer, and shows more variability.

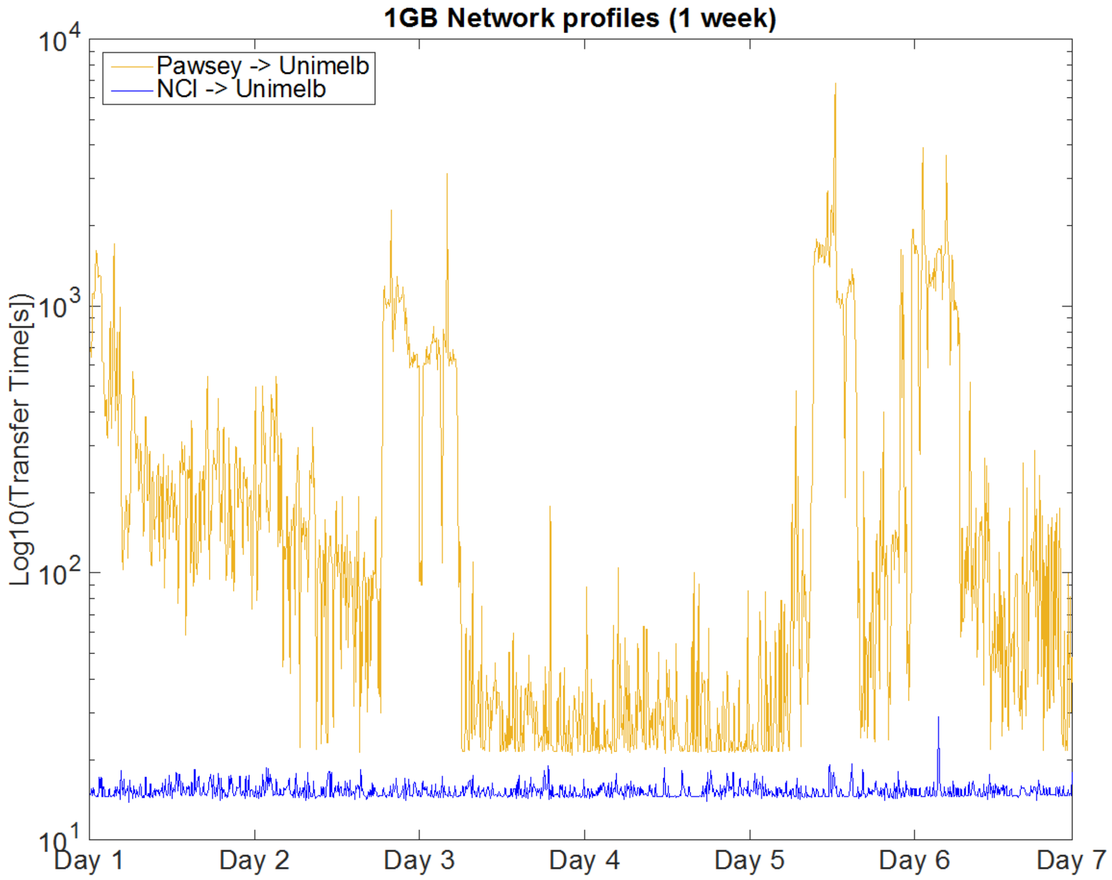

Figure 4. Network profile for 1 GB transfers to the University of Melbourne from NCI and Pawsey.

Conclusion

Combined, the network transfer results provide a compelling case for remote processing, rather than the transfer of large volumes of data to local computers. With a highly stable network such as between National Computational Infrastructure and the University of Melbourne, the growing volume of datasets quickly overwhelms bandwidth and the capabilities of a local computer, both in hard disk storage and RAM and CPU capacity. The situation becomes much more challenging when dealing with additional network instabilities such as those connecting the Pawsey Supercomputing Centre to the rest of Australia.

The stability of the network becomes critical when a fully interactive graphical user interface is being used. The NCI to University of Melbourne link is clearly capable of providing a sustained link of suitable bandwidth for VDI up to UltraHD resolution. However, the link to Pawsey is unlikely to be able to sustain a useable VDI link, and in some cases would be unable to maintain a simple X11 forward. However, this is likely to be a short to medium term issue as the demand for stable network increases and problems are overcome.

Cloud computing has become an essential tool for big data research, and is changing the way research is conducted in many disciplines. Combining computational resources in a shared pool provides far greater performance and economies of scale. But it is also essential to provide solid underlying networks to ensure a high quality user experience, especially in the case of VDI and window forwarding. Fortunately considerable effort is being invested in improving networks across Australia and to international connections. The reality of a fully remote desktop streamed to a local thin client for all researchers is very near.

Acknowledgments

This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia. This research was undertaken on the NCI National Facility in Canberra, Australia, which is supported by the Australian Commonwealth Government. We also acknowledge use of computing resources from the NeCTAR Research Cloud http://www.nectar.org.au NeCTAR is an Australian Government project conducted as part of the Super Science initiative and financed by the Education Investment Fund and National Collaborative Research Infrastructure Strategy (NCRIS). We would also like to acknowledge the assistance of Hans Bauer, Dell Computers, for providing the resources for the VDI testing, and Marko Alat (University of Melbourne) for network support.

References

ACSWeb (2005). *The Advanced Camera for Surveys (ACS)*. Retrieved from http://acs.pha.jhu.edu/

Ahrens, J., Geveci, B., & Law, C. (2005). *The Visualization Handbook, ParaView: An End-User Tool for Large Data Visualization*, ed. CD Hansen & CR Johnson (p. 717). Burlington, MA: Elsevier.

Andrews, C., Endert, A., & North, C. (2010, April). Space to think: large high-resolution displays for sensemaking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 55-64). New York: ACM.

ASVOWeb (2015). *The All-Sky Observatory | NeCTAR*. Retrieved from https://www.nectar.org.au/all-sky-observatory.

Ball, N. M., & Brunner, R. J. (2010). Data mining and machine learning in astronomy. *International Journal of Modern Physics D, 19*(07), 1049-1106.

Ball, R., & North, C. (2005, April). Effects of tiled high-resolution display on basic visualization and navigation tasks. In *CHI'05 extended abstracts on Human factors in computing systems* (pp. 1196-1199). New York: ACM.

Ball, R., North, C., & Bowman, D. A. (2007, April). Move to improve: promoting physical navigation to increase user performance with large displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 191-200). New York: ACM.

Bezerianos, A., & Isenberg, P. (2012). Perception of visual variables on tiled wall-sized displays for information visualization applications. *Visualization and Computer Graphics, IEEE Transactions on*, *18*(12), 2516-2525

Bi, X., & Balakrishnan, R. (2009, April). Comparing usage of a large high-resolution display to single or dual desktop displays for daily work. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1005-1014). New York: ACM.

Czerwinski, M., Smith, G., Regan, T., Meyers, B., Robertson, G., & Starkweather, G. (2003). Toward characterizing the productivity benefits of very large displays. In *Proceedings of INTERACT* (Vol. 3, pp. 9-16).

DeFanti, T. A., Leigh, J., Renambot, L., Jeong, B., Verlo, A., Long, L., ... & Smarr, L. (2009). The OptIPortal, a scalable visualization, storage, and computing interface device for the OptiPuter. *Future Generation Computer Systems*, *25*(2), 114-123.

DeFanti, T. A., Dawe, G., Sandin, D. J., Schulze, J. P., Otto, P., Girado, J., ... & Rao, R. (2009). The StarCAVE, a third-generation CAVE and virtual reality OptIPortal. *Future Generation Computer Systems*, *25*(2), 169-178.

Febretti, A., Nishimoto, A., Thigpen, T., Talandis, J., Long, L., Pirtle, J. D., ... & Leigh, J. (2013, March). CAVE2: a hybrid reality environment for immersive simulation and information analysis. In *IS&T/SPIE Electronic Imaging* (pp. 864903-864903). Bellingham, WA: International Society for Optics and Photonics.

Fluke, C. J., Bourke, P. D., & O'Donovan, D. (2006). Future Directions in Astronomy Visualization. *Publications of the Astronomical Society of Australia*, *23*(1), 12-24.

GVLWeb (2015). *Genomics Virtual Laboratory | NeCTAR*. Retrieved from https://www.nectar.org.au/genomics-virtual-laboratory-0

HyperSuprimeCamWeb (2011). *Hyper Suprime-Cam*. Retrieved from http://www.naoj.org/Projects/HSC/index.html

Keller, S. C., Schmidt, B. P., Bessell, M. S., Conroy, P. G., Francis, P., Granlund, A., ... & Waterson, M. F. (2007). The SkyMapper telescope and the southern sky survey. *Publications of the Astronomical Society of Australia*, *24*(1), 1-12.

Meade, B. F., Fluke, C. J., Manos, S., & Sinnott, R. O. (2014). Are tiled display walls needed for astronomy?. *Publications of the Astronomical Society of Australia*, *31*, e033.

Miller, K., & Pegah, M. (2007, October). Virtualization: virtually at the desktop. In *Proceedings of the 35th annual ACM SIGUCCS fall conference* (pp. 255-260). New York: ACM.

Mohr, J. J., Armstrong, R., Bertin, E., Daues, G. E., Desai, S., Gower, M., ... & Yanny, B. (2012). The Dark Energy Survey Data Processing and Calibration System. *arXiv preprint arXiv:1207.3189*.

Ni, T., Schmidt, G. S., Staadt, O. G., Livingston, M. A., Ball, R., & May, R. (2006, March). A survey of large high-resolution display technologies, techniques, and applications. In *Virtual Reality Conference, 2006* (pp. 223-236). IEEE.

OptipresenceWeb (2009). *Research Projects: OptIPresence Tele-Immersion Testbed – Gravity*. Retrieved from http://vis.ucsd.edu/mediawiki/index.php/Research_Projects:_OptIPresence_Tele-Immersion_Testbed

Peng, Y. J., Lilly, S. J., Kovač, K., Bolzonella, M., Pozzetti, L., Renzini, A., ... & Cassata, P. (2010). Mass and Environment as Drivers of Galaxy Evolution in SDSS and zCOSMOS and the Origin of the Schechter Function. *The Astrophysical Journal*, *721*(1), 193–221.

Quinn, P., Axelrod, T., Bird, I., Dodson, R., Szalay, A., & Wicenec, A. (2015). Delivering SKA Science. *arXiv preprint arXiv:1501.05367*.

Robertson, G., Czerwinski, M., Baudisch, P., Meyers, B., Robbins, D., Smith, G., & Tan, D. (2005). The large-display user experience. *Computer Graphics and Applications, IEEE*, *25*(4), 44-51.

SAGEWeb (2012). SAGE: *Scalable Adaptive Graphics Environment*. Retrieved from http://www.sagecommons.org/

SKAFactsWeb (2015). *Amazing facts - SKA Telescope*. Retrieved from https://www.skatelescope.org/amazingfacts/

Appendix A

Traceroute results for NCI to University of Melbourne Desktop:

traceroute to 128.250.7.99 (128.250.7.99), 30 hops max, 60 byte packets

1 ncihpchub-vlan-256.nci.org.au (130.56.248.4) 0.364 ms 0.314 ms 0.290 ms

2 182.255.121.17 (182.255.121.17) 0.309 ms 0.285 ms 0.293 ms

3 et-5-3-0.pe1.crlt.vic.aarnet.net.au (113.197.15.22) 7.744 ms 7.726 ms 7.703 ms

4 ae9.bb1.b.mel.aarnet.net.au (113.197.15.97) 7.862 ms 7.832 ms 7.813 ms

5 tengigabitethernet2-1.er2.unimelb.cpe.aarnet.net.au (202.158.200.99) 7.791 ms 7.772 ms 7.844 ms

6 gw1.er2.unimelb.cpe.aarnet.net.au (202.158.206.162) 20.547 ms * 11.707 ms

7 * * *

8 * * *

9 * * *

10 * * *

11 128.250.7.66 (128.250.7.66) 8.236 ms 8.360 ms 8.338 ms

12 128.250.7.99 (128.250.7.99) 8.281 ms * 8.239 ms

Traceroute results for Pawsey Centre to University of Melbourne Desktop:

traceroute to 128.250.7.99 (128.250.7.99), 30 hops max, 60 byte packets

1 146.118.80.1 (146.118.80.1) 0.346 ms   0.281 ms   0.281 ms

2 * * *

3 146.118.1.89 (146.118.1.89) 1.005 ms   0.919 ms   0.887 ms

4 ivec-bdr1-te1-4.ivec.org (202.8.32.33) 1.620 ms   1.452 ms   2.574 ms

5 wa-bdr1-te4-4.gw.csiro.au (130.116.129.73) 1.685 ms   1.394 ms   1.408 ms

6 tengigabitethernet2-2.er2.csiro.cpe.aarnet.net.au (202.158.198.233) 1.467 ms   1.391 ms 1.411 ms

7 ge-4-0-0.bb1.b.per.aarnet.net.au (202.158.198.49) 1.442 ms   1.416 ms   1.428 ms

8 ge-6-0-0.bb1.a.per.aarnet.net.au (202.158.194.1) 1.794 ms   1.778 ms   1.715 ms

9 ge-4-0-0.bb1.a.adl.aarnet.net.au (202.158.194.8) 28.008 ms   27.884 ms   27.854 ms

10 so-0-1-0.bb1.a.mel.aarnet.net.au (202.158.194.18) 36.885 ms   36.948 ms   36.911 ms

11 xe-0-0-0.er1.unimelb.cpe.aarnet.net.au (202.158.210.26) 36.919 ms   36.917 ms   36.959 ms

12 gw1.er1.unimelb.cpe.aarnet.net.au (202.158.200.250) 37.015 ms   37.135 ms   37.072 ms

13 * * *

14 * * *

15 * * *

16 128.250.7.66 (128.250.7.66) 37.645 ms   37.434 ms   38.935 ms

17 128.250.7.99 (128.250.7.99) 37.903 ms   37.319 ms   37.330 ms

Corresponding author: Bernard Meade, bmeade@unimelb.edu.au

Please cite as: Meade, B., Fluke, C., Sinnott, R., Manos, S., Killeen, N., Mignone, P., & Wang, M. (2015, May). *Seeing the Big Picture: A Digital Desktop for Researchers*. Paper presented at the THETA 2015 Conference, Gold Coast, Australia.

Note: All published papers are refereed, having undergone a single-blind peer-review process.