Cloud and Web 2.0 Services for Supporting Research

Rob Allan
Computational Science and Engineering Department,
STFC Daresbury Laboratory, Daresbury, Warrington WA4 4AD

Contact e-Mail: `robert.allan@stfc.ac.uk`

January 5, 2012

Abstract

"Cloud computing" can be defined as the flexible provision of computing power, applications, and data storage by a networked pool of hardware resources. In cloud computing, such resources are delivered to users as services.

In this report we present a discussion and analysis of the availablility and uses of cloud services for supporting research. This includes things like e-Mail, Web hosting and data storage for research led organisations, in addition to running applications. Much of the subsequent discussion focusses on issues that should be considered before engaging a cloud service provider.

Whilst "public cloud" services like Amazon EC2 have been widely discussed, there is now a growing interest in "private or partner clouds" which enable more control and re-use of existing skills and resources but nevertheless providing many of the same business advantages.

Contents

1 Introduction

In this report we present a discussion and analysis of the availablility and uses of cloud services for supporting research. This includes things like e-Mail, Web hosting and data storage for research led organisations, in addition to running applications.

"Cloud computing" can be defined as the flexible provision of computing power, applications, and data storage by a networked pool of hardware resources. In cloud computing, resources are delivered to users as a service.

Commercial cloud services offer a "utility" model of computing where individuals do not have to invest in hardware and can instead buy or rent compute cycles and storage capacity from service providers. Costing models vary, but are typically by CPU hours. Components of commercial clouds include "Platform as a Service" and "Software as a Service". In Platform as a Service (PaaS), an layer of capability is provided on top of the basic infrastructure to allow users to develop bespoke applications to run on the cloud. In Software as a Service (SaaS), the software is pre-installed remains the property of the provider and access is provided by subscription or on a pay-per-use basis.

Alternatives to commercial (public) clouds are private or partner clouds, formed from pooled resources within the closed infrastructure of a single or group of organisations. The potential benefits and challenges in commercial and private clouds are different.

Clouds don't have a silver lining, they just make it hard to see where you're going. That would be the opinion of someone engaged in traditional high performance computing research activities.

It has also been noted that *cloud computing is a new way of delivering computing resources, not a new technology* [23].

Web 2.0, whatever that is, is however now proposed (mostly by commercial suppliers hosting services) for almost everything: on-line information, buying, trading, voice communication, publishing, sharing data and full collaboration. Providers include: Yahoo!, Google, Amazon, e-Bay, Skype, SalesForce, mySpace, YouTube, DropBox, etc.

Mell and Grance at NIST have provided the most widely accepted definition of cloud computing [15]. We will use their terminology in this document.

Google arguably hosts the biggest set of services which are now being referred to as "Cloud". Google Search enables over 1 billion searches per day, YouTube has over 20 hours of video uploaded per minute. Gmail is used by millions of people. Google has also become the fourth largest server manufacturer in the world.

With this growth in the industry, it is not surprising that enterprises, both commercial and academic, are looking at how savings can be made by out sourcing some of their services to cloud providers. Percieved advantages include: economy of scale, resilience by design, no need to deal with complexity, agility of "versionless" software (perpetual beta), green (lower your carbon footprint), multiple servers (fast response), multi-tenancy (balances workload).

Migrating legacy applications to the cloud is however not something to be done lightly. It takes a real understanding of your existing systems, a disciplined process for the change management itself,

and the ability to secure both data and access to these systems once they are migrated. Another consequence is that it lowers the IT skills that are required in house, indeed a completely different set of skills are likely to be required which have more to do with managing external contracts, see below.

How are clouds related to grids? They're probably not. Clouds could help to simplify and optimise grid site operations and grid middleware can be used in a transparent way on top of virtualised computing resources, bringing about the development of virtual grid infrastructures. This is still a research topic in itself, see [18]. Some so-called research clouds are really grids with a simplified access layer and a resource broker.

1.1 What are people using Now?

Uptake of services on line is largely driven by the community. This is particularly true of social networking services which groups of peers flock to. The appearance of a new service may mean a mass movement, but the lack of open standards could mean their data and former identities are left stranded. Some examples of the hosted services currently being used by the research community (in the UK) are as follows.

Analytic Bridge: a social networking site for people interested in analytics, e.g. statistical computing. `http://www.analyticbridge.com`

DropBox: used to share and sync files across multiple systems. The syncing mechanism has also been used to invoke tasks, such as grid job submission, but running a deamon which detects the upload of a file or script. For instance this was evaluated by Ian Cottam in Manchester. `http://www.dropbox.com`.

Gmail: `http://www.google.com`

Google Calendar: `http://www.google.com`

Google Maps: e.g. CASA/ NeISS. `http://www.maptube.org/`

Google Search: `http://www.google.com`

Huddle: uses a London based cloud to offer data storage and collaboration, so under easier to meet the requirements of the UK Data Protection Act. `http://www.huddle.net`

JISCMail: is a group based mail and list server which also supports archives, discussion (chat), files, meetings, surveys and newsletters, see `http://www.jiscmail.ac.uk`. MailTalk is also available for commercial users.

myExperiment: developed with RCUK and JISC funding, this is a social networking site for sharing "research objects". It is particularly use by the bio-informatics community sharing workflows. `http://www.myexperiment.org`

NGS Portal: HPC Software as a Service, storage and data management. Also provides mechanisms to allow people to share information about how to run grid jobs. `http://portal.ngs.ac.uk`

Sakai: example of portal or private cloud for collaboration and data sharing. Out of the box has resource folders (files), discussion, chat, wiki, blog, calendar and many more tools for collaboration and education. Can be enhanced with other research services, e.g. for NeISS, National e-Infrastructure for Social Simulation. `http://portal.ncess.ac.uk`

Twitter: e.g. CASA/ NeISS, JISC inf11, DisCo. `http://www.casa.ucl.ac.uk/tom/`. A Twitter feed can be used to update news items on a Web site, e.g. `http://www.cse.scitech.ac.uk/disco/` so could replace RSS for that purpose – there are tools which will convert between them, however there are big differences

Wikipedia: `http://www.wikipedia.org`

YouTube: a subsidiary of Google featuring short on line video clips for fun, education or publicity. Some institutions have YouTube channels, e.g. `http://www.youtube.com/user/SciTechUK` or `http://www.youtube.com/user/EPSRCvideo`

Great New Thing: not being used yet, but 100% sure it will be when available in beta form. Need to use open standards to import identities, social networks and data...

Clouds for Storage and Data Management

Examples – DropBox, Amazon S3, Yahoo! are beginning to feature in this space through their involvement with Hadoop, Sherpa and OpenCirrus.

Key concerns are data protection, access and integrity.

Clouds for Collaboration

Examples – Google, Huddle, Sakai, JISCMail.

Key concerns are identify and role management and data protection.

Clouds for Computing

Examples – Amazon EC2, Microsoft Windows Azure, Penguin on Demand (POD).

A key concern for HPC applications is performance. Penguin claim to have addressed this by removing virtualisation, something expected in other clouds.

1.2 Sources of Information

There is a very good introductory article on Wikipedia at `http://en.wikipedia.org/wiki/Cloud_computing`.

Two reports commissioned by JISC [11, 21].

EU FP7 expert group report [18].

European Network and Information Security Agency (ENISA) analysis of the benefits, risks and recommendations for information security in cloud computing [23].

CloudHousting UK Web site `http://www.cloudhosting.co.uk` contains news and general articles plus comments, blog and forum for feedback.

Cloud Computing Portal `http://cloudcomputing.qrimp.com/portal.aspx` is a source of information about cloud services vendors. It lists over 100 such vendors.

CloudReview.org `http://www.cloudreview.org` is a blog site with citations.

The Cloud Tutorial `http://www.thecloudtutorial.com/`.

HPC Cloud Weekly from Tabor Communications `http://www.hpcwire.com`. You probably have to subscribe to get it. See for example `http://www.hpcwire.com/specialfeatures/cloud_computing/`.

e-Science Institute Theme `http://www.research3.org`.

IBM DeveloperWorks `http://www.ibm.com/developerworks/cloud/`.

"Powered by Cloud" conferences `http://www.poweredbycloud.com/programme/programme.aspx`.

RCUK Workshop on Cloud Computing 20/6/2010.

2 Cloud Services

In the cloud, details are abstracted from the users who no longer have need for expertise in, or control over, the technology infrastructure that supports them. Cloud computing describes a new supplement, consumption, and delivery model for IT services based on the Internet and typically involves over provision of dynamically scalable and virtualised resources.

The term cloud is often used as a metaphor for the Internet. Most cloud computing infrastructure consists of services delivered through common centres and built on hosted servers. Typical cloud providers deliver common business applications on line that are accessed from another Web service or software like a Web browser, whilst the software and data are stored on servers which they host. Clouds often appear as single points of access for all IT services. Commercial offerings are generally expected to meet quality of service (QoS) requirements of customers, and typically include SLAs. This requires management.

In the cloud, almost everything is described as a service. ENISA [23] have considered the division of responsibility for security related factors between customer and supplier in each of these categories. We consider their conclusions for SaaS further below. The most common categories are as follows.

SaaS: Software as a Service

Software is pre-installed and available as a "turnkey" service via the Internet – relevant for well established applications rather than for development. With SaaS, a provider licenses an application to customers as a service on demand, through a subscription or a pay as you go model. Saas is also called software on demand. SaaS was initially widely deployed for sales force automation and customer relationship management (CRM). Now, it has become commonplace for many businesses tasks, including computerised billing, invoicing, human resource management and service desk management.

The pioneer in this field was Salesforce.com offering on-line CRM. Other examples are on-line email providers like Google's Gmail and Microsoft's HotMail, Google docs and Microsoft's on-line version of office called BPOS (Business Productivity On-line Standard Suite). Zoho offer a range of on-line applications which can be integrated with Google.

PaaS: Platform as a Service

Provides a platform and software stack against which applications can be built. Leading examples are are Googles Application Engine, Microsoft's WIndows Azure, Amazon's Web Services, Salesforce.com.

IaaS: Infrastructure as a Service

Typically provides a virtualised infrastructure. Equivalent to "selling cycles". Leading vendors that provide IaaS are Amazon EC2 (Elastic Computing Cloud), Amazon S3, Rackspace Hosting and Flexiscale.

StaaS: Storage as a Service

In addition to the three service models above which were identified by Mell and Grance [15], providers such as Amazon (with S3, Simple Storage Service), AT&T, GoGrid, Rackspace and DropBox also offer storage. Many offer an initial amount of free space, say 10-50 GB, and charge for usage above that level. Some offer incentives for signing up new members.

Client Layer

Access for managers and users is typically, but not necessarily, delivered via a Web browser. Use of the Web has the advantage that services are "pervasive". In addition, a command line interface or some form of terminal, e.g. using VNC, might be provided. All these solutions will give the user a customised environment potentially with turn key applications and/ or a virtual server or cluster. Web services might be used as the underlying integration layer and TLS for security. There might be an issue with bulk data transfer for which other solutions can be provided.

3 Grid vs. Cloud

The name Grid or Cloud could be chosen. They are similar in meaning but subject to differences of interpretation. Cloud has been associated with the provision of services "on demand" as offered by Google, Amazon, Microsoft, etc. who host massive server farms for this purpose. Our meaning is clearly different, but we could use the same term to capture all the services offered over the campus ICT infrastructure as described here.

Cloud computing gained attention in 2007 as it became a popular solution to the problem of horizontal scalability [24]. Our use of Cloud computing naturally evolves from our experience of NW-GRID and the Campus Condor pools and portals described above.

Matching technology to applications.

The purpose of a Cloud interface for DSIC is to allow systems as described above to be introduced and removed dynamically and made accessible independently of where they are physically located,

e.g. they may be at vendor sites or university partners sites such as NW-GRID or located on the Campus, such as in the Tower, the main Computer Room or the Cockcroft Institute. The interface should allow access to a rich variety of computing and storage systems.

The term Cloud Computing derives from the common depiction in most technology architecture diagrams, of the Internet or IP availability, using an illustration of a cloud. The computing resources being accessed are typically owned and operated by a third party provider on a consolidated basis in data center locations, in our case typically somewhere on the Campus. Target consumers are not concerned with the underlying technologies used to achieve the increase in server capability. The Cloud simply provides services on demand. In our case however consumers will be concerned with the architecture and will target their applications to the most appropriate system available at the time, usually to get best performance. Grid computing is a technology approach to managing a Cloud, and one with which we have a lot of experience building on NW-GRID and projects such as eMinerals [20]. In effect, all clouds are managed by a Grid but not all Grids manage a Cloud. More specifically, a Compute Grid and a Cloud are synonymous, while a Data Grid and a Cloud can be different. We also use the term Campus Grid through which we extend the Cloud to cover pools of desktop systems possibly using novel scheduling algorithms such as using spare cycles and back fill. We could also refer to this as Integrated Computing.

Critical to the notion of cloud computing is the automation of many management tasks. If the system requires human intervention to allocate tasks to resources it's not a Cloud.

A compute cluster can offer cost effective services for specific applications, but may be limited to a single type of computing node with all nodes running a common operating system. Alternatively, the canonical definition of Grid is one that allows any type of processing engine to enter or leave the system dynamically. This is analogous to an electrical power grid on which any given generating plant might be active or inactive at any given time. This can be achieved by physically connecting or removing distributed servers or by virtualisation. Since we support many "heritage" applications which are of the traditional MPI parallel type we will keep the notion of clusters and currently support physical rather than virtual resource dynamics. This can however include dual booting of certain servers.

Potential advantages

Potential advantages of any Cloud or Grid computing approach include:

- Location of infrastructure in areas with lower costs of real estate and electricity. We use several buildings;
- Sharing of peak load capacity among a large pool of users, improving overall utilization;
- Separation of infrastructure maintenance duties from domain specific application development;
- Separation of application code from physical resources;
- Ability to use external assets to handle peak loads (not have to design for highest possible load levels). We use resources from other NW-GRID sites;
- Not have to purchase assets for one time or infrequent intensive computing tasks. We can implement rolling upgrades as capital funds become available.

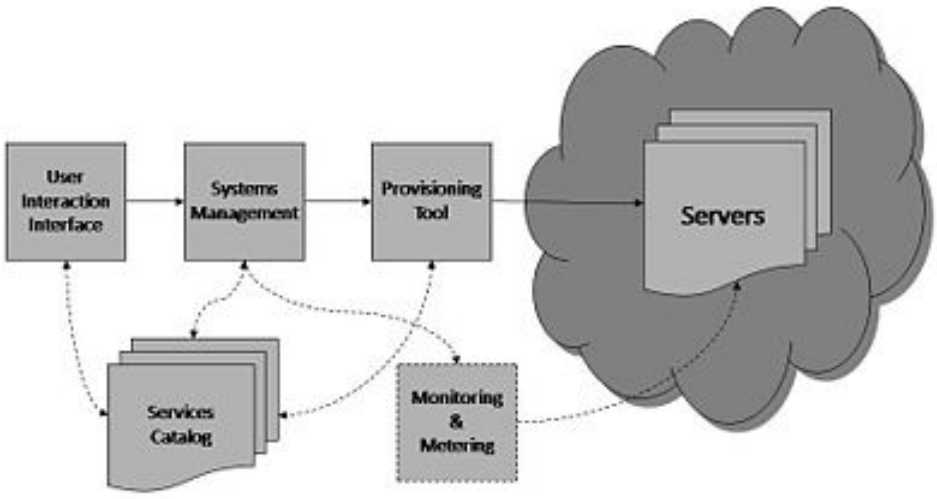

Figure 1: General Cloud Computing Architecture

Architecture

The architecture behind cloud computing, see Figure 1, is a massive network of "cloud servers" interconnected as in a Grid. Virtualisation could be used to maximize the utilisation of the computing power available per server, e.g. to better match the overall workload.

A front end interface such as a Portal allows a user to select a service from a catalogue. This request gets passed to the system management which finds the correct resources and then calls the provisioning services which allocates resources in the Cloud. The provisioning service may deploy the requested software stack or application as well, e.g. via licensing on-demand.

- User interface (Portal or desktop) – this is how users of the cloud interface with the underlying Grid to request services;
- Services catalogue – this is the list of services that a user can request;
- System management – this is the piece which manages the computer resources available;
- Provisioning tool – this tool allocates the systems from the Grid to deliver on the requested service. It may also deploy the required software;
- Monitoring and metering – this optional piece tracks the usage of the Grid so the resources used can be attributed to a certain user;
- Servers – the servers are managed by the system management tool. They can be either virtual or real.

We have considered the use of MOAB from Cluster Resources for some of the above tasks [7].

Cloud storage

Cloud storage is a model of networked data storage where data is stored on multiple virtual servers, generally hosted by third parties, rather than being hosted on dedicated servers. Hosting companies operate large data centers, and people who require their data to be hosted buy or lease storage capacity from them and use it for their storage needs. The data center operators, in the background, virtualise the resources according to the requirements of the customer and expose them as virtual servers, which the customers can themselves manage.

We have achieved this in the past using SRB, the Storage Resource Broker from SDSC [4], which provides a virtual file system interface to distributed storage "vaults". Physically, the resource may thus span across multiple servers. In our case storage services are provided for users of DSIC compute resources and other local initiaties such as POL and NW-VEC, e.g. via NW-GRID. We note that SRB will in the future become iRODS and that other solutions, such as AFS, are available.

4 Public Clouds

There are a number of use cases where public clouds might have a role to play in the research life cycle. In all these security and identity management are common requirements. Federated identity will be a common requirement wherever multiple researchers are concerned and is one reason solutions like the UK Access Management Federation are being used for JISC and other academic services. Some perceived advantages to using a public cloud are as follows.

meet short term requirements: – avoid need to buy a solution;

infrequent use, or no desire to maintain infrastructure: – pay for useage on demand;

cloud bursing: – additional capacity is required on demand at specific times or to cope with unpredictable but temporary peaks in usage;

transfer to commerce: – services being made accessible to commercial partners so cannot be hosted on academic systems;

agile: – flexibility and avoiding effects of system or software upgrades;

data hosting and backup: – using redundancy in the cloud;

research publications: – complement reports with ability to re-run scenarios in a virtual machine;

ad hoc **support activities:** – may be outside local policy limitations, e.g. running an on-line survey or sharing large image files.

4.1 What is Available

Cloud hosting services in the UK include those from the following companies.

Logicalis: `http://www.uk.logicalis.com` in July 2011 announced a shared research cloud directly attached to the JANET network. Logicalis is a UK company based in Slough with offices in

several other countries. The deploy private cloud solutions and also offer hosted cloud services with partners Cisco, HP, IBM, NetApp and CA.

RackSpace: `http://www.rackspace.co.uk` offer public, private and hybrid solutions (formerly known as Mosso)

CloudLab: `http://cloudlab.co.uk`

OutSourcery: `http://outsourcery.co.uk`

Hyve: `http://www.hyve.com` uses HP and Cisco hardware and VMWare virtualisation

RapidCloud: `http://www.rapidcloud.co.uk`

ElasticHosts: `http://www.elastichosts.com/`

UK2.net: (re-branded service from VPS.net) `http://www.uk2.net/virtual-private-servers/` or `http://vps.net/`

Vmhosts: (iomart group) `http://www.vmhosts.co.uk/` or `http://www.iomart.com/cloud_hosting.php`

Amazon: features "cloud front" from fast localised content distribution anywhere in the world including UK. Amazon is currently the most used cloud system world wide. `http://aws.amazon.com`

Google App Engine: that basically auto-tunes its geo-localisation to match the one of the users `http://code.google.com/appengine/`

Windows Azure: (by Microsoft), not released yet, but geo-localization is already offered, and Europe (Ireland at first) will probably be provided in 2010, see Venus-C project below. See `http://www.microsoft.com/windowsazure/` (site did not respond when I last tried it).

OnApp: A London based company offering Cloud software `http://onapp.com/`

Others are described on the CloudHousting UK Web site `http://www.cloudhosting.co.uk`.

4.2 How to use Them 1 – Huddle Case Study

Huddle is a UK based company established in London in 2006. The product is a hosted site for collaboration in business. Huddle is now used by worldwide companies such as Panasonic, Kia Motors, Nokia, Unilever, Kerry Foods, P&G and charities such as UNICEF plus UK and European universities, e.g. UCL and Birmingham and government organisations such as NHS and the Home Office. Huddle also has offices in San Francisco and has recently established a partnership with HP. Huddle interfaces with Microsoft SharePoint.

The Huddle interface looks like a Web portal. It has a dashboard providing access to the main features: tasks, files, calendar (with iCal interface), notifications, news. Additional features of project work spaces include meeting setup, Web conferencing (including shared desktop), discussions, whitboard, teams (with contact details), search, social networks (e.g. LinkedIn), apps and Microsoft Office plugin. Workflows can be implemented to manage processes.

4.3 How to use Them 2 – CoP Platform Case Study

Communities of Practice for Local Government (CoP) is a hosted site provided by Local Government Improvement and Development, part of the LGA Group. CoP Platform is a community platform which supports professional social networking, collaboration and the sharing of information and ideas across local government, the public sector and those working in public service improvement in the UK. A collection of on line facilities and services is provided including discussion forums, blogs, wikis, news feeds and a search facility (known as People Finder) allowing users who have registered to use the CoP Platform to search for and contact peers, advisers and other practitioners who are also users.

Terms and conditions include the following recommendation which could apply to most hosted solutions. *We will use reasonable endeavours to ensure that the CoP Platform is accessible 24 hours per day but the CoP Platform is provided on an "as is" and "as available" basis, and we give no warranties or guarantees that the CoP Platform will meet particular levels of availability or functionality. Therefore, we strongly recommend that you do not post any business critical information or material on the CoP Platform and that you keep copies of all information and content you post on the CoP Platform in accordance with your employer's policies and processes.*

4.4 How to use Them 3 – Australian Research Collaboration Service

A recent example for HPC is the ARCS SaaS Compute Cloud in Australia. The Australian Research Collaboration Service (ARCS), the national provider of inter-operable and collaborative e-Research services, announced the national release of the ARCS Compute Cloud in July 2010, see `http://www.arcs.org.au`. ARCS is a joint venture capital company running a sub-programme of NCRIS, the National Collaborative Research Infrastructure Strategy. The Compute Cloud simplifies using the Australian grid, which is managed by ARCS and networks many of the country's high performance computers. It aims to provide easy access to HPC and complements the Grisu grid submission service which enables more control over which local and remote resources are used.

The ARCS Compute Cloud lets researchers carry out fast analysis of large and complex data by using a number of pre-installed common HPC applications. Its graphical interface tailored to the application is simple to use and enables researchers to submit jobs quickly without requiring extensive technical expertise. This acts as a resource broker and locates an available resource on the grid with the required application. In addition, the service allows users to have a single account that provides seamless access to the compute clusters efficiently, regardless of their location or institutional affiliation. A simple quota system is implemented with time pre-allocated to users.

There is still a disconnect between ARCS and other local HPC providers in universities, much the same situation as in the UK when we compare services provided by HPC-SIG with those of the NGS and its UI-WMS resource broker, see `http://www.ngs.ac.uk/ui-wms`.

4.5 How to use Them 4 – Venus-C Infrastructure

Venus-C, Virtual multi-disciplinary EnviroNments USing Cloud Infrastructures, is one of a number of advanced computing projects currently receiving funding as part of the European Commission's

7th Framework Programme. Its main objective is to demonstrate the feasibility and potential of a pan-Europe scientific cloud that is integrated with the existing European grid system. See `http://www.venus-c.eu/`. In many respects this is comparable to the ARCS and NGS activities (in fact NGS is now part of EGI). It differs from these by using Windows and virtualisation.

The project places a strong emphasis on building a user community, and aims to create, test and deploy an industry quality, service oriented platform based on virtualisation technologies, accessible by researchers across many disciplines. An open call will be issued in late 2010 to broaden its applications and geographical scope. This will fund up to 20 new experiments designed to address the advanced and complex needs of the user communities, in some instances handling complex workflows and data intensive scenarios.

Microsoft is a major partner in, and initiator of, the Venus-C consortium indicating the level of attention being paid to developments in this area. The company's contribution to the project is a substantial Windows Azure data and compute capability, as well as teams of researchers, including one based at the European Microsoft Innovation Centre in Germany.

4.6 How to use Them 4 – Logicalis (UK) Shared Research Cloud

The Logicalis service is located in the UK and built on the IBM POWER-7 platform. It is designed to enable communities of researchers to pool their funding by buying a service on a shared HPC system. The Allocations are met dynamically in real time and each institution can use up to 100% of the pool.

This is one of a range of academic services being provided by Logicalis. They recently announced availability of an Intel cloud platform, and there are plans to launch a research collaboration service in Autumn 2011. This will enable researchers to work more collaboratively on projects both internally and within the wider research eco-system. Institutions can also purchase their own rack(s) which cn be housed in the Logicalis data centre with traffic directed ovre JANET.

See `http://www.uk.logicalis.com/news-and-events/news/shared-research-cloud-on-janet.aspx`

4.7 Cost Models

The major component of the cost model for using cloud services is the trade off between local procurement and maintenance and "pay as you go". Capital expenditure is replaced by something equivalent to rental. Both utility or subscription based billing is available. Indeed IT services are now treated in a similar way to an electricity supply, something that was originally part of the grid vision.

However the principal benefit of converting capex to recurrent is to reduce the barrier to entry and reduce the long term committment. This is not relevant for an enterprise with a large existing in-house system and corresponding expertise.

The benefits and opportunities offered include the following.

Reduced infrastructure costs. Since the user will access resources that are maintained and man-

aged by the service provider, cloud computing has the capacity to cut down infrastructure costs considerably, both in terms of hardware and IT staff costs. Computing is provided as a utility, with users billed typically by CPU hour, storage and bandwidth, thus removing the need for capital investment. For new users the barriers to entry are low, as initial trials cost little. This is an obvious benefit to academic users with more access to resources for consumables than capital. One has to be careful however as there are hidden costs to do with monitoring suppliers and contract management, see below.

Scalability. One of the attractive features of cloud computing is scalability. For example, where charging operates via CPU hours, it is the same cost to rent a 10 node cluster to compute for 40 hours, as it is to utilise a 400 node cluster for one hour. It therefore has the potential to offer high performance computing to researchers who would not otherwise have access. This also has to be considered carefully however, as if you can justify running a 400 node cluster at near to maximum capacity, it is almost certainly cheaper to buy one.

Flexibility. Cloud computing offers considerable flexibility and agility, allowing management of cycles when data flows are uneven, for example in next generation gene sequencing. It can also allow groups that may occasionally need large numbers of cycles to work without needing to purchase high performance computers that may be otherwise be under used. But see note above.

Data sharing. Cloud vendors can provide data facilities, providing alternative strategies for storage, recovery and management of data. Cloud computing also provides potential opportunities in data sharing. Potentially, researchers could place data in clouds and make them accessible for third party use. As software can also be provided through clouds, tools used to interrogate the data can be made available without having to upload them separately. Whilst this is true there are strong legal and commercial reasons why organisations need tight control of their own data, see below.

Competitive pricing. If moving data and software between service providers is relatively straightforward, users may have the opportunity to take advantage of competitive pricing. This applied to raw data – information in a hosted content management system may not be so easy to migrate. Subscriptions or contracts, may also complicate this, see below.

Green computing. Cloud providers may be able to provide similar services using less energy or energy from renewable resources than local provision. This is an effect of critical mass.

4.8 Legal Implications

The cloud model has been criticised by privacy advocates for the greater ease in which the companies hosting the cloud services control, and thus can monitor at will, lawfully or un-lawfully, the communication and data stored between the user and the host company. Instances such as the secret NSA programme, working with AT&T, and Verizon, which recorded over 10 million phone calls between American citizens, causes uncertainty among privacy advocates, and the greater powers it gives to telecommunication companies to monitor user activity. While there have been efforts (such as US-EU Safe Harbor) to "harmonise" the legal environment, providers such as Amazon still cater to major markets, such as the United States and the European Union, by deploying local infrastructure and allowing customers to select "availability zones".

A series of articles in Computing magazine have highlighted legal issues particularly related to protection of business oriented and personal data.

The following notes are taken from [11].

Data protection

The UK Data Protection Act 1998 applies to all personal data, i.e. data that is about a living identified or identifiable individual, irrespective of where he or she lives or works, that is either managed or is held in the UK. For any cloud computing application relevant to a UK based organisation, the Act will apply because the HEI in question is responsible for the processing, i.e. addition, deletion, editing, manipulation or dissemination, of the personal information. This applies even if the actual processing takes place in another country, or indeed in several countries, some of which may or may not be known, as is typical for cloud applications.

The Act imposes on the data controller (a legal term which means the organisation) and on any sub-contractor used by the data controller, i.e. the cloud computing provider, certain obligations. It is a breach of the Act if the organisation fails to fulfil its obligations, or if the organisation fails to impose those obligations on its sub-contractors. This applies wherever the sub-contractors are based and whatever legislative environment they happen to work in. The best way to achieve it is to have a clause in the agreement with the supplier that they shall at all terms observe and obey the requirements of the UK Data Protection Act 1998 while handling personal data belonging to the organisation. An alternative is for there to be an explicit list of obligations, which happen to be those required by the Act, imposed on the cloud service supplier either in the contract or as a schedule to that contract.

Personal data handled by organisations in a research context include material on staff, students, research associates, individuals who happen to be the subject of a research project, and individual contractors, suppliers and partners. The data can range from the most innocuous (e.g. authors' names in a bibliography of a research report, the name of the research associate responsible for particular actions, or the web pages of members of staff) through moderately sensitive (such as e-mails sent and received in connection with the research), through to highly sensitive (such as financial and medical details of individuals, or details of a major research study of law breaking or drug abuse where respondents, who are identifiable, have been assured anonymity). It cannot be stressed too strongly that the degree of sensitivity of the data is irrelevant - all personal data are subject to the Act – but the risk of damage and bad publicity increases with the sensitivity of the data if there is any breach of the Act.

The obligations on the organisation and its cloud computing supplier are the eight data protection principles, enshrined in Schedule 1 of the Act. Organisations should be familiar with them already. They state that personal data: must be processed fairly and lawfully; that it shall be processed only for specified purposes; that the data should be adequate, relevant and not excessive; that it should be accurate and where necessary, kept up to date; that it should not be kept for any longer than is necessary; that the rights of data subjects are paramount (see later); that appropriate technical and organisational measures must be taken to ensure there is no un-authorised processing, loss or destruction of personal data (including no un-authorised accessing by third parties to that data); and that personal data may not be moved to a country or countries with inadequate data protection legislation unless full protection of the data is assured.

The most important of these principles in respect of cloud computing is that the data subject's rights must be respected, the data must be protected against un-authorised disclosure, loss, etc, and that it must not be transferred to a country with inadequate protection in place. These three are considered further below.

Three key Principles

Data subjects, i.e. the individuals who are the subject of the data processing, have the right to inspect the data about them, to know who the data has been disclosed to and where the data has come from, have the right to object to processing of data if they feel it damages them or others, and have the right to sue for any breaches of the Act that has caused them financial damage and/ or distress. Thus, the organisation, and its cloud computing supplier, must be willing and able to provide copies of data to the data subject and to prevent any breach of the Act; they must also keep a record of who has viewed the data (it does not have to be at the level of specific individuals, but broad classes of staff would suffice).

The requirement to respond to data subject requests within a tight time frame is well known in larger organisations and there are well established mechanisms for responding, but the cloud computing supplier may not be familiar with them and might be unable or un-willing, for example, to respond to a query from a data subject, or might fail to do so in time. They may also not even recognise a particular request as falling within the Data Protection Act, as the data subject is under no obligation to use the words "Data Protection Act" in any request. This is particularly an issue in respect of organisations in the USA, as there is no Federal Data Protection Act and the companies may not be geared up to responding to requests.

The requirement to prevent un-authorised disclosure, loss, etc. is significant. Whereas it is clearly impossible to guarantee that third parties can never hack into the account (see Information Assurance, below), many cloud computing contracts go beyond this and include clauses where the supplier states that it accepts no liability for any loss or destruction of data. Whilst this approach is very understandable from the cloud service supplier's point of view, it leaves the organisation exposed to risk if it accepts this. The Act requires that the data controller – the organisation – imposes obligations on its sub-contractors as onerous as the obligations imposed by the Act on the data controller itself. Therefore, a standard cloud supplier's waiver clause should ring alarm bells for an experienced organisation.

Finally, the organisation has potential problems regarding the transfer of data to countries with inadequate data protection laws. The USA is a classic example of a country with inadequate laws, but there are many others. To permit this to happen puts the organisation in potential breach of the Act. Since it is difficult to identify where data is held in a cloud application, the organisation has in effect three choices as follows.

1. Insert a clause in the contract that the cloud supplier will abide by all the terms and conditions of the UK Data Protection Act 1998.
2. Insert a series of clauses into the contract specifying the principles that the cloud supplier must follow – these should ideally be worded exactly as in the UK Data Protection Act 1998. One way the supplier could work with this is to offer a "safe harbour". This is a physical site, perhaps in the USA, where the organisation's data will be kept. The supplier would also need to assure the organisation that the data will not be moved elsewhere and agree that the space where the

safe harbour is will follow the UK principles (there are standard contractual clauses for this on the Web).

3. Insert a clause into the contract confirming that the data will only ever be held in the UK and/ or another member state of the European Economic Area – all have adequate laws. In that way, the data is always subject to the Act or its EEA equivalent. This is sometimes known as a "local cloud". The organisation will require a cast iron reassurance that under no circumstances will the data move away from the local cloud.

In summary, current standard cloud computing contracts do not offer sufficient cover for organisations regarding their obligations under the Act. Organisations that fail to incorporate the appropriate clauses into their agreements with cloud suppliers could find themselves facing action for a breach of the Act for the failure to impose appropriate obligations on their outsourcing supplier. Suppliers also need to understand the requirements of the Act if they are to sell their services successfully in the UK and elsewhere in Europe. Although many suppliers have signed up for the US/EU Safe Harbour scheme, unless their compliance with the scheme is made contractual, there remains a significant risk for institutions.

Information assurance (IA)

Apart from the legal data protection issues discussed above, funders, institutions and individual researchers are concerned about the security of their information, although the definitions of security vary. A recent study by the European Network and Information Security Agency (ENISA) provides extensive analysis of the risks and mitigations for cloud computing [23]. Their security assessment is based on three use cases: 1) SME migration to cloud computing services; 2) the impact of cloud computing on service resilience; and 3) cloud computing in e-Government.

Many potential users expect cast iron guarantees that their data cannot be accessed without their authorisation, but it is never possible to give these guarantees. For example, it is reasonable to expect services to protect against common attacks and to not release user data to the internet. But what about skilled and well resourced attackers who might be targeting an organisation? New vulnerabilities are constantly discovered in all elements of the internet, and until they are disclosed, they will be exploitable. The real requirement is to make sure that information is protected proportionately to the risk it is under.

The security arrangements put in place by a cloud provider may or may not be adequate for any particular application or dataset. Potential users should apply good risk management approaches to ensure that their own risk appetite is met. Most cloud providers describe their security approaches publicly, and many have completed some type of external audit. Holding ISO27001/ 27002 accreditation is regarded as an excellent demonstration of good information assurance policy and practice. However there is really no standard at present for cloud security services.

These issues are being addressed by the Cloud Security Alliance (CSA), `http://www.cloudsecurityalliance.org/`. They already provide a number of guidance documents in various languages. v2.1 of the *Security Guidance for Critical Areas of Focus* [6] should be read by everyone considering outsourcing to cloud providers. It describes the overall cloud architecture and potential issues before focussing on twelve separate domains. CSA is already promoting best practice and will move to offering training, certification and accreditation by the end of 2010.

A full treatment of the IA aspects of cloud computing is beyond the scope of the present document. Some common concerns are described below; it is informative to consider issues against the traditional IA dimensions of confidentiality, integrity and availability.

Confidentiality

Confidentiality is usually the first concern expressed by potential users of cloud services, and may be the only concern that has been considered. There is a perception that there is increased risk in transferring data to an external, usually foreign, service provider, where it will be hosted on a system which is used by many other users simultaneously and over which the user has no ownership.

There are undoubtedly some new risks in adopting cloud provision - most obviously, the shift to a hypervised multi-tenant system brings the potential for attacks against the virtualisation layer. If a cloud based virtual server is compromised, conducting forensics can be very hard - it is not possible to simply turn off the machine and recover the disks for analysis.

However, this must be balanced with the concentrations of both risk and expertise within the cloud computing providers. These are specialist service delivery and hosting organisations, which have extensive in house security expertise. Hosting data locally (be it on a personal laptop, departmental server, or university SAN) requires local security expertise that may not be available.

Note that hosting virtual servers with an IaaS provider still requires security expertise - although the shared infrastructure may be secure, the security of the virtual server is largely determined by configuration which is left to the end user.

Integrity

Cloud hosting of data creates new concerns and opportunities for the integrity of data, ensuring that data is not corrupted, either maliciously or accidentally. Cloud providers typically do not conduct backups in the traditional sense, rather they synchronise data between multiple centres. Whereas this helps ensure that integrity is maintained, it does not address issues of long term recovery, which may be required for some audit activities. Historical backups allow the data as it stood at some point in the past to be recovered.

For comprehensive assurance of integrity, it would be necessary to host the same datasets on multiple providers, and locally, and conduct regular bit level comparisons. This degree of re-assurance is much greater than most current provision, and is probably un-necessary and too expensive to implement for the majority of uses.

Availability

It is important to define what availability means for any given task. Availability of compute facilities is typically given as an up time guarantee within a Service Level Agreement (SLA). But the notion of up time might not be adequate to consider the availability of cloud resources. For example, if an institution's up link to the internet fails, cloud services will become un-available to users at the institution. This is outside the control of the cloud service provider, but must be considered. Alternatively, a hosted virtual server may be on line (and therefore "up"), but if a hosted database server is down, or the performance of the server is degraded, whilst still remaining up, the service may be compromised. These availability issues require consideration.

Although these issues are expressed when considering cloud Computing, it is evident that they have often not been carefully considered even for current provision. Few institutional IT services provide an SLA to their users, and we are not aware of any that match the delivered availability of the major cloud providers. The NGS has implemented a system of resource provision based on on SLDs from partner institutions which could be a model for this.

Contract management

It is challenging for any organisation to manage contractual relationships with vendors, particularly when the vendor is very much larger than the organisation itself. Few institutions have the legal and negotiation expertise to contract effectively for mission critical cloud services. These services are relatively new, and their business models are still immature and evolving. Standard contracts are however typically balanced toward the provider and relatively inflexible. It is therefore unlikely that an organisation considering larger scale procurement, for example, buying cloud services centrally for use by multiple researchers, will find much opportunity for variation. Nevertheless, we are aware of one major cloud vendor that has altered its contract for SaaS applications to meet the demands of a UK institution - in particular their requirements under the DPA.

5 Private or Partner Clouds

Whilst "public cloud" services like Amazon EC2 have been widely discussed, there is now a growing interest in "private or partner clouds" which enable more control and re-use of existing skills and resources but nevertheless providing many of the same business advantages.

Note on research competitiveness [2].

A key concept is virtualisation and scalability. Private or partner clouds does not benefit so much from the key cost model characteristics of public cloud, but do admit some economy of scale by sharing resources and policies. Open source software over existing infrastructures are typically used. Such software includes Eucalyptus and OpenNebula as will be described below.

To meet the needs of researchers using computational science, we require useful software over and above simply providing a virtualised hosting platform.

There are three principal justifications for private clouds as follows.

- Compliance – private clouds mitigate the security and privacy issues related to regulated work loads;
- Culture – private clouds are better aligned to the command and control cultured expectations and existing skill base of IT departments in large organisations;
- Economics – private clouds re-use existing infrastructure and staff investments and are significantly more cost effective for long running work loads than are public clouds.

What matters to the end user is that IT service departments in large organisations will offer self

service, on demand infrastructure, platforms and applications with research departments buying into a cost effective solution and paying for what they use.

It is possible to combine a variety of internal and external resources to create integrated hybrid clouds that allow work loads to dynamically "spill over" from private to public cloud resources to optimise for price, policy, performance and various service level characteristics. Care must however be taken not to compromise privacy aspects of the private cloud solution.

5.1 Virtualisation and Green IT

Cloud pundits claim green credentials. This is achieved partly through virtualisation. In this way it is claimed that a small number of resources can appear to meet the requirements of a large work load. Additional resources can be switched when there is sufficient demand. Really this has nothing to do with cloud. There is no reason why, in a traditional compute cluster or campus grid, a queuing system such as Sun Grid Engine, Condor or Platform LSF could not preferentially target jobs at nodes which are powered on and power on or off nodes as required. Virtualisation can however be used to improve resource utilisation.

On the other hand, most clouds use para-virtualisation. This carries less overheads than simply running a virtual machine image as certain operations can be executed natively. This however requires hooks from the virtual operating system to bypass the normal code. Hypervisors such as Xen and VMWare can use this method.

Note: there are some worries about data security in a virtualised system, i.e. one running multiple clients' apps on the same servers.

5.2 Shared Service Clouds

In mid-2010 the coalition government wasted no time in setting out severe and comprehensive spending cuts of 20%, with Chancellor George Osbourne introducing a nine point spending review for all government departments. Ambitious schemes such as the implementation of a cloud based infrastructure and services across government are seemingly more relevant than ever. In fact this was referenced in the Digital Britain report of 2009 [22].

The "G-Cloud" strategy has been suggested to save the government £3.2bn of its annual £16bn IT budget, meeting the 20% savings target. The proposal is to replace the current *ad hoc* network of systems hosted in separate departments with a dozen or so dedicated government secure data centres, costing £250M each. By 2015 it is estimated that up to 80% of government departments could be using this system.

In Oct'2011, the government announced a Strategic Implementation Plan (SIP) for its ICT strategy, which minister for the Cabinet Office Francis Maude argued will deliver a more realistic 1.4bn in savings in the next four years.

Part of the SIP looks at plans for the G-Cloud. It claimed the government currently has *an expensive and fragmented ICT infrastructure which often duplicates solutions and impedes the sharing and re-use*

of services and solutions.

SIP indicates that the G-Cloud could save the government £20M between 2012-13; £40M between 2013-14; and £120M between 2014-15. Most savings arise from linking the services to form a Public Service Network as a new integrated and ICT based delivery mechanism.

According to the proposers, G-Cloud not only offers hosted solutions but also standardisation, commoditisation and elasticity. The hope is that cloud computing will provide part of the solution to this problem, and it is argued in the SIP that the G-Cloud will *increase public sector agility and reduce the cost of its ICT*. Faced with large budget cuts, issues that initially deterred some public sector IT managers, such as data risks, are now being re-assessed. Potential benefits of this strategy are to develop more efficient, green practices, improve reliability and obtain much needed cost savings.

A £60M tender notice was issued on 21/10/2011 for four contracts: infrastructure as a service (IaaS); platform as a service (Paas); software as a service (SaaS); specialist cloud services.

The Cabinet Office will also release four more detailed reports by the end of Oct'2011. They will look at end user devices, cloud, ICT capability and greening government ICT strategies.

Of course there are other suggestions for shared services clouds for the research sector.

5.3 Comparison – Sakai Case Study

Sakai was developed, starting at University of Michigan, as an open source collaborative tool for teaching and learning. It is also used for support of e-Research and business activities. A guiding principle of the Sakai developers is that *if your business depends on it, you need to have the capability to modify, develop and maintain it in house* (after Chuck Severance, 2005). Sakai is now the second largest open source portal project in the world managed by the Sakai Foundation with over 200 production installations including Universities of Oxford (`http://weblearn.ox.ac.uk`) and Cambridge (`http://camtools.caret.cam.ac.uk`) in the UK, see `http://sakaiproject.org`. Sakai is a pluggable Java framework and is downloaded and installed using the Tomcat Web server and a database such as mySQL. It is distributed under the Educational Community License.

There is a very large range of available tools provided with Sakai which can be configured to appear as portlets on pages belonging to project work sites. The ones for collaboration include: announcements, blog, chat room, community links, drop box, email archive, forums, mailtool, messages, RSS news reader, polls, resource folders, calendar, search, wiki, site members, glossary, web content (iFrame). There are many more tools developed for educational purposes and there is a well documented procedure for developing and adding other tools. The international community is rapidly developing Sakai-3 which will have many more social networking features, apps, workflow and content management capabilities. An upgrade path will be provided for existing projects and data.

There is internally a strong role based security model with the capability to have moderated or joinable work sites and individual roles of site members granting permissions in them [2].

A number of projects in the UK are developing additional tools to plug into Sakai to enable it to be used as a Virtual Research Environment [1]. These connect to grids and clouds for computational and data storage resources.

5.4 Comparison – the enCore Partner Cloud on the Daresbury Campus

EnCore is a compute on demand service hosted at Daresbury Laboratory and manged by OCF plc using Platform Computing ISF, see below. There is provision for users from both the commercial and academic research sectors, as appropriate to the mixed Daresbury Science and Innovation Campus.

OCF is responsible for pre-sales qualification with business customers to discover required volumes of processing power and benchmarks to demonstrate that enCore can run specific applications or problems faster than their existing infrastructure.

OCF also has a number of turn key applications ready for use with the service. OCF can also work with independent software vendors (ISVs) to get application licensing for the term of a contract with customers, or it can potentially access the end users' licences directly, thus ensuring adherence to the ISV's licensing policy.

Data transfer between the customer and enCore is handled by enCore's simple secure Web interface (Platform HPC Enterprise Portal) or, in the case of extremely large files, by secure shuttle service.

Contracts with OCF are flexible, and use of enCore involves a small annual subscription plus a cost per core hour used.

The service is aimed at UK businesses of any size and from any sector primarily to satisfy the following needs.

- act as an overflow service for businesses to meet a temporary requirement for more processing power;
- enable SME's design consultants for example – to pitch for larger projects than would otherwise be possible, due to the limitations of their IT infrastructure;
- serve as a courtesy service for customers whilst they wait for tenders to complete or for a new HPC system to arrive;
- act as a direct hardware replacement by businesses in order to reduce their capital expenditure.

Academic use of the system is initially for Daresbury Laboratory staff and researchers from University of Huddersfield. It is expected that the service will be extended as more partners come on line. Pricing, usage and access policies are tailored for each user group with SLAs in place. Virtualisation is not used, but resources are managed in a flexible way with Platform Cluster Manager and LSF.

Other companies are now beginning to offer very similar partner cloud service whch target HPC users and remove the overheads of virtualisation [13]. These includ Penguin Computing, the SGI Cyclone service and the Bull Extreme Factory service.

5.5 Comparison – the University of Loughborough Cloud

Loughborough has an on site private cloud is built by Logicalis from Cisco, NetApp and CA technologies to create a self contained, highly virtualised and extremely compact environment. This provides

enough compute, storage and network capacity to meet immediate local demand, while long term future capacity, on demand burst capacity and disaster recovery capability is provided by the Logicalis Research Cloud hosted at Slough.

5.6 Software

There are a number of open source software offerings available to build and manage private clouds. There are also commercial packages such as those from Platform Computing.

Eucalyptus

Eucalyptus, Elastic Utility Computing Architecture Linking Your Programs to Useful Systems, is arguably the best known and certainly the oldest open source cloud solution, `http://open.eucalyptus.com/`.

The components of Eucalyptus are as follows.

Cloud controller – provides a Web interface and Amazon EC2 compatible SOAP interface for virtual machine management. Written in Java.
Walrus – implements Amazon S3 compatible SOAP and REST storage interface. Also written in Java.
Cluster Controller – manages one node group (one ethernet segment). Written in C.
Storage Controller – an Amazon EBS style repository for virtual images. Written in Java.
Node Controller – an abstraction layer over KVM or Xen hypervisors. Written in C.

A Eucalyptus cloud can be managed using tools written in Python which are compatible with Amazon. It is therefore easy to create a hybrid cloud. In many ways Eucalyptus could be considered to be an open source version of EC2.

Evaluation work using Eucalyptus is going on at University of St. Andrews, see StACC Web site `http://www.cs.st-andrews.ac.uk/stacc`. Their private cloud will be used for Ph.D. students to carry out research into cloud computing and its applications.

Eucalyptus also forms the basis of two cloud pilot projects for the NGS, one based in Edinburgh and one in Oxford. See `http://www.ngs.ac.uk/news/research-communities-on-the-ngs-cloud-pilot`.

OpenNebula

OpenNebula is an open source toolkit which uses the Apache 2.0 license. It was developed at the Universidad Complutense de Madrid, `http://www.opennebula.org/`.

It contains a core daemon plus abstraction drivers for network, storager and hypervisors including Xen, KVM and VMWare. There is a user front end which includes management and a node image repository. OpenNebula can integrate with public cloud solutions such as Amazon EC2 and OCCI.

Nimbus

Nimbus is designed as an open source cloud computing IaaS tool kit for science, see `http://www.nimbusproject.org`. It uses the Apache 2.0 license.

Some parts of Nimbus will seem familiar, as it is partly built on the Web services instantiation of Globus. It includes the following.

Three sets of remote interfaces: Amazon EC2 WSDLs; Amazon EC2 Query API; and grid community WSRF.
Storage implementation compatible with S3 REST API.
Virtualisation based on Xen and KVM uses images from the Cumulus repository.
A Web interface is being developed using Python Django.

Nimbus can be configured to use schedulers like PBS or SGE to schedule virtual machines. It launches self configuring virtual clusters from the users command line. The VM images are handled through Cumulus. It defines an extensible architecture that allows you to customise the software to the needs of your project, i.e. a tool kit.

It could be argued that Nimbus is a natural evolution of the grid. From a user perspective it simply launches a remote virtual work space over a secure TLS connection using the WSRF factory mechanism with a defined lease time. This requires Java to be installed.

It is not clear how accounting and user management are handled on Nimbus enabled servers.

OpenStack

OpenStack is supplied by Rackspace Hosting and NASA, see `http://www.openstack.org`. It is a collection of open source technologies available under the Apache-2.0 license delivering a scalable cloud operating system. It is designed to be easy to implement. OpenStack is currently developing three inter-related projects: OpenStack Compute, OpenStack Object Storage and OpenStack Image Service.

OpenStack Compute (Cactus from 15/5/2011, previously Nova) is a fabric controller which supports Xen, KVM, QEMU and user mode Linux hypervisors. Security groups are implemented and it uses the Glance image service described below. It is designed to make it easy to provision and manage large networks of virtual machines, creating a redundant and scalable cloud computing platform. It gives you the software, control panels, and APIs required to orchestrate a cloud, including running instances, managing networks, and controlling access through users and projects.

OpenStack Object Storage (Swift) is used to create redundant, scalable object storage using clusters of servers. It is not a file system or real time data storage system, but rather a long term storage system for relatively static data. Examples include virtual machine images, photo storage, e-mail storage and backup archives.

The system is distributed and scalable. Objects are written to multiple hardware devices with the OpenStack software responsible for ensuring replication and integrity across the storage cluster.

OpenStack Image Service (Glance) provides discovery, registration and delivery services for virtual disk images. The Image Service API server provides a standard REST interface for querying information about virtual disk images stored in a variety of back end stores, including OpenStack Object Storage.

The service is a multi-format image registry supporting a variety of formats, including: raw; machine (kernel/ramdisk outside of image, a.k.a. AMI); VHD (Hyper-V); VDI (VirtualBox); qcow2 (Qemu/KVM); VMDK (VMWare); and OVF (VMWare, others).

Penguin Computing

Penguin offers a public cloud known as POD, Penguin on Demand, but the software can also be installed locally. Unlike other clouds there is no virtualisation, so processes are targeted direct to physical cores which can yield improved performance. Scyld ClusterWare from Penguin Computing was one of the cluster management products reviewed by Cable and Diakun [8]. ClusterWare enables a Linux based private cloud to be installed, provisioned and managed. The system addresses a large homogeneous collection of nodes, workers being effectively diskless clones of the head or master node with a minumum of services. All work is carried out on the master node, including running jobs, which are subsequently spawned to the workers. A special set of commands and user interface is provided for this which makes it easy to manage high throughput tasks with a simple workflow script. For MPI tasks there is beoMPI, which is based on MPICH. It is designed to enable the running of turn key applications using relatively simple scripts with just a few parameters additional to the normal mpirun command. PVFS, Parallel Virtual File System, is also available.

Whilst this IaaS may be fine for new or relatively portable applications it does not provide such a flexible environment as many researchers, particularly developers, have come to expect with a number of supported compilers, numerical libraries and combinations of MPI and OpenMPI versions.

There is a copy of the Scyld HPC Programmer's Guide on-line here `http://cougar.triumf.ca/scyld-doc/programmers-guide/`. This includes a section on porting applications.

Platform Computing

Platform claim to be the leading independent cloud management software provider building on their experience of cluster and grid management, and have recently announced the proprietary Infrastructure Share Facility, ISF v2.1, a new release of their modular software for building and managing enterprise private clouds. This is designed to support the entire application life cylcle from development and testing to turn key applications. See `http://www.platform.com/privatecloud`. A white paper is available from the Web site [25].

ISF is expected to support a variety of work loads including: test and development; HPC; J2EE; others. It is built as a three layer architecture as follows.

Service delivery layer: contains application services; self service portal and APIs; reporting and accounting. This top layer provides interfaces to users and applications as well as supporting the life cycle of cloud service management. A self service portal enables users to request and obtain physical servers and VMs. Platform ISF has a set of APIs that can be called by applications, middleware and work load managers to request and return resources without human intervention. Templates can be configured for simple and complex N-tier business applications to automate their life cycle management. ISF allows for the starting of all the components of an N-tier application, the adding or removal of a resource, and monitoring and failure recovery. ISF supports middleware such as J2EE, SOA, CEP and BPM, and workload schedulers such as AutoSys and Platform LSF. No change to the application using this software is needed. The service offerings can be structured as: complete application environments (e.g. application packages, CPU, memory, storage and networking); as bare metal servers with an operating system installed; or as virtual machines. SLAs can be associated with each service offering. ISF collects all resource usage data and provides reports and billing information. Alternatively, the

cloud administrator may choose to feed the usage data into site specific reporting and charge back tools.

Allocation Engine: includes reservation and on demand scheduling; resource aware allocation policies; and self service resource planning. Once a pool of shared resources is formed, a set of site specific sharing policies is configured in the allocation engine layer to ensure that applications receive the required resources. The policies ensure that the organisation's resource sharing priorities are applied, and that the quota constraints applicable to business groups sharing the cloud are reinforced. The allocation engine matches IT resource supplies to their demands based on resource aware and application aware policy management.

Resource Integration: contains VM manager adapters; provisioning tool adapters; and external service adapters. This foundation layer integrates distributed and heterogeneous IT resources to form a shared system. Resource integration is the opposite of server virtualisation - instead of creating multiple VMs on one physical server, this capability creates one shared computer out of many heterogeneous servers, storage devices and interconnects. All major industry standard hardware, operating systems (both Linux and Windows) and VM hypervisors (including VMware ESX, Xen, Citrix XenServer, Microsoft Hyper-V and Red Hat KVM) are supported. The resource integration layer also uses provisioning tools to set up application environments on demand. It integrates with many 3rd party tools for various systems management tasks out of the box, including directory services, security, and monitoring and alert. Its extensible framework of resource and management adapters enables ISF to fit into an existing data centre systems environment. This layer can also transparently integrate resources from external providers whilst maintaining its private cloud management environment.

The range of solutions which can currently be integrated with ISF is as follows.

VM: VMWare ESX; Microsoft Hyper-V; Citrix Xen; Red Hat KVM; Sun Solaris Containers; Xen open source.
Provisioning: BMC BladeLogic; HP Opsware; Tivoli Provisioning Manager; Symantec Altiris; IBM xCAT; Platform Cluster Manager; Scalent IM.
External services: Amazon EC2; IBM CoD; HP Enterprise Services.

Contrail

Contrail is a new EU funded 3 year project aiming to provide an open source solution for managing infrastrucuture and linking to other clouds, see `http://contrail-project.eu`. This extends work on the open source XtreemOS system developed in a previous project with STFC as a partner.

It is currently not clear how this will work alongside the Venus-C project.

6 Some Conclusions

Some of the gaps and challenges to cloud computing have been identified as follows.

Capability computing. Due to the process parallel nature of the service and calculations performed, very tightly coupled capability computing may be poorly served by cloud computing. For this reason,

some of the research clouds illustrated are really a simplified interface to clusters on a grid.

Charging systems. According to a recent study, the charging systems operated by service providers may be cost effective for small and medium sized users, but not for heavy users who own their own compute infrastructures already. More flexibility is required, particularly for commercial users, where a quota basede system should be replaced by pay-as-you-go, e.g. using a PayPal service.

Access and usability. One of the primary algorithms used in cloud computing is MapReduce, developed by Google and used in the open source Hadoop file system. It is unclear how easy utilisation of software based on such algorithms will be for users with little coding experience. Conventional software may require significant modification to use in conjunction with clouds which use MapReduce.

Security and risk. The most highly discussed concerns are with data security, as noted here. Concerns arise because data are secured in the servers of the service providers, and the user has much less direct control over security. Cloud computing providers can potentially offer a range of possible security levels to users, but discussions are ongoing over the holding of sensitive and personal data by third parties. Other risks, such as company failures whilst holding research data, should also be compared with alternative provisions in a full risk assessment. Despite these issues, all kinds of security measures are cheaper when implemented on a larger scale. Therefore the same amount of investment in security buys better protection [23].

Bandwidth issues. As the hardware is remote from the user, there is the potential for users to access their work using much more lightweight, portable Web enabled interfaces. However, for users with large data sets, bandwidth can be a considerable problem. There are a number of potential solutions to bandwidth problems, from using data distribution tools to sending physical drives with data on them in a van to the service provider (e.g. Penguin Computing offer to plug in a user's 2TB SATA disk if delivered this way). Linking physically to a hub via a dedicated link bought or leased from a tele-communications company is a further option, but apart from being expensive, reduces the agility of being able to change service providers if a rival service becomes less expensive. There may also be a need for software applications that deal with data submitted to clouds, particularly where the data are particularly large or complex.

Virtual machines. Most cloud computing uses virtual machines which encapsulate the users' software and data. Prepared collections of software to run on these machines are called "images". There may be benefits to organising arrangements for finding, maintaining and creating images to meet particular research requirements. This is referred to as "provisioning" and is done by the HPC community already, see [8].

The EU FP7 programme commissioned an expert group report [18] which was published in Jan'2010. This indicates the need for further work. The ENISA report [23] also lists security related research areas.

Another FP7 project, the Partnership for Advanced Computing in Europe (PRACE), disagrees that cloud is the future of high performance computing and is instead focusing efforts on the implementation of super-computers with a combined computing power in the multi-petaflop/s range. EPSRC is the formal UK partner in PRACE and EPCC and Daresbury Laboratory are leading work packages.

Two independent studies commissioned by JISC have drawn conclusions from investigating the use of clouds for research [21, 11]. These investigated using the cloud for actually doing research, rather

than broader research support activities.

Wills et al. [21] concluded that in order to understand how cloud computing can be used for research, there are a few facts which need to be appreciated as follows.

- Cloud computing for research is in its infancy. Cloud computing for research is still in a very early stage of development. Its concepts, characteristics, underlying technologies, and applicability for research are not clear to many. Although early cloud adopters have shared some useful information, including example case scenarios demonstrated in this report, in the "Using Cloud for Research" project, and in those published by commercial public CSPs (Cloud Service Providers), most of this information is constrained by the interests involved and only show that cloud computing can run specific research applications. We need more information, e.g. performance benchmarks, economic savings, etc. before proceeding to define a global strategy for using cloud computing in research.
- Commercial cloud or "research cloud", there is no right answer yet. There is still no right answer to the way in which we should use cloud. Whether public or commercial, interoperability is a key issue for cloud computing in research.
- Current offerings are not research friendly. The current cloud computing offerings come from two main sources, public CSPs and open source cloud-ware. The public CSPs provide well defined cloud service APIs or platforms which allow users to interact and develop cloud applications. These APIs and platforms were designed to support business application development, however, and offer limited support for research applications, such as workflow and parallel computing applications. Although public CSPs provide powerful self management facilities and well defined programming APIs, these enhanced features also "lock" applications in to a specific cloud infrastructure. On the other side, researchers are starting to think of building specific cloud environments by using open source cloud-ware. These open source offerings only provide low level programming APIs and service interfaces, however, leaving resource management tasks to application developers.

In order to design a global cloud computing strategy for UK research in a sensible way, there are some things we need to learn.

- Researchers need to learn for themselves the reasons for cloud computing, what benefit it may provide over other computing models being employed in organisations, and what key technologies enable the migration to cloud computing.
- There are many commercial CSPs that are successful in different ways. Amazon is the highest profile IaaS CSP at present, while IBM exhibited a successful service delivery model for its on-premises software products. We need to learn from these successful experiences and use them for research cloud service delivery.
- In using open source cloud-ware to develop customised private environments and considering only limited functionalities and low level self service APIs provided by these open source offerings, researchers can learn from advanced technologies being used by other computing models, e.g. grid computing, to enhance self management capabilities of private cloud infrastructure.

Hammond et al. [11] made the following recommendations.

- any organisation considering adopting any cloud services for mission critical applications, or for processing personal or otherwise sensitive information, should obtain specialist legal advice regarding their contracts and SLAs.
- JISC should investigate the issues surrounding the management and preservation of research data in cloud systems and produce guidance aimed at researchers. This should support risk assessment and management and should not design or develop technical solutions.
- JISC should investigate mechanisms for national engagement, negotiation and procurement of cloud services, primarily with AWS (Amazon Web Services), Google and MS (Microsoft), but allowing for the engagement of smaller, niche providers.
- The NGS, and its funders, should consider whether there is a role for that organisation in supporting the development of virtual machine images for common research codes, to allow users to deploy them easily within commercial and private clouds. This may include liaising with or funding the developers or maintainers of the codes.
- unless backed by clear evidence of demand and a robust and revenue neutral business case, JISC should not support the development of a production UK academic research cloud.

To meet the needs of a growing community of computational scientists, private or partner clouds offering Software as a Service can appeal to those with limited experience and access to resources. This is equally true for acadmic or commercial end users. For those considering offering such a service with turn key applications pre-installed, the following division of responsibilities with respect to security was identified by ENISA [23].

Customer	Provider
Compliance with data protection law in respect of customer data collected and processed	Physical support infrastructure, facilities, rack space, power, cooling, cabling, etc.
Maintenance of identity management system	Physical infrastructure security and availability, servers, storage, network bandwidth, etc.
Management of identity management system	OS patch management and hardening procedures, check also any conflict between customer hardening procedure and provider security policy
Management of authentication platform, including enforcing password policy	Security platform configuration, firewall rules, IDS/IPS tuning, etc. Systems monitoring Security platform maintenance, firewall, host IDS/IPS, antivirus, packet filtering Log collection and security monitoring

The SaaS model dictates that the provider manages the entire suite of applications delivered to end users. Therefore SaaS providers are mainly responsible for securing these applications. This is no different on a cloud to what it was on a shared national or other service. Customers are

normally responsible for operational security processes, complying with user and access management requirements. However the following questions might be asked by prospective customers.

- What administration controls are provided and can these be used to assign read and write privileges to other users?
- Is the SaaS access control fine grained and can it be customised to their organisations policy?

The above responsibilities have to be seriously considered and discussed with each customer to ensure that they are satisfied before offering such a service.

With PaaS or IaaS more responsibility is transferred to the customer for security around installed software and applications.

7 Acknowledgements

This work was partly funded by EPSRC through the SLA with Daresbury Laboratory. Funding from the e-Science Institute. Funding from JISC as part of the CRIB VRE project *Collaborative Research in Business*.

The author would like to thank the following people for input:

Mark Baker and others participating in the Research3 Theme *The Influence and Impact of Web 2.0 on e-Research Infrastructure, Applications and Users*, see `http://www.research3.org`.

Jerry Dixon (OCF plc) and Terry Fisher (Platform Computing) for discussions on the technical implementation and business models of the enCore service.

Participants of the UK Campus Grids Special Interest Group.

References

[1] R.J. Allan *Virtual Research Environments: from Portals to Science Gateways* (Chandos Publishing, Oxford, 2009) 230pp `http://www.woodheadpublishing.com/en/book.aspx?bookID=1892&ChandosTitle=1`

[2] R.J. Allan and X. Yang *Using Role based Access Control in the Sakai Collaborative Framework* (STFC, Mar'2008)

[3] A. Apon, S. Ahalt, V. Dantuluri, C. Gurdgiev, M. Limayem, L. Ngo and M. Stealey *High Performance Computing Instrumentatino and Research Productivity in US Universities* Journal of Information Technology Impact 10:2 (2010) 87-98

[4] C. Baru, R. Moore, A. Rajasekar and M. Wan *The SDSC Storage Resource Broker* Proc. CASCON (Toronto, 30/11-3/12/1998) `http://www.npaci.edu/DICE/Pubs/srb.pdf`

[5] D. Bradshaw *Western European Software as a Service Forecast 2009-13* LT02R9 (Apr'2009)

[6] G. Brunette and R. Mogull (eds.) *Security Guidance for Critical Areas of Focus in Cloud Computing* v2.1 (Cloud Security Alliance, 2009) 76pp `http://www.cloudsecurityalliance.org/csaguide.pdf`

[7] D. Cable *Using MOAB Cluster Suite for Cluster Management* (STFC, 2009) `http://www.cse.scitech.ac.uk/disco/publications/moab_final_report.pdf`

[8] D. Cable and M. Diakun *A Review of Commodity Cluster Management Tools* DisCo report (Nov'2010) 26pp `http://www.cse.scitech.ac.uk/disco/publications/publications.shtml`

[9] N. Carr *The Big Switch – our new Digital Destiny* (W.W. Norton and Co., 2008) 276pp ISBN 978-0393062281

[10] F. Gens, R.P. Mahowald and R.L. Villars *Cloud Computing 2010* IDC update report TB20090929 (Sep'2009)

[11] M. Hammond, R. Hawtin, L. Gillam and C. Oppenheim *Cloud Computing for Research* Report to JISC (Curtis and Cartwright Consulting Ltd., 2010) `http://www.jisc.ac.uk/whatwedo/programmes/researchinfrastructure/usingcloudcomp.aspx`

[12] J.-P. Laisné, P. Aigrain, D. Bollier, M. Tiemann, *et al.*, *2010 FLOSS Roadmap* (Open World Forum, 2010) PDF document available from Web site `http://www.2020flossroadmap.org/`

[13] G. Law (ed.) *Cloud in HPC* Scientific Computing World (Dec'2011-Jan'2012) 20-22 `http://www.scientific-computing.com`

[14] T. Mather *Cloud Security and Privacy: An Enterprise Perspective on Risks and Compliance* (O'Reilly, 2009) 336pp ISBN 978-0596802769

[15] P. Mell and T. Grance *NIST Working Definition of Cloud Computing* (NIST, 2009) `http://www.csrc.nist.gov/groups/SNS/cloud-computing/index.html`

[16] M. Miller *Cloud Computing: Web-Based Applications That Change the Way You Work and Collaborate Online* (QUE, 2008) 312pp ISBN 978-0789738035

[17] G. Reese *Cloud Application Architectures: Building Applications and Infrastructure in the Cloud: Transactional Systems for EC2 and Beyond* (O'Reilly, 2009) 208pp ISBN 978-0596156367

[18] L. Schubert (rapporteur) *The Future of Cloud Computing: Opportunities for European Cloud Computing Beyond 2010* EU FP7 Programme expert group report (Jan'2010) 71pp `http://cordis.europa.eu/fp7/ict/ssai/docs/cloud-report-final.pdf`

[19] S.N. Shah *Cloud Computing by Government Agencies* (IBM DeveloperWorks, Aug'2010)

[20] J.M.H. Thomas, R.P. Tyer, R.J. Allan, J.M. Rintelman, P. Sherwood, M.T. Dove, K.F. Austen, A.M. Walker, R.P. Bruin and L. Pettit. *Science carried out as part of the NW-GRID Project using the e-Minerals Infrastructure* `http://epubs.cclrc.ac.uk/bitstream/1678/rmcs_and_science.pdf`

[21] G. Wills, L. Gilbert, X. Chen and D. Bacigalupo *Technical Review of using Cloud for Research (TeciRes)* Report to JISC (University of Southampton, June 2010) `http://www.jisc.ac.uk/whatwedo/programmes/researchinfrastructure/cloudcomptechreview.aspx`

[22] *Digital Britain Final Report* UK Departments for Culture Media and Sport and Business Innovation and Skills (June 2009) ISBN 978-0-101765022. `http://www.culture.gov.uk/what_we_do/broadcasting/6216.aspx`

[23] D. Catteddu and G. Hogben (eds.) *Cloud Computing – benefits, risks and recommendations for Information Security* European Network and Information Security Agency (ENISA) (Nov'2009) `http://www.enisa.europa.eu/act/rm/files/deliverables/cloud-computing-risk-assessment`

[24] The Cloud Computing Portal is a public database of cloud computing providers, news and resources. `http://cloudcomputing.qrimp.com/portal.aspx`

[25] *Enterprise Cloud Computing – transforming IT* Platform Computing White Paper (July 2009)