



CLOUD COMPUTING: Opportunities and Challenges for Australia

**REPORT OF A STUDY BY THE AUSTRALIAN
ACADEMY OF TECHNOLOGICAL SCIENCES
AND ENGINEERING (ATSE)**



CLOUD COMPUTING: Opportunities and Challenges for Australia

This report discusses cloud use by government, business and universities – both overseas and in Australia – and reveals there are valuable opportunities for Australia in cloud computing: for government, researchers and business. But the Australian Government needs to ensure that these opportunities are grasped and unnecessary barriers removed.

SEPTEMBER 2010

CLOUD COMPUTING: OPPORTUNITIES AND CHALLENGES FOR AUSTRALIA

Report of a Study by the Australian Academy of Technological Sciences and Engineering (ATSE)

© The Australian Academy of Technological Sciences and Engineering (ATSE)

ISBN 978 1 921388 15 6

This work is copyright. Apart from any use permitted under the Copyright Act 1968, no part of it may be reproduced by any process without written permission from the publisher. Requests and inquiries concerning reproduction rights should be directed to the publisher.

Publisher

The Australian Academy of Technological Sciences and Engineering (ATSE)

Level 1, 1 Bowen Crescent

Melbourne, Victoria 3004

GPO Box 4055, Melbourne, Victoria 3004)

Telephone +613/03 9864 0900

Facsimile +613/03 9864 0930

Website www.atse.org.au

Principal author: **Dr J Craig Mudge FTSE**

This work is also available as a PDF document on the ATSE website, www.atse.org.au

This project was funded, in part, by CSIRO and National ICT Australia (NICTA)

Design and production: Coretext, www.coretext.com.au

Printing: FSG Print Management, www.fsg.com.au

Date of Publication: September 2010

Cover : iStockphoto

Executive Summary

Cloud computing provides a means of accessing a shared pool of configurable computing resources (including networks, servers, storage applications and services) that can be rapidly provided, used and released with minimal effort on the part of users or service providers. There are different types of clouds, ranging from those that are publicly accessible to private clouds with restricted access. These clouds can utilise software, platforms and infrastructure services as needed. The world cloud market is estimated to be currently worth more than US\$20 billion and is expected to grow rapidly.

Cloud computing offers important advantages:

- users pay per use for services and infrastructure, reducing the requirement for capital investment;
- cloud computing is scalable – when additional resources are needed they can be accessed;
- cloud computing software facilitates the manipulation of large databases;
- the increased scale of processors and data storage devices provides economies of scale and energy use; and
- cloud computing can enable entirely new innovative business services.

Governments, businesses and researchers can all benefit from the adoption of cloud computing services. Governments can use the advantages of cloud computing to provide services more efficiently to a broad range of customers.

Governments in other countries are moving quickly to ensure the rapid adoption of cloud computing. In the USA, the Obama administration has launched projects to identify services and solutions that can use cloud computing. A number of US Government agencies, including the National Aeronautics and Space Administration (NASA), the General Services Administration and the Department of Defence are taking initiatives in cloud computing.

The UK Government is adopting cloud computing to reduce its administrative costs. Other examples of the government use of cloud computing range from the census to the management of health services contracts. Japan has established a Government cloud to enable various ministries to consolidate hardware and create shared platforms. Sweden, France and Spain are among the leaders in adopting cloud computing in Europe.

Businesses can develop new services based on cloud computing as well as using the cloud to manage data-intensive activities more efficiently. The major reduction in capital costs that cloud computing provides makes it attractive to small and medium sized enterprises (SMEs) with limited access to capital.

An example of a notable success in the use of cloud computing to create new business opportunities is provided by Animoto, which enables customers to upload images and music and automatically creates customised web-based video presentations. Another example, which illustrates the cost-saving potential of the cloud, is the *New York Times*' use of the Amazon Web Service to create an archive of published material from 1851 to 1980, accomplished in 24 hours at a cost of only US\$240.

Many areas of research are becoming increasingly data-intensive. Cloud computing facilitates the efficient management and use of very large databases. It also greatly reduces the cost of computation when segmentation of the task and parallel processing are possible. One example of an application of cloud computing that offers new possibilities for researchers is in environmental monitoring, where large numbers of sensors can, over time, provide large amounts of data which needs to be collected and

analysed. Another example, from India, is the use of the cloud to compare protein sequences in order to determine evolutionary linkages and predict molecular structures.

CONCLUSIONS

Governments have important roles to play in encouraging the adoption of cloud computing and facilitating the use of cloud-based services by researchers and the business sector. For example, in the USA, the National Science Foundation (NSF) promotes the use of cloud services by researchers, though a program launched in 2008. Governments in the United Kingdom, Japan and Sweden are also encouraging and facilitating the use of cloud computing services by researchers and business.

The use of cloud services raises new issues in regard to privacy, security, trust, data transfer capacity and lock-in with service providers. Privacy legislation which pre-dates the development of cloud-based services needs to be reviewed. Trust arrangements between service providers and users of cloud services may need new and additional elements to cover all critical interactions. Service providers need to ensure that data is not lost. Data transfer capacity will need to be enhanced and the cost reduced. These issues can all be addressed and managed, but some changes to laws and regulations may be required.

It will be particularly important for government to protect the interests of consumers and SMEs. Both should be able to seek redress for failures on the part of cloud service providers. For SMEs, the cloud offers exciting opportunities, but they will need government assistance to manage the risks.

Network pricing will also be important in determining the adoption of cloud based services in Australia. At present Australia's broadband prices are significantly higher than the UK – the OECD leader.

This report discusses a number of examples of cloud use by government, business and universities – both overseas and in Australia, and reveals that this technology has much to offer. While Australia does not yet have the extensive broadband links, large-scale data centres and requisite regulatory environment, these are possible in the near future. There are valuable opportunities for Australia in cloud computing – for government, researchers and business. However, action by the Commonwealth Government is needed to ensure that these opportunities are grasped and unnecessary barriers removed.

Recommendations

RECOMMENDATION 1

The Commonwealth Government should take a whole-of-government approach to new security and privacy issues arising from the use of cloud computing, by establishing a new taskforce to review the adequacy of current legislation and identify what steps need to be taken to ensure a supportive regulatory environment.

RECOMMENDATION 2

The Commonwealth Government's Commercialisation Australia Program, which assists researchers, entrepreneurs and innovative companies to convert ideas into successful commercial ventures, should actively encourage new businesses that are cloud-focused in their internet-scale applications.

RECOMMENDATION 3

The Commonwealth Department of Broadband Communications and the Digital Economy should review broadband pricing and policies of commercial and other providers to ensure that unnecessary impediments to the uptake of Cloud Computing are minimised.

RECOMMENDATION 4

Australian universities should expand their courses to build knowledge and skills in cloud computing. They should seek the support of cloud service providers to offer short courses for industry, government and university researchers on cloud computing, including its data-parallel programming models.

RECOMMENDATION 5

The Commonwealth Government should create and fund an Australian equivalent of the NSF's Cluster Exploratory Program CluE and the NSF-Microsoft Program to actively encourage the use of cloud computing. This could be administered by the Australian Research Council.

RECOMMENDATION 6

The National Research Infrastructure Council (NRIC) should refine its investment plans to reflect the benefits that cloud computing can provide.

RECOMMENDATION 7

The Commonwealth Government should ensure that proposals for research data storage using Super Science funds have evaluated cloud computing services.

Acknowledgements

The Academy of Technological Science and Engineering (ATSE) is most grateful for the contributions made by the principal author of this report, Dr Craig Mudge FTSE, and members of the ATSE Steering Committee established to oversee its finalisation (the Steering Committee was formed following the completion of a draft Working Group report). Membership of the Steering Committee was:

- Mr Peter Laver FTSE (Chair);
- Dr Vaughan Beck FTSE;
- Dr John Bell FTSE;
- Professor Michael Manton FTSE;
- Dr J Craig Mudge FTSE;
- Dr John O'Callaghan FTSE; and
- Professor Robin Stanton FTSE.

Dr John Bell FTSE contributed to the redrafting and editing of this report and Dr Vaughan Beck FTSE, ATSE's Executive Director – Technical, was the project manager.

ATSE thanks the Working Group formed to prepare a draft report and for inputs received from a variety of sources during the course of the project. ATSE is also most grateful to CSIRO and NICTA for providing financial support for the project, which has enabled travel as well as covering a student prize, a subscription to a wiki and other expenses.

To obtain more diversity, experts from outside the Academy were invited to join the Working Group on Cloud Computing at Peta Scale (the Working Group). All members of the Working Group were volunteers who offered their time, and in some cases, resources from their institutions. The Working Group members were:

- Dr J Craig Mudge FTSE, (Chair), Pacific Challenge and the University of Adelaide;
- Professor David Abramson, Monash University;
- Mr Gordon Bell FTSE, Microsoft Research;
- Associate Professor Jim Hogan, Queensland University of Technology;
- Dr Anna Liu, NICTA and the University of NSW;
- Mr Alan Noble, Google Engineering Director, Australia;
- Professor Robin Stanton FTSE, Pro-Vice Chancellor, the Australian National University;
- Dr Andrew Wendelborn, University of Adelaide;
- Mr Glenn Wightwick, IBM Australia Development Laboratory;
- Professor Tony Williams, ARCS, Platforms for Collaboration;
- Dr Darrell Williamson, CSIRO eResearch;
- Dr John Zic, CSIRO ICT Centre; and
- Professor Albert Zomaya, the University of Sydney.

In addition to regular meetings, some Working Group members undertook several small projects including measuring latency in using overseas clouds, extending ARCS grid capability to include Amazon's Elastic Compute Cloud (EC2), trialling an accountability mechanism to improve trust in outsourced data management, and an eHealth project. One university research laboratory located its source-code control system SVN in Amazon's storage service.

Working group members from QUT, Monash, and CSIRO hosted 2009-10 summer intern projects. Projects in machine learning for profiling, a cloud-based scanner on a smart phone, and accountability as

CLOUD COMPUTING

a cloud service were selected through a competition for student project proposals that drew 23 entries. At the end of the summer, the students submitted their results of their projects in a further competition. The winner was Jinhui Yao of Sydney University with a project in an important area of security, “Accountability as a service for the cloud” hosted by CSIRO ICT Centre. These activities of the Working Group informed the understanding, and potential use, of cloud computing for Australia.

Contents

EXECUTIVE SUMMARY	i
RECOMMENDATIONS	iii
ACKNOWLEDGEMENTS	v
1. INTRODUCTION	1
1.1 Project goals and cloud definitions	1
1.2 The definition of cloud computing	1
2. CLOUD COMPUTING: CHARACTERISTICS, APPLICATIONS & BARRIERS	3
2.1 Characteristics	3
2.2 Cloud service provider features	4
2.3 MapReduce – a new data-parallel programming model	6
2.4 The cloud as a data aggregator	7
2.5 Barriers to the uptake of cloud computing	9
3. ROLES AND OPPORTUNITIES FOR GOVERNMENT	13
3.1 The roles of Government in cloud computing	13
4. OPPORTUNITIES FOR BUSINESS	17
4.1 Examples of cloud use by business	18
4.2 Opportunities for Australia	19
5. OPPORTUNITIES FOR SUPPORTING UNIVERSITY RESEARCH	23
5.1 An emerging public-private research infrastructure services sector.	23
5.2 Innovations in data-intensive computing	24
5.3 Funding strategy for university ICT research infrastructure	25
5.4 Acquisition of computer and storage infrastructure	25
ABBREVIATIONS AND ACRONYMS	29
BIBLIOGRAPHY	31

1 Introduction

1.1 PROJECT GOALS AND CLOUD DEFINITIONS

Cloud computing at internet scale is an emerging technology of importance to Australia's information and communications technology capacity and the competitiveness of its innovation capability. For this reason, ATSE established an informal Working Group to report on relevant aspects of this disruptive technology.

The Working Group's goals were to:

- clearly communicate the essence of cloud computing;
- identify the benefits, costs, risks and opportunities provided by cloud computing at internet scale;
- examine current barriers such as lack of interoperability between different clouds, geo-political constraints on moving data globally, and perceptions on limitations in security and privacy; and
- comment on impediments to the adoption of cloud computing in Australia.

The study was limited to cloud computing as an applications enabler for Australian government, business, and scientific research, as opposed to an area of research in computer science in its own right.

1.2 THE DEFINITION OF CLOUD COMPUTING

There are many definitions of cloud computing. In 2009 the US National Institute of Standards and Technology (NIST) Information Technology Laboratory developed a considered and well-written definition distilled from a number of perspectives.

NIST Definition of Cloud Computing

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models. The emerging cloud abstracts infrastructure complexities of servers, applications, data, and heterogeneous platforms.

Source: Mell and Grance 2009

The NIST definition identifies essential characteristics, service models and deployment models (Table 1.1).

Table 1.1 Summary of cloud computing features

Essential characteristics	Service models	Deployment models
On-demand self-service	Software as a service (SaaS)	Private cloud
Broad network access	Platform as a service (PaaS)	Community cloud
Resource pooling	Infrastructure as a service (IaaS)	Public cloud
Rapid elasticity		Hybrid cloud
Measured Service		

Source: Mell and Grance 2009

The deployment models are of particular importance (Table 1.2). The different types of cloud are described in Table 1.3.

The NIST definition arguably comes closest to capturing in a minimal number of words all of the essential ideas of cloud computing. Some other definitions below are more industry or commercially oriented. For example, Wikipedia states that "Cloud computing describes a new supplement, consumption and delivery model for IT services based on the internet, and it typically involves the provision of dynamically scalable

and often virtualised resources as a service over the internet.”¹ Others have suggested that cloud computing is a general term for anything that involves delivering hosted services over the internet.

Table 1.2 Categories of Cloud Services

Category	Description
Software as a Service (SaaS)	The capability provided to the consumer is to use the provider’s applications running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a web browser (e.g. web-based email). The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.
Platform as a Service (PaaS)	The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications, created using programming languages and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations.
Infrastructure as a Service (IaaS)	The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components (e.g. host firewalls).

Source: Wyld, 2010)

Table 1.3 Types of clouds

Cloud type	Description
Private cloud	The cloud infrastructure is operated solely for an organisation. It may be managed by the organisation or a third party and may exist on premise or off premise.
Community cloud	The cloud infrastructure is shared by several organisations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be managed by the organisations or a third party and may exist on premise or off premise.
Public cloud	The cloud infrastructure is made available to the general public or a large industry group and is owned by an organisation selling cloud services.
Hybrid cloud	The cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardised or proprietary technology that enables data and application portability (e.g. cloud bursting for load-balancing between clouds).

Source: Wyld, 2010

The NIST definition takes a broader view, oriented towards the notion of transparent access to resources, some of which may be services, rather than simply adopting the notion of service provision. A more service-oriented approach is often found in SaaS – this delivery of applications over the internet:

- offers significant cost savings because resources in massive warehouse-sized data centres are pooled at scale, built from low-cost commodity chips and disks², and share overheads of cooling, refrigeration, physical security, and backup power,
- is presented as a utility with a matching business model, namely pay-per-use; and
- allows a new data-parallel programming framework.

Estimates of the size of the cloud computing market have been provided by a number of leading IT consulting firms. One recent report estimates that the cloud market was worth US\$20.3 billion in 2009 is anticipated to reach US\$100.4 billion by 2016 (Wintergreen Research, 2010). A recent survey of 895 technology experts and stakeholders by the Pew Research Centre has found that they expect that they will live mostly in the cloud by 2020 (Pew, 2010).

¹ http://en.wikipedia.org/wiki/Cloud_computing

² The systems involved are using innovative software to create reliable, available, and scalable web services.

2 Cloud Computing: Characteristics, Applications and Barriers

Every Google or Bing search uses cloud computing. The internet, accessed via a web browser, is used to access storage and computers located in massive warehouse-size data centres, currently all located outside Australia. More generally, cloud computing allows an organisation to outsource the management and location of its IT equipment to such facilities, on a scale and usability that has not been available before.

2.1 CHARACTERISTICS

The Working Group has identified four major drivers of cloud computing:

a) 'Everything as a service'

Service-oriented architectures define standard interfaces and protocols (e.g. for discovery, quality, and data transfer) to allow developers to utilise information tools and functions as services that users can access without knowledge of, or control over, their internal workings. Services in the cloud can be set up to form business workflows. The advent of XML in the nineties and standards, such as BPEL in the early 2000s, fuelled this growth.

b) Availability of broadband

Both the reach and capacity of Internet Protocol networks are drivers of cloud computing. Moreover, as capacity grows, more demanding types of data can be used in commerce, entertainment, government, and business. In the early days of the web, data and an occasional image were retrieved by web browsers. Today users routinely download music and short video clips. However, full-length movies at high quality are still beyond most broadband connections.

c) Warehouse-size data centres

Another driver of cloud computing is an evolution from small, distributed, data-oriented computing centres (e.g. 1000 node³, one petabyte⁴) to more cost-effective, very large scale commercial cloud services (e.g. 100,000 node, 100 petabyte). This is likely to continue.

d) Energy efficiency

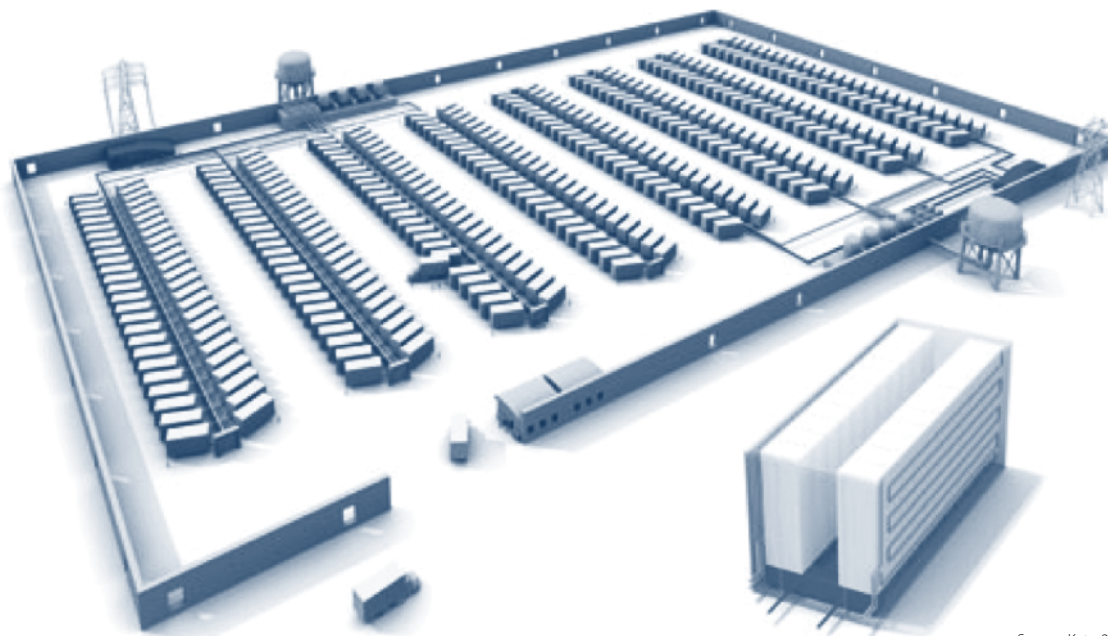
In some data centres, only 30 per cent of the electric power is used by the IT equipment, with the remaining 70 per cent going to cooling, backup, etc. The last few years has seen rapid learning in warehouse-size data centres on the design of cooling systems (Google, 2009), environment monitoring (Liu, 2007), and backup with the result that power utility efficiency is improving.

Figure 2.1 shows in diagrammatic form, how these warehouse size data centres can be configured. Figure 2.2 shows two Google facilities.

3 An example of a node in this context is the Australian Antarctic Data Centre (AADC), which was established in 1996 to fulfil Australia's role in addressing Article III.1.c of the Antarctic Treaty which states that scientific observations and results from Antarctica shall be exchanged and made freely available. The AADC automatically captures metadata from approved research projects and is updated annually in conjunction with progress reports. Data must be submitted to the AADC within two years of collection and is made freely available on the Internet.

4 A petabyte is 1000 terabytes, or 10^{15} bytes.

Figure 2.1 Container-packaged computers and storage



Source: Katz, 2009

Figure 2.2 Google's data centres at The Dalles, Oregon and Mountain View, California



Sources: Web2.Ohhh and Google.

2.2 CLOUD SERVICE PROVIDER FEATURES

The major commercial suppliers of cloud computing services are Amazon Web Services, Google, Microsoft Azure, Yahoo, and Salesforce.com. The essential properties of their services are:

- software as a service – applications are delivered over the internet;
- significant cost savings because resources in massive warehouse-sized data centres are pooled at scale, built from low-cost commodity chips and disks⁵, and share the overhead of cooling, refrigeration, physical security, and backup power;
- presented as a utility with a matching business model, namely pay-per-use; and
- a new data-parallel programming framework.

Pay per use

The pay-per-use property has business value to start-up companies, because they can replace their scarcest resource, capital, with operating expenses. Of course, this is also of value to established businesses starting new divisions, and government departments wishing to save costs.

⁵ The systems are running innovative software to create arbitrarily reliable, available, and scalable Web services.

In December 2009, Amazon Web Services introduced a variant on pay-per-use. ‘Spot instances’ enable bidders to purchase unused Amazon EC2 capacity. ‘Instances’ are charge a Spot Price set by Amazon. Spot instances are a new, potentially low cost pay-per-use approach to accessing computing capacity where the user has some time flexibility. It is ideally suited to work with soft deadline requirements. Spot instance price fluctuates with the market demand. When demand is low, the spot instance price is low. When demand is higher, the price will increase as the market functions.

Utility

The elasticity of the pooled resources was illustrated by Google to satisfy a rapidly increasing demand on the Web site of the Victorian Country Fire Authority during the catastrophic bushfires in February 2009. Within 10 days, their copy of the Authority’s page was among the top 10 trafficked sites in the world.

It is also a property that allows Australian university researchers to acquire overflow resources when their “free” infrastructure cannot meet their demand.

Massive scale leads to lower cost

When tens, sometimes a few hundreds, of thousands of cheap personal computer-type processors and commodity disks are housed in warehouse-size data centres, scale means low cost, sometimes one-fifth the cost of the usual data centre (Armbrust et al, 2010). Sharing the overhead of cooling, refrigeration, power backup, and security also reduces the cost of computing and data storage. The warehouse-size systems are running innovative software to create arbitrarily reliable, available, and scalable Web services.

Shipping containers are now being use to house racks of servers, together with cooling and power for the container. Modular data centres are now being built from these containers (see Figure 1.1). According to the main cloud service providers, the trend to larger and larger numbers of computers in one centre will continue. The results for Google are illustrated in Table 2.1.

Table 2.1 Google operations – change over the 10 years 1999 to 2009

Parameter	Growth factor	Comment on change
Numbers of documents	100 times	From about 70 million to many billion
Queries processed per day	1000 times	
Information in index per document	3 times	
Update latency ⁶	1000 times	From months to minutes
Average query latency	5 times	From < 1 second to < 0.2 seconds
More faster machines	1000 times	

Source: Based on Dean 2009

CSIRO is in the process of consolidating its data-centre footprint to reduce its operating costs and to make available a private cloud whereby it can deliver Software-as-a-Service⁷ (SaaS) and Infrastructure-as-a-Service (IaaS) for CSIRO science as part of its eResearch agenda.

The adoption of the cloud model will result in more cost-effective ICT services with scope for the creation of new businesses, and new business models. There is potential to do things that were not previously possible, such as drawing on the entire resources of the web (images, books, music, videos, all scientific data, etc.). In the view of the Working Group, enabling both existing and new Australian firms to compete on ideas, rather than the size of available resources, is more than the lowering of the cost of ICT services.

⁶ Latency is a measure of the time delay experienced in a system.

⁷ SaaS is software that is deployed over the Internet and/or is deployed to run behind a firewall in a local area network. It is used by businesses for Customer Relationship Management, human resource management and other tasks.

2.3 MapReduce: A NEW DATA-PARALLEL PROGRAMMING MODEL

MapReduce is a software framework introduced by Google to support distributed computing on large data sets on clusters of computers. This framework facilitates the processing of very large datasets where there is scope for distributed processing. MapReduce involves two steps:

- a 'Map' step where a master node takes the input, chops it up into smaller sub-problems, and distributes these to nodes; and
- a 'Reduce' step where the master node takes the responses and combines them to provide the answer to the problem it was asked to solve.

Of equal or possibly greater importance is the runtime library⁸ in its implementations, first at Google (Dean and Ghemawat, 2004) and now in an open source version called Hadoop. This library takes care of several intricate and complex tasks, such as allocating a job across a large number of commodity computers⁹ and disks in warehouse-size data centres. The runtime library also manages fault tolerance and inter-computer communication. Microsoft Research's version of MapReduce is called Dryad (Isard *et al.*, 2007).

Commenting on the use of MapReduce in 2007, Google experts said that:

"Many real world tasks are expressible in this model. Programs written in this functional style are automatically conducted in parallel and executed on a large cluster of commodity machines. Our implementation of MapReduce runs on a large cluster of commodity machines and is highly scalable: a typical MapReduce computation processes many terabytes of data on thousands of machines. Programmers find the system easy to use: hundreds of MapReduce programs have been implemented and upwards of one thousand MapReduce jobs are executed on Google's clusters every day." (Dean and Ghemawat, 2004).

Historically, for computationally intensive applications there have been two main approaches taken to exploiting data parallelism. The first was to build compilers which automatically discover parallelism in sequential programs. The second required the program developer to explicitly expose the data dependencies of a computation. Message Passing Interface, developed in the 1990s, has become the most popular model used in the second approach, whereas parallelising compilers have never been very successful. MapReduce follows the second approach.

Hadoop is now widely used in the USA. Its attractiveness jumped when a couple of leading Google engineers left to form Cloudera, a company providing Hadoop support and training. Cloudera is to Hadoop what Red Hat was to Linux a decade ago.

The programming languages involved are stylistically quite different from conventional languages. As a consequence, in spite of Dean's optimism, they face barriers to adoption. MapReduce is a functional language, like Lisp and APL developed in the 1950s, and users should therefore be comfortable with the underlying concepts.

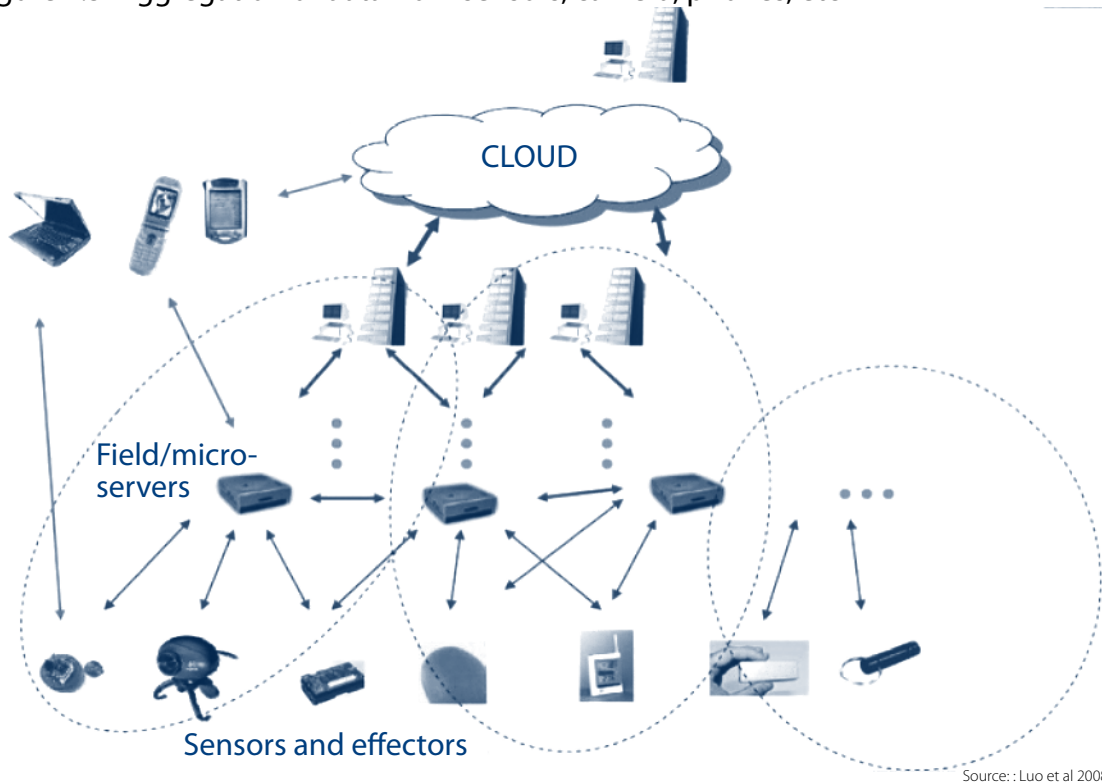
MapReduce is not a silver bullet for solving all the problems of scale. However it is a good technique for calculations involving large sets of data when there is scope to process small pieces of that data in parallel. An improved understanding of the MapReduce model will enable Australian users to apply it to solving multi-core problems,¹⁰ as shown by recent work at Stanford (Ranger *et al.*, 2007).

⁸ A program library used by a compiler to implement functions built into a programming language, during the execution (runtime) of a computer program.

⁹ Commodity computers are computer systems manufactured by multiple vendors, incorporating components based on open standards. Based on commodity components, these computers are less expensive and there is less differentiation between vendor's products.

¹⁰ The multi-core problem is significant in microprocessor design, and hence in computer hardware and applications in general. Power constraints have led microprocessor designers to place multiple cores (processors) on each chip. Multiple processors require parallelism in the workloads presented to these chips. Microsoft Research and Intel have invested tens of millions of dollar to fund university research on this problem, with UC Berkeley's new Parallel Computing Laboratory a major recipient of the funding.

Figure 2.3 Aggregation of data from sensors, camera, phones, etc



The Google File System

The Google File System (Ghemawat *et al*, 2003) contributes to the essential properties (fault tolerance, scalable distributed file system, clustered storage, huge files) of Google's approach to large distributed data-intensive applications. Because low-cost commodity chips and discs populate massive data centres, component failures are the norm rather than the exception. The Google File System is available on Amazon in its Elastic MapReduce service. A widely used open source version is the Hadoop Distributed File System (HDFS).

These file systems play an important role in one of the requirements of data-intensive computing, namely to move computation to the data, as the Hadoop job scheduler co-locates computation with data where possible.

2.4 THE CLOUD AS A DATA AGGREGATOR

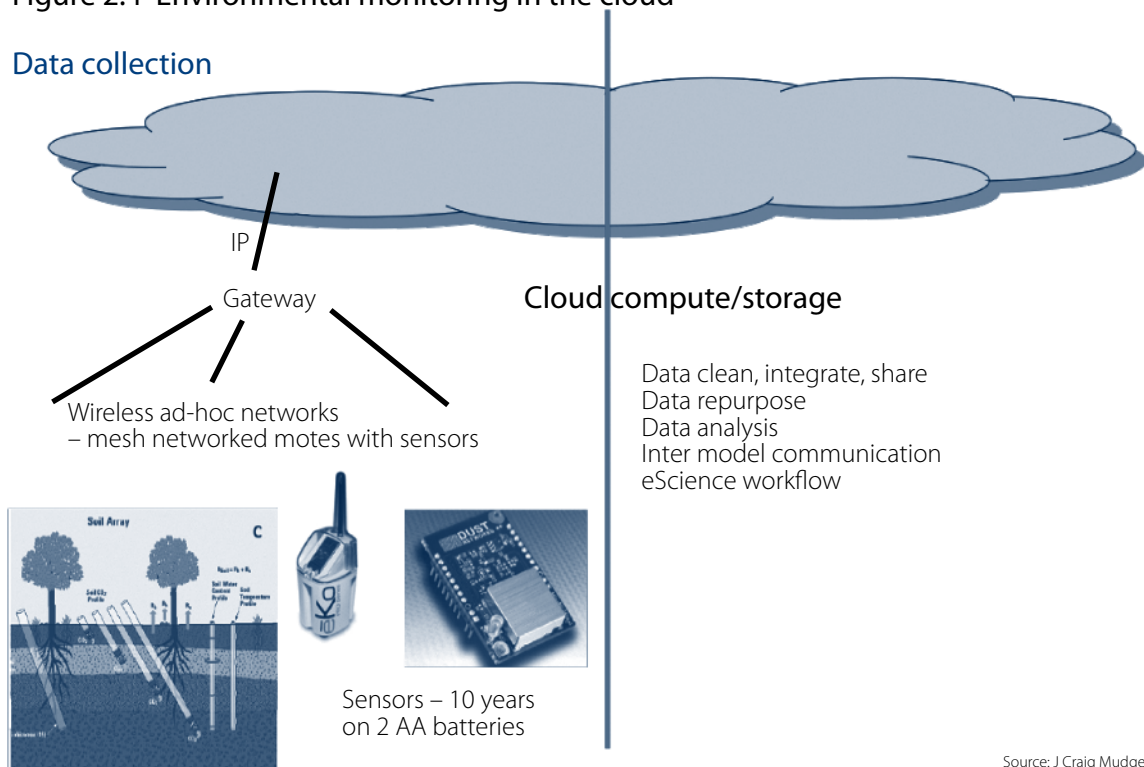
Although this report has emphasised processing, sharing, and other uses of cloud-resident data, the loading and aggregation of data into the cloud is of equal importance, although lagging in practice. The rapid evolution and widespread deployment of sensors – in the soil, tree canopies, in gene sequencing machines, in telescopes, on the sea floor, in point-of-sale terminals, in pallets in containers, and so on – are causing an explosion in data volumes. Within a short time, businesses, governments, and science will structure their information sources and needs as shown in Figure 2.3.

This diagram envisages several applications of wireless sensor networks, which are collections of tiny nodes which communicate their location, identity, and local measurements of their environment wirelessly over low-power computer networks. One simple form of this, Radio Frequency Identification (RFID),¹¹ is beginning to be used in supply chain management in the retail sector. Richer forms are being planned for factory management, agriculture, mine safety, energy exploration, and defence. Battery life and cost are two key design parameters of tiny nodes, often called 'motes', as researchers will want to deploy thousands of them over their domain of interest, and let them gather data unattended for long periods.

¹¹ RFID is a tracking system that uses intelligent bar codes to track items.

Figure 2.4 Environmental monitoring in the cloud

Data collection



Source: J Craig Mudge

A special case is the use of tiny sensors to monitor the environment (Figure 2.4).

Actuation, on the other hand, where commands will be sent to motes equipped to change the environment in which they sit, for example, by adjusting a flow valve, has shown little progress to date. Microelectromechanical systems will play a role in the fabrication of actuators. A closed loop system will sense, determine appropriate responses through models in the cloud, and transmit commands to actuators sitting alongside the sensors.

This report has discussed data sensing and aggregation in the cloud because it illustrates another property of the cloud, namely universality of access. No matter where a sensor, or a smart phone, is located, whether in the desert or in London's central business district, the same access methods (authentication, access control, and data integrity during data transmission to the cloud) will be used.

During the next five years or so, Australia needs to prepare, in terms of infrastructure and organisation, for the time when the world's entire collection of scientific data will be stored on the web. Another scenario to prepare for is an 'Internet of Things', enabled by tiny smart tags wirelessly communicating readings of sensors. Three examples of applications that might be self-managed are:

- vehicle traffic for energy efficient movement of people in cities;
- energy management at the suburb level by networking smart electricity meters on the walls of instrumented homes, offices, and factories; and even
- using Google to find lost car keys.

2.5 BARRIERS TO THE UPTAKE OF CLOUD COMPUTING

The major barriers to the adoption of cloud computing are concerns about security and privacy. Assuring trusted behaviour of both clients and providers is also an important issue. Security has always been an important concern for ICT managers, and the protection of personal, scientific, commercial and intellectual property information must be assured by cloud service providers.

A “one-size-fits-all” approach to providing the latest state-of-the art in security, privacy and trust technologies is neither appropriate nor even possible. For example, Australians may be willing to have personal information stored in Facebook’s cloud in the USA. Yet these same users will cite the multi-tenancy (shared resources) of a warehouse-size data centres as a reason not to store their business data with a cloud provider. Each application of the use of cloud computing technology must be critically and carefully evaluated and considered in terms of identified threats, risks and mitigation requirements.

To strengthen their contribution to sensitive, regulated information, cloud service providers need to define boundaries and responsibilities, noting which security and information management functions and policies they provide, and what security and information management still needs to be provided by a customer or other partner service provider. Typically, any service-level agreements (SLAs) must reflect this partitioning.

Security

A number of the current applications of cloud computing involve consumer services, including e-mail, and social networks. These services collect and store large volumes terabytes of personal information, in data centres in countries around the world. The protection of personal data and management of privacy issues may well determine the success or failure of many cloud services.

The issues of privacy and security are complicated by the location of data in a number of different jurisdictions with differing levels of protection. Wiretapping and other surveillance activities by government agencies is also an issue. There have been examples in the past where information gathering activities bordered on commercial espionage. On the other hand, legitimate access by law enforcement agencies may be difficult. Intellectual property protection, particularly in relation to copyright protection will also raise issues. There have already been attempts to hold internet service providers responsible for preventing the unauthorised sharing of copyright material. How these issues will play out in a cloud environment is unclear.

As an OECD paper has noted:

Companies that wish to provide Cloud services globally must adopt leading-edge security and auditing technologies and best-in-class practices. If they fail to earn the trust of their customers by adopting clear and transparent policies on how their customers’ data will be used, stored, and protected, governments will come under increasing pressure to regulate privacy in the Cloud. And if government policy is poorly designed, it could stymie the growth of the Cloud and commercial Cloud services. (OECD 2009).

Securing information within a cloud computing environment requires three levels of security: network security, host security, and application security. These security needs are also present within in-house infrastructure, and are impacted directly by access policies and workflows of an entity which owns and manages its resources. When an entity moves to cloud computing there are security challenges at each of the three levels, as well as those dealing with the operation of the business and the individuals involved in the system’s deployment and management. Although these security challenges are exacerbated by cloud computing, they are not specifically caused by it.

Encryption is not a complete solution because data needs to be decrypted in certain situations – so that computation can occur and the usual data management functions of indexing and sorting can be carried

out. Thus although data in transit and data at rest are effectively encrypted, the need to decrypt, generally by the cloud service provider, can be a security concern.

Nevertheless cloud services can be augmented by email filtering (including back-up, and spam), Web content filtering, and vulnerability management, all of which improve security. Some threats are better dealt with by warehouse-size data centres (e.g. Distributed Denial of Service attacks, which involve attempts to prevent an internet site or service from functioning). Applications running in the cloud are less vulnerable to these attacks.

Identity and access management and associated policies for the use of cloud services must be equivalent to current enterprise-grade practices, and provide the ability to interoperate with existing applications.

Privacy

Privacy and data protection laws and regulations, such as those of European Union countries and the US Safe Harbor Program, require knowledge of where data is stored at all times. Data may be stored in multiple copies across several jurisdictions with differing privacy and data protection legislation. This is a particular issue for government agencies handling personal information. This issue may encourage some cloud service providers to locate their data centres in jurisdictions with minimal legal requirements (potentially outside Europe and the USA). However users of cloud services may need the legal protection provided by strong data protection legislation.

Again, integrating agreed-upon behaviour between provider and client, demonstrating to each other for each transaction that the agreement is being maintained, and an externally auditable record of critical interactions, will raise the trust and confidence in the use of cloud services.

Trust

Ultimately, however, the uptake of cloud services is dependent upon providers and their clients trusting each other. Trust need to have a legally enforceable or standards basis if it is to work in practice. *Trusted behaviour* includes security and privacy measures. Trust in cloud computing needs to use an integrated approach that builds on the ideas underpinning Service Level Agreements (SLAs):

- Agreement upon policies for sharing information ahead of any interaction between the client(s) and cloud service providers. This may be achieved through the use of contracts or SLAs defining access to specialised services or access control policies.
- Rigorous proofs of agreed good behaviour (as defined in the SLA) between the cloud service provider and their clients during all critical interactions, and after the policy agreements have been reached.
- Ensuring externally verifiable and auditable “good behaviour” records are kept over the life span of interactions, from commencement to termination.

Taken on their own, each of these approaches raises the confidence between all parties that there is some degree of trust in place. It is only when all three components are linked together that a truly trusted system can be achieved.

Service lock-in and Standards

Service lock-in is a concern to some users. Because of a lack of standardisation, consumers cannot easily extract their data and programs and move to another supplier in the cloud. The success of the internet is largely due to the adoption of the Internet Protocol, based on open standards and could be implemented in open source software. For cloud computing to succeed, Governments will need to promote open, international standards for the Cloud so that users will be able to switch Cloud service providers with a minimum of cost and risk.

Other issues

There have also been concerns expressed about:

- Data transfer bottlenecks – as applications become more data-intensive and computing is spread across the cloud, the need to move large volumes of data may increase;
- The costs of network traffic – some organisations have cost models that restrict or penalise access to external and overseas facilities;
- Loss of data – if a cloud service provider loses data, this is a major problem for users relying on the service; and
- Inability of consumers to be able to seek effective redress.

Management of these issues

A 2009 report of the European Network and Information Security Agency (ENISA) on cloud computing concluded that the cloud's economies of scale are 'both a friend and a foe', from a security perspective. A massive concentration of resources and data presents an attractive target to attackers but cloud-based defences can be more robust scalable and cost-effective (ENISA, 2009a).

For all the reasons noted above, companies, government agencies, and academic researchers need to have their security staff study the cloud environment, just as they did when deciding to outsource some of their infrastructure and applications. Policies, processes, and adherence to standards such as ISO/IEC 27000 may need to be extended or modified. A portion of the cost savings obtained by the use of cloud computing services can be used to fund this effort.

A very useful framework for doing this is the *Security Guidance for Critical Areas of Focus in Cloud Computing V2.1* (Cloud Security Alliance, 2009). That document addresses 13 domains, namely:

- Cloud Computing Architectural Framework;
- Governance and Enterprise Risk Management;
- Legal and Electronic Discovery;
- Compliance and Audit;
- Information Lifecycle Management;
- Portability and Interoperability;
- Traditional Security, Business Continuity, and Disaster Recovery;
- Data Centre Operations;
- Incident Response, Notification, and Remediation;
- Application Security;
- Encryption and Key Management;
- Identity and Access Management; and
- Virtualisation.

3 Roles and Opportunities for Government

The main opportunities arise from the benefits of lower cost, universality of access for public sector information, greater flexibility to innovate, and greater energy efficiency.

3.1 THE ROLES OF GOVERNMENT IN CLOUD COMPUTING

US Government adoption of cloud computing

In May 2009, the Obama administration's 2010 budget requests indicated that White House officials want government departments and agencies to launch pilot projects that identify common services and solutions that focus on using cloud computing. The request (Office of Management and Budget, 2009) included:

"Cloud-computing will help to optimize the Federal data facility environment and create a platform to provide services to a broader audience of customers. Cloud-computing and "work-at-a-distance" represent major new Government-wide initiatives, supported by the CIO Council under the auspices of the Federal CIO (OMB's E-Government Administrator), and funded through the General Services Administration (GSA) as the service-provider. Of the investments that will involve up-front costs to be recouped in out-year savings, cloud-computing is a prime case in point".

In September 2009 at NASA's Ames Research Center, the US Chief Information Officer Vivek Kundra unveiled the website www.apps.gov operated by the General Services Administration (GSA) to help US Government agencies adopt cloud computing. The GSA has contracted Terremark Worldwide to provide cloud-based hosting of the US Government's primary e-government portals (Wyld, 2010).

At the same event, Google announced it was building a cloud for the US government and that it had been working on achieving the US Government's Federal Information Security Management Act (FISMA) certification, required of government IT contractors.

Many US Government agencies active in cloud computing initiatives include the Defense Information Systems Agency, the Department of Energy, the Department of the Interior, and NASA (which has recently launched the NEBULA cloud computing platform) (see Wyld, 2010 for more details).

US NSF initiatives

The US Government is leading the way in promoting the adoption of cloud computing. In February 2008 the US National Science Foundation (NSF) invited the first proposals for the new *Cluster Exploratory (CLuE) Initiative* as a part of a relationship between Google, IBM and NSF. NSF expects that CLuE will lead to innovations in the field of data-intensive computing, as well as serve as an example for future collaborations between the private sector and the academic computing research community.

CLuE is providing NSF-funded researchers with access to software and services running on a Google-IBM cluster to explore innovative research ideas in data-intensive computing. Funding is directed to proposals that are likely to contribute to science and engineering research and produce applications which promise to benefit society as a whole.

NSF recognised that private sector companies had launched a number of highly effective internet-scale applications powered by massively scaled, highly distributed computing resources known as data clusters.

These clusters include large-scale server farms with hundreds of gigabytes of disk capacity. The resulting network capacity and fundamental changes in computer architecture provide software developers with opportunities to take new approaches to problem solving.

Such resources had not been easily available or affordable for academic researchers. In October 2007, Google and IBM created a large-scale computer cluster of approximately 1600 processors to give the academic community access to otherwise prohibitively expensive resources. CLuE developed from this initiative.

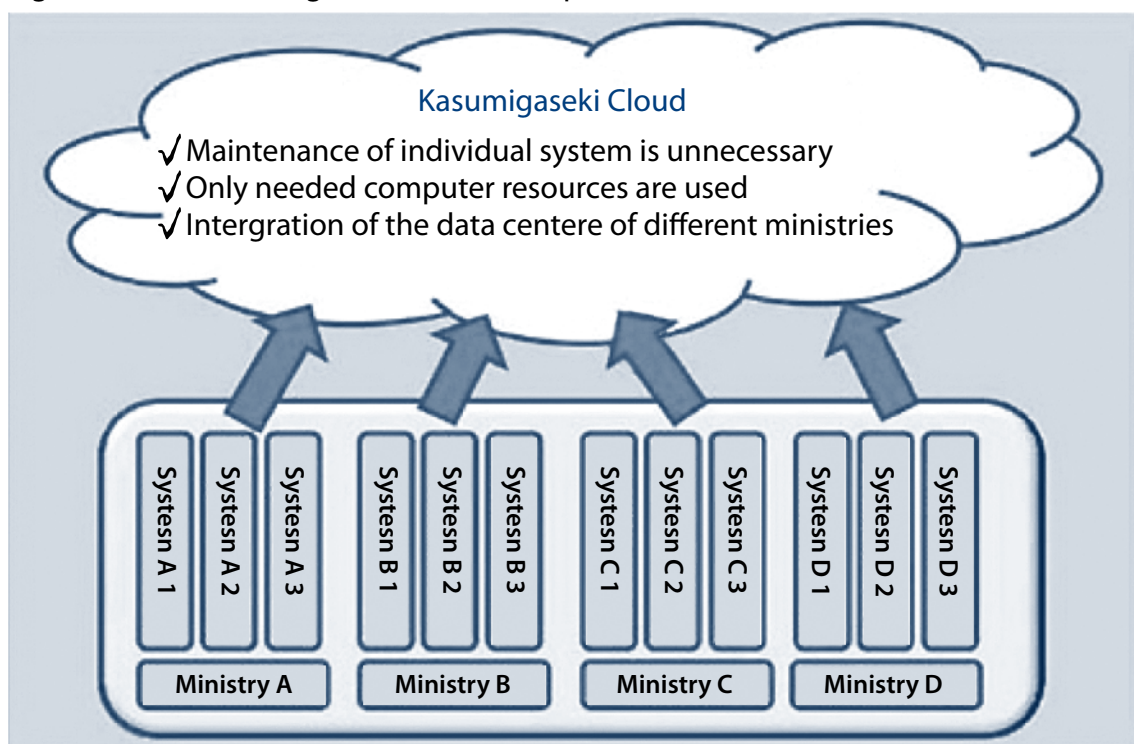
NSF sought proposals from academic researchers, some of whom have then be selected by NSF to have access to the cluster. NSF offered support to the researchers to conduct their work while Google and IBM would cover the costs associated with operating the cluster and provide other support to the researchers. In April 2009 NSF announced 14 awards totalling US\$5 million for research over a period of up to two years.

In February 2010 Microsoft and the NSF announced an agreement to offer individual researchers and research groups selected through NSF's merit review process free access to advanced cloud computing resources. The program is designed to help broaden researcher capabilities, foster collaborative research communities, and accelerate scientific discovery.

Japan

The 2009 Digital Japan Creation Project (ICT Hatoyama Plan) seeks to create new ICT markets to boost Japan's economy. Within this plan is an outline to create a nation-wide Cloud Computing infrastructure referred to as the Kasumigaseki Cloud. The plan proposed that the Government take initiatives to implement cloud computing, in stages, by 2015. The Kasumigaseki Cloud (see Figure 3.1) will enable various ministries to collaborate to integrate and consolidate hardware and create platforms for shared functions (Japanese Ministry of Internal Affairs and Communications, 2009).

Figure 3.1 The Kasumigaseki Cloud concept



The United Kingdom

The United Kingdom Government's Chief Information Officer has outlined the Cloud Strategy for the UK. It has six major elements which include the adoption of open standards and the rationalisation of data centre real estate.¹² These elements are seen as underpinning the establishment of a UK onshore private Government Cloud providing Infrastructure as a Service, middleware, Platforms as a Service and Software as a Service. There are plans to establish a Government Application Store. The report *Digital Britain* (DCMS & DBIS, 2009) provides supporting details of the strategy outlined above.

Early in 2010, the UK Government revealed plans to cut £3.2 billion from its annual spending by 2013-14 by sharing software and IT services in the cloud. One measure involves building the Government Application Store (see www.elasticvapor.com/2009/06/more-details-on-uk-governments-cloud.html).

Europe

Other European countries where government agencies are implementing cloud initiatives include Sweden, France and Spain. Examples of applications include management of public sector housing, transportation service networks, the census, health services, contracts, and education services (Petrov, 2009).

An expert report to the Commission of the European Communities (CEC, 2010) undertook an analysis of strengths, weakness opportunities and threats (SWOT) in relation to cloud computing in Europe. Table 3.1 summarises the SWOT analysis. Many of the weaknesses and threats in the SWOT table would be applicable to Australia.

The report recommends:

- The European Community (EC) should stimulate research and technological development in the area of Cloud Computing;
- The EC together with Member States should set up the right regulatory framework to facilitate the uptake of Cloud computing;
- The European Union needs large scale research and experimentation test beds;
- The EC together with industrial and public stakeholders should develop joint programs encourage expert collaboration groups;
- The EC should encourage the development and production of:
 - Cloud interoperation standards; and
 - an open source reference implementation; and
- The EC should promote the European leadership position in software through commercially relevant open source approaches

Australia

A recent Commonwealth Government report (Government 2.0 Taskforce, 2007) touched on issues relevant to cloud computing but, apart from concerns about the archiving of government records (Recordkeeping Innovation, 2007), failed to address the opportunities which cloud computing offers some government agencies.

While it is clear that governments in leading OECD countries are moving to adopt cloud computing, there are a few signs that the Commonwealth¹³ and State Governments have started to implement cloud computing solutions. It appears that regulatory, security and privacy issues have yet to be addressed.

¹² See http://www.cabinetoffice.gov.uk/cio/greening_government_ict.aspx

¹³ See for example http://www.cio.com.au/article/349566/immigration_department_cio_talks_cloud_computing/ and <http://www.itwire.com/business-it-news/technology/37791-anz-cios-focus-on-pedestrian-technologies-gartner>

Table 3.1 SWOT analysis – cloud computing in Europe

Strengths	Weaknesses
<ul style="list-style-type: none"> • Knowledge background and expertise in related technological areas • Significant expertise in building high-value industry specific applications • On-going research projects and open source technologies • Strong SOA and distributed systems research community • Strong synergies between research and industry; technological platforms • Concerted government effort (legislation etc.) • Selling products & telecommunications (as opposed to selling new technologies) • Provisioning of complex processes as services, rather than of low level infrastructures • Strong telecommunication industry (research, consumer focus, investment capabilities) • Commercial success-stories 	<ul style="list-style-type: none"> • Few resource infrastructures available in Europe • Comparatively weak development of new (cloud) technologies in comparison to US • Primarily consumer; main Cloud providers are not European • Research timelines vs. fast moving markets • No market ecosystem around European providers • Subsidiaries and fragmentation of key industries • No platform to find / select cloud providers
Opportunities	Threats
<ul style="list-style-type: none"> • Strong experience and involvement in standardisation efforts • European companies use (and need) their own clouds (private clouds) (cf. location) • Growing interest from both industry and academia in cloud technologies (cf. "readiness") • Existing infrastructures with strong resources and in particular with strong communication networks (e.g. telecoms) • Clouds provide an excellent backend for mobile phone applications (which have usually low power local resources). • Increase competitiveness and productivity, of service providers by adoption of local/hybrid/public computing platforms • Application provisioning instead of technology orientation • Support SMEs and start-ups with improved ROI (elasticity), reduced time to market and easy adoption • New business models for cloud improved products and cloud adopters • High awareness for the green agenda and new approaches to reduce the carbon footprint • Similar business incentives and infrastructure requirements between Grid and Cloud, facilitating the movement from Grid to Cloud Provider 	<ul style="list-style-type: none"> • Better developed cloud infrastructures (mainly in the US) already exist • High investment and funding required to build up infrastructure • Investment/economic benefit asymmetry (IPR, OSS, commercialisation) • Lacking IaaS provider(s) • Dependency on external (non-European) providers • Technological impact / development underestimated • Latencies (federation too inefficient)

Source: CEC, 2010, page 36.

RECOMMENDATION 1

The Commonwealth Government should take a whole-of-government approach to new security and privacy issues arising from the use of cloud computing, by establishing a new taskforce to review the adequacy of current legislation and identify what steps need to be taken to ensure a supportive regulatory environment.

Open access to public data

The Gov 2.0 Taskforce has recommended improved access to public sector information. Implementation of this recommendation could provide opportunities for data aggregating, data repurposing, data mining, indexing, and analysis. This could encourage innovation, and opportunities for cloud services as questions of scientific and social value are asked, and techniques for asking them are developed.

4 Opportunities for Business

Businesses can benefit from all aspects of cloud computing, although cost savings for individual firms are large only when there is massive scale use of the cloud – when large numbers of commodity machines are used in warehouse-size data centres. Some business value can be achieved via pay-per-use and MapReduce using a data centre with a few servers. However factors of five in cost reduction or real elasticity in the scale of use require large-scale facilities.

Barriers to adoption of cloud computing by business include the cost of migration to a cloud model, fear of lock-in to one cloud service provider, data security and the lack of large-scale data centres in Australia. These are some of the reasons for believing that the uptake of cloud computing in large business is slower in Australia than in the USA.

For small and medium sized enterprises (SMEs) cloud computing offers significant opportunities. Using the cloud, SMEs can overcome some of the advantages that their bigger competitors have had in the past such as the capital-intensive hardware investments needed to provide services. However, as a European survey shows, while SMEs are enthusiastic about cloud computing, they also have concerns (see Table 4.1). While some of these concerns would also be shared by larger companies, SMEs face greater difficulties in addressing them.

Table 4.1 SME concerns ranked in declining order of importance

Concern
Privacy
Availability of services and/or data
Integrity of services and/or data
Confidentiality of corporate data
Repudiation
Loss of control of services and/or data
Lack of liability of providers in case of security incidents
Inconsistency between trans national laws and regulations
Unclear scheme in the pay per use approach
Uncontrolled variable cost
Cost and difficulty of migration to the cloud (legacy software, etc)
Intra-clouds (vendor lock-in) migration

Source: ENISA 2009b

Nevertheless, some large Australian companies are evaluating the cloud, and some are using it already, often through Amazon Web Services. An Adelaide software company which provides optimisation solutions to its customers is using Amazon Web Services for its computationally intense jobs. Job completions are being achieved in a week instead of two or three months on the clusters it has access to in Australia. Moreover, as the company's customer is interested in energy efficiency, the customer is happier for the firm to use Amazon, rather than an Australian cluster which is not likely to be able to match Amazon's efficiency. An Australian company developing software for the insurance industry is using Amazon for program development, testing on dummy data. Production runs are done in-house in Sydney on real data. A New Zealand start-up company is using Amazon for its fee-for-service supply of public images, which it makes more accessible than the original government source.

Data mining is an application well-matched to the distributed data-parallel computing of the cloud. Teradata, a large IT company providing parallel data bases, is exploring MapReduce, yet another example of the importance of cloud computing to business and government.

Some functions that fit naturally with the cloud, such as hierarchical storage management, are being extended to the cloud. This type of evolutionary innovation is expected to become common.

Private clouds, located in Australia and built for a single customer, can counter security and privacy concerns. However massive scale is needed to get the most business value from the cloud computing concept. On the other hand, SMEs may be able to benefit from multi-tenancy clouds where other tenants have high security requirements.

4.1 EXAMPLES OF CLOUD USE BY BUSINESS

Some of the reasons that business is attracted to using cloud computing are illustrated in the examples listed in this section. The elasticity of Amazon's service is illustrated by the example below.

The Animoto story

A vivid example of cloud power comes from Animoto, a (then) 18-month-old start-up in New York that lets customers upload images and music and automatically creates customised Web-based video presentations from them; earlier this year about 5000 people a day were trying it. Then, in mid-April, Facebook users went into a small frenzy over the application, and Animoto had nearly 750,000 people sign up in three days. Animoto had neither the money to build significant server capacity nor the skills – and interest – to manage it. However, it had designed the application for Amazon's cloud. That paid off during the three-day surge in growth, when Animoto did not buy or configure a single new server. It added capacity on Amazon, at the cost of about 10 cents a server per hour, as well as some marginal expenses for bandwidth, storage and some related services. Animoto traffic doubled every 12 hours, and its cloud service provider scaled from 50 servers to over 3500, then just as importantly, scaled back down when the excess demand ceased a few days later.

Source: Based on *New York Times* article, 25 May 2008.

As Amazon has stated:

"Our customers can instantly provision as much or as little capacity as they like to perform data-intensive tasks for applications such as Web indexing, data mining, log file analysis, machine learning, financial analysis, scientific simulation, and bioinformatics research" (Amazon entry, Apache Hadoop website).¹⁴

Journey Dynamics, based in Guildford, United Kingdom, provides technology and information that enables it to deliver dramatically improved forecasts of journey times, personalised and smarter routing, on a global basis. Devices, application and services powered by Journey Dynamics technology are able to:

- more than triple the accuracy of journey time forecasts;
- enable smarter routing options that reflect the vehicle, the time of day and the day of week for the journey; and
- learn and understand individual driving styles so that they can offer personalised routes.

These features translate into the delivery of real benefits to the users, to reduce journey times, provide fuel cost savings, reduce emissions and deliver an improved driving experience. Journey Dynamics uses Hadoop/MapReduce to analyse billions of lines of GPS data.

BabaCar is a car hire comparison search engine for both luxury and economy car hire. It works with leading and independent car hire companies to provide quotes which compare prices from over 450 car hire companies. BabaCar has been online since 1999 and has the facility to book a car hire online at more than 6000 locations in 134 countries worldwide. To do this, BabaCar uses cloud computing to search and analyse millions of rental bookings.

¹⁴ See <http://wiki.apache.org/hadoop/PoweredBy>

The *New York Times* decided to make their archives 1851–1980 free, and their delivery strategy required all 11 million articles to be available as images in PDF format. By using Hadoop on Amazon Web Service, they spread the job over multiple machines and finished in just under 24 hours at a cost of only US\$240.

At **Detikcom**, Indonesia's largest news portal, search log analysis and determination of most-viewed pages is performed using Hadoop.

Eyealike is a leading visual-based search platform for facial similarity, image detection, and video copyright surveillance. It is based in Seattle. The Eyealike Visual Search Platform, based on its own research and development and technologies from the University of Washington, offers a new approach to searching, finding, and accessing people, images, and other rich media content.

At **Palo Alto Research Centre Inc (PARC)** (formerly Xerox PARC) Hadoop has been used to analyse Wikipedia conflicts. A complete history dump of the English Wikipedia of 2 July 2006 was analysed. This included over 58 million revisions, of which 2.4 million were article entries in the encyclopaedia, totalling approximately 800 gigabytes of data. To process this data, the raw text was imported into the Hadoop distributing computing environment running on a cluster of commodity machines. The Hadoop infrastructure allowed the PARC team to explore new full-scale content analysis techniques quickly (Suh *et al*, 2007).

Facebook has tens of millions of users and more than a billion page views every day. As a result, Facebook ends up accumulating massive amounts of data. One of the challenges has been to develop a scalable way of storing and processing all these bytes, as this data plays a key role in improving the users' experience. Facebook is a big user of Hadoop to store copies of internal log and dimension data sources and use them for reporting, analysis and machine learning. Currently Facebook uses a 600-machine cluster with 4800 cores and about seven petabytes of storage, where each node has eight cores and 12 terabytes of storage.

PSG Tech in Coimbatore, India, is a higher education institute. Protein sequence analysis is one of its areas of interest. Multiple alignment of protein sequences helps to determine evolutionary linkages and to predict molecular structures. The dynamic nature of the algorithm coupled with data and computing parallelism of Hadoop data grids improves the accuracy and speed of sequence alignment. This is of interest to university and industry researchers.

4.2 OPPORTUNITIES FOR AUSTRALIA

Australians are capable of coming up with ideas for the next YouTube, or Facebook, and it is important that these ideas are not held back by lack of infrastructure. The properties of shared cost, elasticity, and pay-per-use are attractive to start-up companies.

Given the low cost of infrastructure for users of the cloud, Australian start-ups can compete on ideas, rather than on the amount of resources (cash or equipment) they have. The cloud will catalyse development of new products and services.

Furthermore, the elimination of capital for computer equipment is expected to result in at least an order of magnitude increase in new cloud-based ventures. For example, this can be seen in the extreme with the creation of over 200,000 applications that run on the iPhone and the many new ventures that have been created using computational and storage resources at Amazon. There are already examples of cloud facilities resident abroad that have enabled numerous Australian start-ups to provide services to Australian customers.

If local cloud facilities and platforms are developed, either by local companies or appropriate strategic partnerships, this could mitigate some of the legal and security issues described in Chapter 2. Moreover, cloud computing is a disruptive technology which, if implemented fully (data centres of massive scale, utility computing, pay-per-use, etc.) could also help promote an innovative culture.

RECOMMENDATION 2

The Commonwealth Government's Commercialisation Australia Program, which assists researchers, entrepreneurs and innovative companies to convert ideas into successful commercial ventures, should actively encourage new businesses that are cloud-focused in their internet-scale applications.

Broadband networks

Broadband internet access is essential to cost-effective use of cloud computing. The Commonwealth Government's decision to fund a National Broadband Network may assist in this regard. OECD data¹⁵ illustrate the higher cost of broadband in Australia compared with some other countries. The UK is the leader, with broadband subscription prices in the latest OECD survey range from US\$7.57 to US\$56.57. Australian ranks eighteenth, with prices ranging from US\$21.02 to US\$129.83. In the USA, the range is from US\$19.99 to US\$144.95.

There are pricing anomalies in the existing networks in Australia which penalise commercial users and research organisations that sit outside our universities. This includes certain state government-funded collecting institutions which are research intensive.

RECOMMENDATION 3

The Commonwealth Department of Broadband Communications and the Digital Economy should review broadband pricing and policies of commercial and other providers to ensure that unnecessary impediments to the uptake of Cloud Computing are minimised.

Response of the IT industry in Australia

The year 2010 has seen a lot of activity in Australia, as local IT vendors of all types move to seize opportunities presented by cloud computing. An important step towards outsourcing to the cloud requires applications to be delivered as Software-as-a-Service (SaaS) and this provides opportunities for providers of virtualisation software such as VMWare.

Data centres are being built in Australia to satisfy the need for SaaS, particularly for those customers who are unwilling to house their data in foreign data centres, as is the case with current cloud service providers. As a result, there is activity by cabling service providers, modular container providers, such as SGI and Sun, security services, and heating, ventilation and air conditioning suppliers. An interesting business alliance, the Virtual Computing Environment Coalition involving VMWare, Cisco, and EMC, has been formed in the USA and this approach is now found in Australia with Brennan IT, a company founded in 1997 to meet the business and technology needs of Australian mid market businesses.

Importantly, because of the smaller scale of Australian data centres, the SaaS customers of these data centres are not receiving the cost benefits nor can they receive the business model benefits warehouse-size data centres can offer. This issue has been illustrated by James Hamilton, a computer scientist with extensive experience in software development and more recently in cloud infrastructure development at

¹⁵ See http://www.oecd.org/document/54/0,3343,en_2649_34225_38690102_1_1_1_1,00.html

Microsoft and Amazon. His 2006 comparison of costs for different sized data centres shown in Table 4.1 gives this comparison between medium-sized and large-scale data centres.

Table 4.1 Data centre economies of scale

Technology	Cost (medium sized DC)	Cost (very large DC)	Ratio
Network	\$95 per Mbit/sec/month	\$13/ per Mbit/sec/month	7.1
Storage	\$2.20 per GByte/month	\$0.40 Gbyte/month	5.7
Administration	~140 servers/administrator	>1000 servers/administrator	7.1

Note: 1. US dollars, 2006 prices.

2. Medium sized data centre ~ 1000 servers, very large data centre ~ 50,000 servers.

Source Hamilton, 2008

Nevertheless, data centres housing SaaS in Australia do provide more work for Australian developers and outsourcing providers.

However, 2010 has seen two business actions that deliver true cloud services to the Australian market:

- Virtual Ark, a start-up founded by Marty Gauvin, the founder of a successful hosting business Hostworks, is offering cloud services to independent software vendors. Virtual Ark operates globally on a pay-per-use basis by using the cloud services of overseas providers, or private clouds in Australia as required. (Virtual Ark, 2010).
- Amazon Web Services announced an increase in activity in the Australian market at CeBit in May, together with appropriate tutorial sessions, and the construction of a warehouse-sized data centre in Singapore.

Other Australian developments include decision by two major banks to cut capital investment and move to buying computing power, including software as a service, at contestable spot prices. Telstra is reported to have initiated arrangements to provide cloud services to packaging company Visy and heavy equipment manufacturer Komatsu.¹⁶

¹⁶ Reported in the Australian Financial Review, 10 July 2010.

5 Opportunities for Supporting University Research

University researchers in the sciences, engineering, medicine, and the social sciences make extensive use of computing and communication and the Web, whether it be for genome sequencing, modelling in nuclear physics, geophysics computation, seismic modelling, data mining, climate change modelling, water management, developing new composite materials, or simply publishing research results.

In Australia, university researchers obtain large-scale computing and storage resources from a variety of facilities, which may be located within the researcher's university, at a state-based facility supporting multiple universities, at major national facilities or through the specialised IT national infrastructure programs.

At a national level, these resources are supported mainly by programs in the National Collaborative Research Infrastructure Strategy (NCRIS), particularly through Platforms for Collaboration (PfC),¹⁷ and by the Research Councils through their infrastructure programs. PfC supports a number of programs for computation, data management and collaborative environments. These programs are focal points for e-research developments which are occurring throughout research support infrastructures shaped by the NCRIS Roadmap and implemented through coordinated use of resources from multiple sources across the sector.

5.1 AN EMERGING PUBLIC-PRIVATE RESEARCH INFRASTRUCTURE SERVICES SECTOR

Over the past decade or so, partnerships and collaborative ventures have been created to provide researchers with services which support IT based innovation in research methods. For the most part, Partners in these ventures have been the universities, CSIRO and the following six state-based organisations:

Table 5.1 State-based research services organisations

State	Organisation
NSW	Intersect Australia Ltd
Vic	VPAC – Victorian Partnership for Advanced Computing
SA	eRSA – eResearch SA, formerly the South Australian Partnership for Advanced Computing
WA	iVec – formerly the West Australian Interactive Virtual Environments Centre
Qld	QCIF – Queensland Cyber Infrastructure Foundation
Tas	TPAC – Tasmanian Partnership for Advanced Computing

Funding for these state-based facilities is provided by a combination of Commonwealth Government (Department of Innovation, Industry, Science and Research) and the universities in each state, generally with help from state governments. The facilities provided these organisations support a wide range of users, applications and requirements, including high-performance computing, high-throughput computing, and data-intensive computing.

Federal funds currently support an experimental national storage infrastructure known as the ARCS Data Fabric, based on the Integrated Rule-Oriented Data System (iRODS)¹⁸ for distributed data management. Storage for the ARCS Data Fabric is mainly provided by the organisations mentioned above.

¹⁷ For more information see <http://ncris.innovation.gov.au/Capabilities/Pages/PfC.aspx>

¹⁸ iRODS™ is a data grid software system developed by the Data Intensive Cyber Environments research group and collaborators. For more information see <http://www.irods.org.au>

The computing facilities provided by the National Computational Infrastructure (NCI), a project established through the PfC, are complementary to cloud based services which, for some research support services, are integrated as part of NCI support for particular research projects.

Computer science researchers in the Nimrod Project¹⁹ at Monash University have created cloud oriented research computing system which reflects the value of resources available on a pay-per-use basis (Bethwaite *et al*, 2009):

“These resources open the opportunity for university researchers to buy compute time on an ad-hoc basis – shifting university funding models from capital expenditure to recurrent costs. This transition poses many policy issues as well as a range of technical challenges. Existing resources that are free will not disappear – there is clearly a role for continued investment in university infrastructure. On the other hand, commercial clouds could provide an overflow, or elastic, capability for individual researchers. One could easily imagine a research group performing much of their base-load computations on ‘free’ resources, but resorting to pay-as-you-go services to meet user demand”.

Professor David Abramson and his team at Monash University have built an extension to their Nimrod tool set to add Amazon’s Elastic Compute Cloud (EC2), so users can mix commercial cloud resources and conventional grid resources and university level infrastructure. This is a “hybrid cloud” in NIST terminology. Their first case study uses Bayesian statistics for training a predictive model within a ‘recommender system’ which suggests exhibits to museum visitors.

Training

Given that cloud computing is a step change in the cost/performance of ICT systems, the higher education sector has a special role in preparing Australia for its adoption. The US experience suggests that it may be possible to engage with cloud service providers to offer short courses which would accelerate the uptake of cloud computing in Australia.

RECOMMENDATION 4

Australian universities should expand their courses to build knowledge and skills in cloud computing. They should seek the support of cloud service providers to offer short courses for industry, government and university researchers on cloud computing, including its data-parallel programming models.

5.2 INNOVATIONS IN DATA-INTENSIVE COMPUTING

The US National Science Foundation, recognising that cloud computing would lead to innovations in the field of data-intensive computing, mounted a program which provided NSF-funded researchers with access to cloud computing resources provided by Google and IBM.

Given the clear signs that cloud computing is an area of emerging importance, it is important that Australia is not left behind in adopting this new approach. Australia’s universities are in an ideal position to use cloud computing and to train graduates in its use.

RECOMMENDATION 5

The Commonwealth Government should create and fund an Australian equivalent of the NSF’s Cluster Exploratory Program CluE and the NSF-Microsoft Program to encourage the use of cloud computing. This could be administered by the Australian Research Council.

¹⁹ Nimrod/G is a widely adopted Grid middleware environment for building and managing large computational experiments over distributed resources. One of the goals of the Nimrod project is to create a user-friendly interface to give scientists access to large computational experimentation.

5.3 FUNDING STRATEGY FOR UNIVERSITY ICT RESEARCH INFRASTRUCTURE

In October 2009, the Minister for Innovation, Industry, Science and Research announced the formation of the National Research Infrastructure Council (NRIC) to provide strategic advice on future research infrastructure investments, including those to be funded through the Super Science Initiative. The Australian eResearch Infrastructure Council (AeRIC) informs NRIC's consideration of issues relating to eResearch infrastructure.

There are many areas that NRIC and AeRIC will be addressing in the near future to better serve Australian science. Three such areas are:

- a) Massive deployment of tiny sensors and actuators in wireless sensor nets, which are immediately relevant in environmental sensing, and in smart physical infrastructure. These will generate large volumes of data. Moreover, NCRIS Project TERN (Terrestrial Ecosystem Research Network) will be extended significantly, for example, by making use of mesh networks and ultra low power sensors;
- b) The forthcoming data deluge from a range of new instruments such as a new telescope (Bell *et al.*, 2009); and
- c) The best computer systems architecture structure for 'taking the computation to the data', which is a central issue in internet-scale computing. File systems of Google and Microsoft Azure provide a workable solution, by means of their replication and co-location at the server level. However this is still an area requiring research in computer systems design.

Recognising that new technology is changing infrastructure requirements, the laboratories of Google, Microsoft, Yahoo, and Amazon Web Services have made substantial contributions to computer science in the last five years, particularly to distributed systems architecture.

The current business plan of PfC recognises the need for AeRIC to be more strategic in its strategy formulation, as it states (PfC, 2007):

"To ensure the widest possible input, AeRIC will host a broadly inclusive e-Research forum to gather strategic input in refining its business plan and activities."

Advances in distributed systems architecture, the arrival of utility computing, the rapid change in the technology base of eResearch, the emergence of new business models, such as usage-based pricing, and the cost savings possible from cloud computing can provide major benefits for Australian researchers.

NRIC needs to gather strategic input in all these areas to refine its investment plans. In doing so, NRIC should seek input from leading commercial cloud providers.

RECOMMENDATION 6

NRIC should refine its investment plans to reflect the benefits that cloud computing can provide.

5.4 ACQUISITION OF COMPUTER AND STORAGE INFRASTRUCTURE

Over the years, Australian science has been fortunate in the provision of Commonwealth Government funded grid services, state government-supported and state-based facilities, its peak computer facilities, and its network capability. However, changes in the technology base suggest a reappraisal of this approach.

Australian consumers are familiar with, and adapt to the constant year-after-year price decline in consumer electronics, whether they are flat panel displays, personal computers, or digital cameras. The

same price decline applies to ICT infrastructure for science – originally it applied to computers and storage, then network switches and flat panel displays, and now sensors.

The following simple model illustrated in Figure 5.1 shows a savings of 50 per cent if usage-based procurement is used for an acquisition of storage valued at \$50 million. There are two factors at work in this model. It factors in a price decline of 40 per cent per year and assumes that storage is not purchased until it is needed.

This simple situation assumes usage increases over the five years such that the capacity needed grows from 10 to 20, 35, then 50 petabytes in year four, and that in year one, 50 petabytes would cost \$50 million.

As discussed earlier in this report, scale impacts on cost. This is most clearly in the warehouse-sized data centres that are deployed by Google, Microsoft and Amazon. The evolution of Australian data centres and the impact of scale are important to understand. During this decade, it is expected that there will be an evolution from small, distributed, data-oriented computing centres (e.g. 1000 node, one petabyte) to more cost-effective, very large scale commercial cloud services (e.g. 100,000 node, 100 petabyte). As a result, investment in distributed, small scale centres, such as Australia currently has with one-per-state may not be the best way to proceed.

Understanding economies of scale in the realities of a fast-changing ICT infrastructure world, driven by massive warehouse-sized data centres, will not only save money, it could stimulate the location of large-scale facilities in Australia.

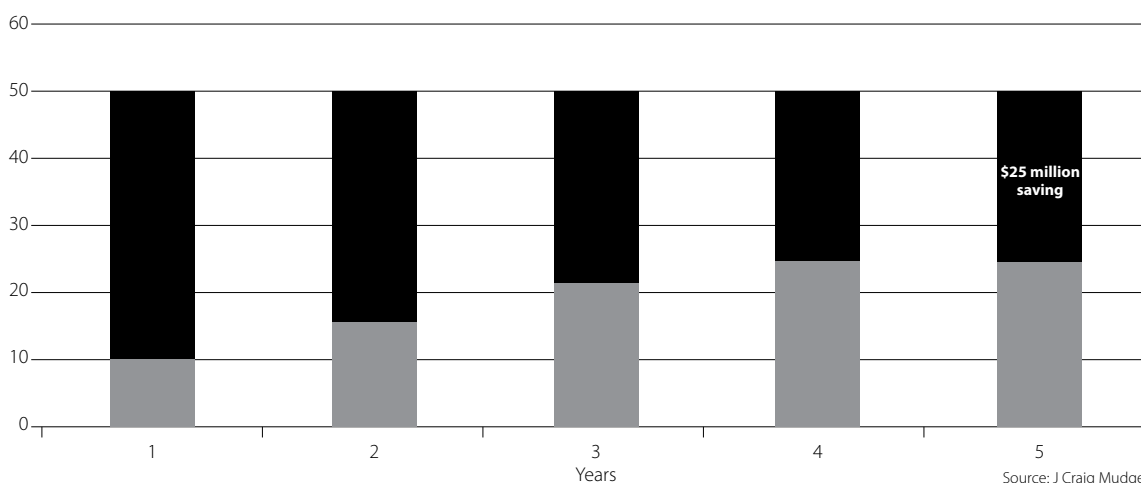
NICTA is working on the costing of the various commercial cloud offerings, and conducted a two-day workshop, Enterprise Cloud Computing – Understanding Costs, Managing Risks and Realising Business Value, in April 2010.²⁰

NICTA is also:

- assisting businesses and the Commonwealth Government in evaluating cloud computing platforms, opportunities, risks, technical challenges, proof of concepts;
- building modelling technologies that assist with forecast of resource usage in cloud and corresponding costs;

Figure 5.1 Acquisition of storage with a current value of \$50 million

Cumulative expenditure (\$ millions)



²⁰ See http://www.nicta.com.au/news/events/event_repository/upcoming_events/enterprise_cloud_computing__understanding_costs_managing_risks_and_realising_business_value

- benchmarking cloud platforms from various perspectives, performance, consistency characteristics, virtualisation technology overhead, network implications etc;
- using artificial intelligence-based techniques for moderating workloads across local servers and cloud, enabling 'intelligent cloudburst' scenarios;
- building cloud monitoring and management technologies; and
- working closely with large cloud platform vendors on various cloud computing research topics, through various researcher exchange and resource support arrangements.

A recent document, *Super Science – Research Data Storage Infrastructure Implementation Discussion Paper* (DIISR, 2010) addresses the issue of storing Australian research data. This Discussion Paper presents an option that allows for clouds and indicates that the Commonwealth Government should consider the use of cloud computing as an option in determining a solution to the problem outlined. Further, proposals for research data storage using NCRIS funds should evaluate cloud computing services and the major cloud service providers (such as Amazon Web Services, Google, IBM and Microsoft) should be invited to submit proposals.

RECOMMENDATION 7

The Commonwealth Government should ensure that proposals for research data storage using Super Science funds have evaluated cloud computing services.

Abbreviations and Acronymns

AeRIC	Australian eResearch Infrastructure Council
ANU	Australian National University
ARCS	Australian Research Collaboration Service (part of the Australian Government's National Collaborative Research Infrastructure Strategy)
ATSE	Academy of Technological Sciences and Engineering
BPEL	Business Process Execution Language is an OASIS standard executable language for specifying interactions with Web Services
CEC	Commission of the European Communities
CIO	Chief Information Officer
CLuE	Cluster Exploratory Program (of NSF)
CSIRO	Commonwealth Scientific and Industrial Research Organisation
DCMS	Department for Culture, Media and Sport (UK)
DIISR	Department of Innovation, Industry, Science and Research
EC2	(Amazon's) Elastic Computer Cloud
ENISA	European Network and Information Security Agency
FTSE	Fellow of the Academy of Technological Sciences and Engineering
IaaS	Infrastructure as a Service
IPR	Intellectual Property Rights
IT	Information Technology
MARCS	Members of the ARCS
NCI	National Computational Infrastructure
NCRIS	National Collaborative Research Infrastructure Strategy
NICTA	National Information and Computer Technology Australia
NIST	National Institute of Standards and Technology (USA)
NRIC	National Research Infrastructure Council
NSF	National Science Foundation (USA)
OASIS	Organization for the Advancement of Structured Information Standards
OMB	Office of Management and Budget (USA)
OSS	Open Source Software
PaaS	Platform as a Service
PfC	Platforms for Collaboration (an element of NCRIS)
QUT	Queensland University of Technology
ROI	Return on Investment

CLOUD COMPUTING

SaaS	Software as a Service
SLA	Service Level Agreement
SMEs	Small and Medium Sized Enterprises
XML	Extensible Markup Language – a set of rules for encoding documents in machine-readable form.

Bibliography

- Anderson, J Q and L Rainie, 2010, *The future of cloud computing*, Pew Research Centre, accessed on 22 June 2010 at http://pewinternet.org/~media//Files/Reports/2010/PIP_Future_of_the_Internet_cloud_computing.pdf
- Armbrust, M, A Fox, R Griffith, A D. Joseph, R H. Katz, A Konwinski, G Lee, D A. Patterson, A≈Rabkin, I Stoica and Mi Zaharia, 2009, *Above the Clouds: A Berkeley View of Cloud Computing*, Technical Report No. UCB/EECS-2009-28, accessed on 22 June 2010 at <http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html>
- Bell, G, T Hey and A Szalay, 2009, *Beyond the Data Deluge*, *Science* 323, 1297.
- Bethwaite, B, D Abramson, F Bohnert, S Garic, C Enticott and T Peachey, *Mixing Grids and Clouds: High-throughput Science using the Nimrod Tool*, in *Cloud Computing: Principles, Systems and Applications*, Springer, to be published August, 2010.
- Cloud Security Alliance, 2009, *Security Guidance for Critical Areas of Focus in Cloud Computing V2*, December 2009, accessed on 22 June 2010 at <http://www.cloudsecurityalliance.org/csaguide.pdf>
- Commission of the European Communities, 2010, *The Future of Cloud Computing: Opportunities for European Cloud Computing Beyond 2010*, Brussels, accessed on 22 June 2010 at <http://cordis.europa.eu/fp7/ict/ssai/docs/cloud-report-final.pdf>
- Department of Broadband, Communications and the Digital Economy, 2009, *Realising our broadband future*, Conference held December 10-11 2009, details accessed on 22 June 2010 at <http://www.broadbandfuture.gov.au/>
- Dean, J, 2009, *Challenges in Building Large-Scale Information Retrieval Systems*, keynote presentation to the Second ACM International Conference on Web Search and Data Mining (WSDM09), Barcelona Spain, accessed on 22 June 2010 at http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en//people/jeff/WSDM09-keynote.pdf
- Dean, J and S Ghemawat, 2004, *MapReduce: Simplified data processing on large clusters*. In *Proc of Operating Systems Design and Implementation (OSDI)*. San Francisco, CA. 137-150 access on 22 June 2010 via <http://labs.google.com/papers/mapreduce.html>
- Department for Culture, Media and Sport (DCMS) and Department for Business, Innovation and Skills (DBIS), 2009, *Digital Britain: Final Report*, London, accessed on 22 June 2010 at <http://webarchive.nationalarchives.gov.uk/+http://www.culture.gov.uk/images/publications/digitalbritain-finalreport-jun09.pdf>
- Department of Innovation, Industry, Science and Research, 2010, *Super Science – Research data storage infrastructure; Implementation Discussion Paper*, accessed on 22 June 2010 at <http://www.innovation.gov.au/Section/science/Documents/DataStorageDiscussionpaper13April2010FINAL.pdf>

- European Network and Information Security Agency 2009a, Cloud computing: Benefits risks and recommendations for information security, access on 22 June 2010 at <http://www.enisa.europa.eu/act/rm/files/deliverables/cloud-computing-risk-assessment>
- European Network and Information Security Agency 2009b, *An SME perspective on Cloud Computing, Survey*, accessed on 22 June 2010 at http://www.enisa.europa.eu/act/rm/files/deliverables/cloud-computing-sme-survey/at_download/fullReport
- Francis, R, 2009, Presentation, Meeting 13, 10 December 2009, accessed on 22 June at https://www.pfc.org.au/pub/Main/AeRIC-13/Item_5_eResearch_Issues.pdf
- Ghemawat, S, H Gobioff, and S-T Leung, 2003, *The Google File System*, 19th ACM Symposium on Operating Systems Principles, Lake George, NY, October, 2003, accessed on 22 June 2010 at <http://labs.google.com/papers/gfs.html>
- Government 2.0 Taskforce, 2009, *Engage: Getting on with Government 2.0, Report of the Taskforce*, December 2009, Department of Finance and Deregulation, accessed on 22 June 2010 at <http://www.finance.gov.au/publications/gov20taskforcereport/index.html>
- Hamilton, J, 2008, *Internet-Scale Service Efficiency*, presentation to LADIS 2008 conference, accessed on 22 June 2010 at http://mvdirona.com/jrh/TalksAndPapers/JamesRH_Ladis2008.pdf
- Hey, T, S Tansley and K Tolle (Editors), 2010, *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Microsoft Research Redmond, Washington, accessed on 22 June 2010 at <http://research.microsoft.com/en-us/collaboration/fourthparadigm/default.aspx>
- IBM, 2010, Cloud Computing accessed on 22 June at http://www.ibm.com/ibm/cloud/ibm_cloud/
- Isard, M, M Budiu, Y Yu, A Birrell and D Fetterly, 2007, *Dryad: distributed data-parallel programs from sequential building blocks*, in Proceedings of the 2007 EuroSys Conference, Lisbon, March 2007, P Ferreira, T R. Gross and L Veiga (Eds), accessed on 22 June 2010 via <http://www.informatik.uni-trier.de/~ley/db/conf/eurosys/eurosys2007.html>
- Japanese Ministry of Internal Affairs and Communications, 2009, *Digital Japan Creation Project (ICT Hatoyama Plan): Outline*, Tokyo March 2009, accessed on 22 June 2010 at http://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/Releases/Topics/pdf/090406_1.pdf
- Katz, R H, 2009, *A computer scientist looks at the energy problem*, paper presented to the 3rd International Universal Communication Symposium (IUC09), Tokyo, 2009 accessed on 22 June 2010 via <http://portal.acm.org/citation.cfm?id=1667780.1667784>
- Luo, L, A Kansal, S Nath and F Zhao, 2008, *Sharing and exploring sensor streams over geocentric interfaces*, Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems, Irvine CA accessed on 22 June 2010 at <http://portal.acm.org/citation.cfm?doid=1463434.1463439>
- Mather, T, S Kumaraswamy, S Latif, 2009, *Cloud Security and Privacy: An Enterprise Perspective on Risks and Compliance*, O'Reilly Media, Inc., Sebastopol, CA. 2009.
- Microsoft Research, 2010, *Cloud Futures 2010: Advancing Research with Cloud Computing*, April 2010, accessed on 22 June 2010 at <http://research.microsoft.com/en-us/events/cloudfutures2010/videos.aspx>

Musaloiu-E, R. A. Terzis, K Szlavec, A. Szalay, J. Cogan, R J Gray, 2006, *Life under your feet: Wireless sensors in soil ecology*, Proceedings of the Third Workshop on Embedded Networked Sensors (EmNets 2006) May 2006 Harvard University, Cambridge, Massachusetts, accessed on 22 June 2010 at <http://hinrg.cs.jhu.edu/uploads/Main/lifeunderyourfeet06.pdf>

Mell, P and T Grance, 2009, *The NIST Definition of Cloud Computing*, Version 15, October 2009, National Institute of Standards and Technology, Information Technology Laboratory, accessed on 22 June 2010 at <http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc>

National Science Foundation, 2008, *A CluE in the Search for Data-Intensive Computing*, Press Release 08-066, accessed on 22 June 2010 at http://www.nsf.gov/news/news_summ.jsp?cntn_id=111470

National Science Foundation, 2009, *National Science Foundation Awards Millions to Fourteen Universities for Cloud Computing Research*, Press Release 09-082, accessed on 22 June 2010 at http://www.nsf.gov/news/news_summ.jsp?cntn_id=114686

National Science Foundation, 2010, Microsoft and National Science Foundation Enable Research in the Cloud, Press Release 10-023, accessed on 22 June 2010 at http://www.nsf.gov/news/news_summ.jsp?cntn_id=116336

Office of Management and Budget, 2009, *Crosscutting Programs*, Washington DC, accessed on 22 June 2010 at <http://www.whitehouse.gov/omb/budget/fy2010/assets/crosscutting.pdf>

Organisation for Economic Cooperation and Development, 2009, *Briefing paper for the ICCP Technology Foresight Forum: Cloud computing and public policy*, DSTI/ICCP(2009)17 prepared by Prof M R Nelson, accessed on 22 June 2010 at <http://www.oecd.org/dataoecd/39/47/43933771.pdf>

Petrov, O, 2009, *Financial Crisis and Cloud Computing: Delivering More for Less: Demystifying Cloud Computing as Enabler of Government Transformation*, World Bank Government Transformation Initiative, June 2009, accessed on 22 June 2010 via <http://search.worldbank.org/all?qterm=cloud%20computing>

Ranger, C, R Raghuraman, A Penmetsa, G Bradski, and C Kozyrakis., 2007, *Evaluating MapReduce for Multi-core and Multiprocessor Systems*, Proceedings of the 13th Intl. Symposium on High-Performance Computer Architecture (HPCA), Phoenix, AZ, February 2007, accessed on 22 June 2010 at <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.137.2453>

Recordkeeping Innovation Pty Ltd, 2007, Project 9 Report: Preservation of Web 2.0 content, accessed on 22 June 2010 at <http://gov2.net.au/projects/project-9/>

Suh, B, E H Chi, B A Pendleton and A Kitur, 2007, *Us vs. them: Understanding Social Dynamics in Wikipedia with Revert Graph Visualizations*, paper presented to IEEE Symposium on Visual Analytics Science and Technology, October/November 2007, Sacramento, CA, vied on 22 June 2010 at <http://sites.google.com/site/ascparc/2007-10-28-VAST2007-RevertGraph-Wiki.pdf?attredirects=0>

Virtual Ark, 2010, accessed on 22 June at www.virtualark.com

Wang, C, S Chen and J Zic, 2009, *A Contract-based Accountability Service Model*, presented at IEEE 7th International Conference on Web Services (IEEE ICWS 2009); Los Angeles, CA.

Williams, T, 2009, Australian Research Collaboration Service 2009/10 Annual Business Plan for the Interoperation and Collaboration Infrastructure.

Wintergreen Research 2010, Worldwide Cloud Computing Market Opportunities and Segment Forecasts 2009 to 2015, summary accessed on 20 June 2010 at

<http://www.wintergreenresearch.com/reports/CloudOpportunities.htm>

Wyld, D C, 2010, *The cloudy future of government IT: Cloud computing and the public sector around the world*, International Journal of Web & Semantic Technology (IJWesT), Vol 1, Num 1, January 2010, accessed on 22 June 2010 at

<http://airccse.org/journal/ijwest/papers/0101w1.pdf>

Yu, Y, M Isard, D Fetterly, M Budiu, Ú Erlingsson, P Kumar Gunda and J Currey, 2008, *DryadLINQ: A System for General-Purpose Distributed Data-Parallel Computing Using a High-Level Language*, OSDI'08 Eighth Symposium on Operating Systems Design and Implementation, accessed on 22 June 2010 at

http://www.usenix.org/events/osdi08/tech/full_papers/yu_y/yu_y_html/

ATSE – in brief

The Academy of Technological Sciences and Engineering (ATSE) is an independent, non-government organisation, promoting the development and adoption of existing and new technologies that will improve and sustain our society and economy.

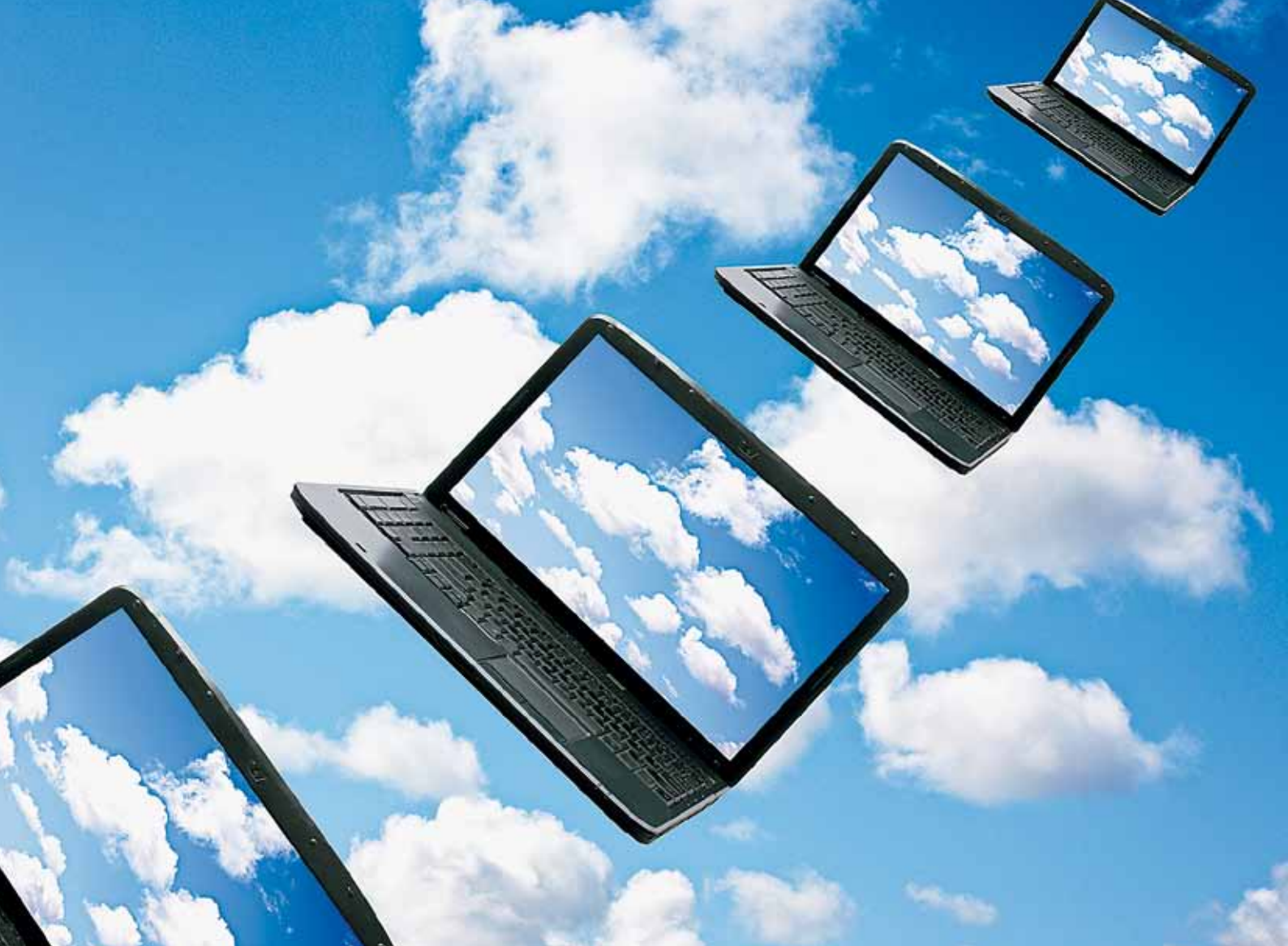
ATSE consists of some 800 eminent Australian Fellows and was founded in 1976 to recognise and promote the outstanding achievement of Australian scientists, engineers and technologists.

ATSE provides a national forum for discussion and debate of critical issues about Australia's future, especially the impact of science, engineering and technology on quality of life.

ATSE links Australia with leading international bodies and worldwide expertise in the technological sciences and engineering.

ATSE fosters excellence in science, engineering and technology research and the critical education systems that underpin Australia's capacity in these areas.

ATSE tackles many of the most difficult issues governing our future, by offering fresh ideas, practical solutions and sound policy advice – and putting them on the public record.



CLOUD COMPUTING: OPPORTUNITIES AND CHALLENGES FOR AUSTRALIA
Report of a Study by the Australian Academy of Technological Sciences and Engineering (ATSE)

© Australian Academy of Technological Sciences and Engineering

The Australian Academy of Technological Sciences and Engineering (ATSE)

ATSE Office

Level 1, 1 Bowen Crescent
Melbourne Victoria 3004
AUSTRALIA

Mail address

GPO Box 4055
Melbourne Victoria 3001
AUSTRALIA

Phone

+61 (0)3 9864 0900

Fax

+61 (0)3 9864 0930

Email

vaughan.beck@atse.org.au
bill.mackey@atse.org.au

Websites

www.atse.org.au
www.crawfordfund.org
www.cluniesross.org.au