TECIRES REPORT

USING CLOUD FOR RESEARCH: A TECHNICAL REVIEW

Xiaoyu Chen, Gary B. Wills, Lester Gilbert, David Bacigalupo
{xc2, gbw, lg3, db1f}@ecs.soton.ac.uk
School of Electronic and Computer Science
University of Southampton

June, 2010

EXECUTIVE SUMMARY

The purpose of the TeciRes project was to conduct a technical review of the current landscape within cloud computing to establish the extent to which existing solutions meet encountered and envisioned requirements for using emerging cloud technologies, in particular those which enable computing and storage cloud facilities for research in Higher Education (HE) institutions, and to make recommendations on further development, guidance, and standardisation.

The TeciRes project is one of three parallel projects funded by JISC, the other two complementary projects being a review of "using cloud computing" focusing on cloud use cases from users' perspectives, and a review of "environmental and organisational" implications of cloud computing" in Higher Education (HE) and further education sectors.

This project combined expert consultancy and desk research approaches. The project started with stakeholder analysis, case studies from HE institutions and research departments, and a close collaboration with the "using cloud for research" group, while paying particular attention to technical requirements and issues. The project team attended research conferences and organised workshops to enrich its understanding of technical requirements, challenges, available solutions, and ongoing technical research activities. Continuous desk research was ongoing throughout the project for information consolidation, reviewing available and emerging standards, and technical prototyping.

This technical review and two guidance notes (cloud computing for research, technical issues on cloud computing for research) are available on the project website[1].

During the course of the TeciRes project, the number of Cloud-related workshops and conferences, along with white papers and publications, has greatly risen. In part this shows the interest, both commercial and academically, of this field. The outputs from the project are a reflection of an extensive literature research focused on the issues that practitioners in the field are facing. Where possible the report tries to show how they have overcome these problems. There are still many questions unanswered, in part because cloud computing for research is still only a concept to many or has only been tried on a small scale.

Recommendations

Recommendation 1: JISC SHOULD continue support for proofs of concept of using cloud for research. Cloud computing is in its infancy; hence a number of proof of concept application are need in order to give a baseline understanding/measure from which to assess the usefulness of cloud computing for research.

Recommendation 2: JISC SHOULD stimulate cloud-based shared service delivery. Various JISC committees have invested considerable time and resources in developing shared services in a non-cloud context. Hence there are a number of existing shared services in the Grid and VRE space that can be migrated to the cloud for research, and there are new services to be developed, including the need to provide management software for virtual machine images. This will allow Universities to explore the appropriateness of various cloud types and also their 'green' credentials. More detailed recommendations related to environmental issues can be found in the final report[2] of review of the environmental and organisational implications of cloud computing in Higher and Further Education.

Recommendation 3: JISC COULD work with international standard bodies, for instance Open Cloud Computing Interface, a working group in OGF, or similar non-grid

[1] http://tecires.ecs.soton.ac.uk
[2] http://www.jisc.ac.uk/whatwedo/programmes/greeningict/environmentalreviewcloudcomp.aspx

organisations. This will require funding and efforts, however, on standards adoption through reference implementations.

Recommendation 4: JISC COULD fund various cloud service facilities, particularly storage and data cloud facilities, to explore data security and privacy as highlighted areas of concern from practitioners and stakeholders. Funding projects in this area will allow researchers to provide alternative methods of sharing research data inside UK research communities, with enhanced security and data privacy.

Recommendation 5: JISC MAY fund large scale institutional and cross institutional clouds for research. This could be informed by the work of the NGS cloud pilot project currently underway, with a view to providing cloud services on the NGS structure.

Acknowledgements

The authors thank in particular the following individuals and organisations for their time, support and input during the course of this review. We apologise to others who have also contributed, but whose names have been omitted.

Arjuna	Steve Caughey, Stuart Wheater
Curtis and Cartwright	Max Hammond, Rob Hawtin
Cardiff University	Omer Rana, James Osborne
EPCC	Rob Baxter, Steve Booth, Sean Mcgeever, Kostas Kavoussanakis
IBM	
Intellect UK	Ian Osborne
GridPP	Tony Doyle, Akram Khan
JANET	Daniel Perry
Microsoft	Matthew Deacon
NeSC	Malcolm Atkinson, Jano van Hemert, Steve Thorne
Open Grid Forum	Alexis Richardson, Thijs Metsch
OMII UK	Neil hue Hong
StACC	Alexander Voss, Ian Sommerville, James Smith, Ali Khajeh-Hosseini
University of Newcastle	Aad van Moorsel, Hugo Hidden, Simon Woodman, Paul Watson, Steve McGough
University of Oxford	Matteo Turilli, David Wollam
University of Reading	Mark Baker
University of Southampton	Steve Johnston, David De Roure, Oz Parchment
University of Warwick	Stephen Jarvis, Gihan Mudalige,Simon Hammond, Adam Chester, Roger Packwood, Matthew Ismail

Contents

1 INTRODUCTION

The emerging cloud computing is driven by and contributes to the realisation of four major computing concepts (Figure 1-1): utility computing, on-demand computing, autonomic computing, and Green IT.

Utility computing was firstly proposed in the 1960s by John McCarthy, who envisioned that future organisations would simply plug in to a computing grid for computational resources rather than providing their own computing power, just like connecting to an electrical grid, and pay fees based on what is used.

On-demand computing, sometimes also referred as utility computing, focuses on dynamic provision or elastic scalability of computing capabilities to efficiently meet fluctuating demands. On-demand computing resources can either reside within an organisation or be outsourced to third-party service providers.

In 2001, IBM initiated research on autonomic computing that aimed at developing an intelligent computing system that was capable of self management, and at reducing the complexity of system management particularly for large-scale computing environments. An autonomic system is able to monitor, make decisions, and adjust the underlying system environment on behalf of system administrators in order to fulfil pre-defined Quality of Service (QoS).

With increasing recognition of the importance of reducing greenhouse gas emissions, the concept of Green IT or Green Computing was proposed recently to tackle environmental issues through environmentally and economically sustainable IT infrastructure.

Cloud computing therefore can be understood as a logical evolution of these envisioned computing concepts, and aims at providing an elastic, self-managed, cost- and energy-effective computing environment.

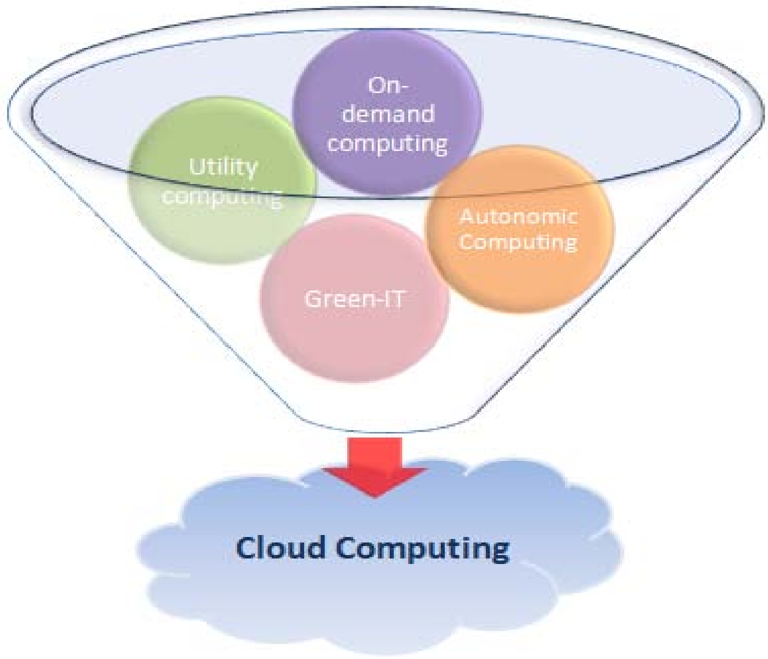

FIGURE 1-1: CLOUD COMPUTING AND "WATER-DROP" DRIVERS

Common misunderstandings came from thinking of cloud computing as a new technology. Cloud computing is better understood as a new business model that delivers computing capabilities at different levels, ranging from computing infrastructure to high-end applications. With cloud computing, an enterprise can outsource IT infrastructure and management tasks to third-party cloud providers rather than owning a data centre, can adapt to changing demands much more quickly, and only pays for what is used.

Recently, there has been a rapidly increasing adoption of cloud computing for research activities in the UK Higher Education (HE) sector. These early activities include outsourcing IT infrastructure to public Cloud Service Providers (CSPs), "bursting" out and offloading to public cloud services, migrating existing research applications to be cloud-ready, and building cloud infrastructures using open-source cloud software (Cloudware).

Despite all these recent uses of cloud computing in research communities, the term itself and its technological underpinnings are still confusing to many. For instance, there is a lack of clarity about the differences from or similarities to pre-existing computing models such as grid computing and high-performance computing which have mostly been employed in research communities.

This report is one of three reports commissioned by JISC regarding Cloud Computing. While this report focuses on the technical aspects of using cloud for research, the other two reports have a different focus:

- Review of Using Cloud Computing for Research focuses on "using cloud computing" and provides detailed use cases from the users' perspectives. It also provides advice and guidance on the governance, legal and economic issues around using cloud computing for research.
- Review of Environmental and Organisational implication of Could Computing in HE and Further Education focuses on the environmental costs and benefits of cloud computing. In addition it reviews the necessary changes for the adoption of cloud computing.

This report is the principal output of the TeciRes project. It provides a comprehensive review of the technical requirements of using cloud for research and of current cloud offerings. In particular, this report gives answers to the following key questions:

- What is cloud computing and how is it different from other computing models employed in research communities?
- What are current practices and technical requirements of using cloud for research in the UK HE sector?
- How would current offerings of public cloud services, open-source Cloudware, standards, and active research activities help to accommodate identified technical requirements?
- What would be the practical next steps and long-term goals to accelerate the adoption of cloud for research communities?
- How could JISC support cloud computing for research?

Details of use cases and environmental and organisational implications of using cloud computing for research can be found in the final reports of the two other review projects.

2 THE EMERGING CLOUD COMPUTING

Cloud computing, as an emerging computing paradigm, has received considerable interest. We argue that cloud computing is nothing new, in that it is built on long-established trends and underpinned by pre-existing and evolving computing technologies. We also argue that cloud computing is indeed new because it changes the way IT infrastructure is procured and IT applications are built. In this section, we address cloud computing's core characteristics and differences from other computing models.

2.1 DEFINITION AND CHARACTERISTICS

As shown in Table 2-1, there are various definitions and interpretations of "Cloud" or "Cloud computing". Major IT companies such as IBM, Sun Microsystems, and Microsoft have recognised cloud computing as the next-generation business model and defined custom cloud strategies. Definitions of cloud computing are therefore given in various ways based on their own understanding and business strategies. Recently a group of domain-specific experts have proposed definitions that aim at abstracting common characteristics of the "Cloud" and "Cloud computing". The four representative definitions listed in Table 2-1 came from Gartner, US National Institute of Standards and Technology (NIST), EU commission Expert Group, and Berkeley Institute.

In this report, we define cloud computing:

> Cloud computing is an emerging business model that delivers computing services over the Internet in an elastic self-serviced, self-managed, cost-effective manner with guaranteed Quality of Service(QoS).

Based on our definition, a typical cloud environment consists of a set of hardware, software, and manageability interfaces, which collectively deliver computing as a service with the following four essential characteristics:

- **Elastic** and dynamic scaling
- **Self-service** provisioning and management
- **Self-management** and automatic scaling
- **Cost effectiveness** and multi-tenancy on per-usage basis

It is worth noting that these essential characteristics have different meanings to different types of cloud stakeholders. In order to ensure common understanding, we categories potential cloud stakeholders into four general roles: End Users, Cloud Service Providers (CSPs), Cloud Tool Providers (CTPs), and Cloud Application Vendors (CAVs). The definition of these roles and the role-specific implications of individual characteristics are listed in Table 2-2.

TABLE 2-1: LIST OF DEFINITIONS AND INTERPRETATIONS OF CLOUD COMPUTING

Organisation	Definitions
amazon webservices	"Cloud Architectures are designs of software applications that use Internet-accessible on-demand services. Applications built on Cloud Architectures are such that the underlying computing infrastructure is used only when it is needed (for example to process a user request), draw the necessary resources on-demand (like compute servers or storage), perform a specific job, then relinquish the unneeded resources and often dispose themselves after the job is done. While in operation the application scales up or down elastically based on resource needs." [1]
IBM	"The concept of cloud computing has developed from earlier ideas such as grid and utility computing, and aims to provide a completely Internet-driven, dynamic and scalable service-oriented IT environment, which can be accessed from anywhere using any Web-capable device." [2]
Sun microsystems	"We define cloud computing as services that are encapsulated, have an API, and are available over the network. This definition encompasses using both compute and storage resources as services." [3]
Microsoft	"We believe cloud computing represents the platform for the next generation of business. Cloud Computing is driving the transformation of the IT industry across the entire stack: hardware model delivering incredibly powerful and efficient hardware at a fraction of the cost; application model allowing developers to rapidly create highly available secure cloud applications; operations model keeping cloud applications available 24X7 with 9-to-5 management." [4]
ubuntu	"Cloud computing is a consequence of economic, commercial, cultural and technological conditions that have combined to cause a disruptive shift in information technology (IT) towards a service-based economy. The underlying driver of this change is the commoditisation of IT." [5]
vmware	"At VMware we find that cloud computing is best understood from the perspective of the consumer of services provided by such a computing approach. The following four attributes are core to cloud computing based services: radically improved economics through shared infrastructure; pricing based on consumption: You only pay for what you use; flexible access; a lightweight entry and exit service acquisition model." [6]
Gartner	"The term 'Cloud Computing' refers to any computing capability that is delivered as a service over the Internet. While there is no authoritatively accredited definition of the concept, one of the most frequently used definitions is the one given by Gartner, who describe cloud computing as "a style of computing where massively scalable IT-related capabilities are provided 'as a service' across the Internet to multiple external customers." [7]
NIST National Institute of Standards and Technology U.S. Department of Commerce	"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." [8]
	A 'cloud' is an elastic execution environment of resources involving multiple stakeholders and providing a metered service at multiple granularities for a specified level of quality [of service]. [9]
Berkeley UNIVERSITY OF CALIFORNIA	Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the data centres that provide those services. The services themselves have long been referred to as Software as a Service (SaaS), so we use that term. The data centre hardware and software is what we will call a Cloud. [10]

TABLE 2-2: CLOUD CHARACTERISTICS AND IMPLICATIONS TO PARTICIPANT ROLE TYPES

Role		Characteristics			
Name	**Description**	**Elasticity**	**Self Service**	**Self Management**	**Cost Effectiveness**
End User	Consumer of cloud application services	•reliability •availability	•service instance management	•guaranteed QoS	•pay for what is used
CSP	Provider of cloud capabilities	•virtualisation •resource pooling	•manageability interfaces •programming APIs and platform	•Service Level Agreement (SLA) enforcement	•resource re-use and reallocation •multi-tenancy
CAV	Vendor of cloud application services	•on-demand scaling	•self-provisioning •full software lifecycle support	•efficiency and performance	•zero start-up investment
CTP	Third-party provider of cloud support and manageability tools	•auto-scaling	•accounting •monitoring •historic usage reporting	•automated scaling and management	•zero start-up investment
(NOTE: A cloud stakeholder may take more than one role at the same time**)**					

2.2 Service delivery models

Current commercial CSPs deliver cloud computing capabilities at three hierarchical service layers (Figure 2-1): the infrastructure layer, the platform layer, and the application layer. Each of the cloud delivery layers target different groups of stakeholder. Each layer provides self-service and self-management facilities.

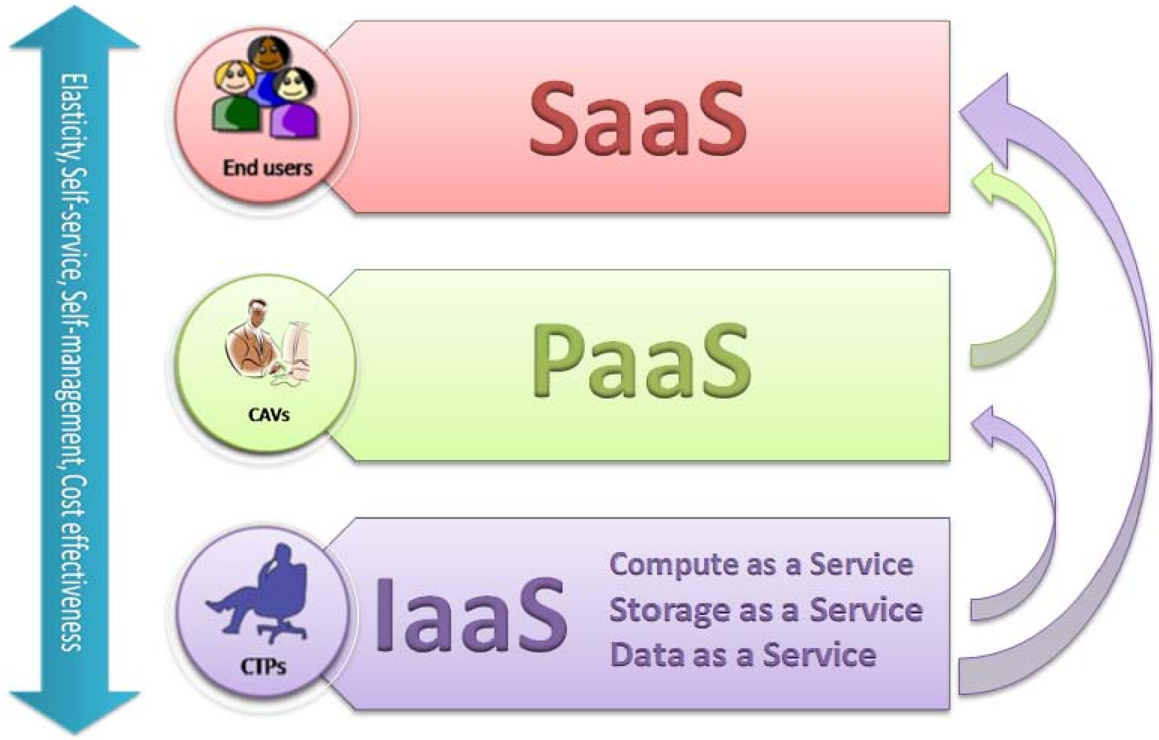

Figure 2-1: Relations and characteristics of cloud service delivery models

2.2.1 Infrastructure as a Service

Cloud services that deliver infrastructure resources (i.e. compute, storage, networking, and operating systems) as a service are known as Infrastructure as a Service (IaaS). Instead of going through procurement, capacity planning, installation, and configuration processes, the IaaS model allows customers to start a new project quickly by renting computing resources. The key characteristic of an IaaS cloud is elasticity and scalability, enabling computing resources to scale up and down. Most IaaS cloud providers offer scalability under customers' control with direct self-service interfaces, through which consumers can request to scale, control, and manage computing resources. IaaS self-service interfaces can also be used to provide automatic scaling features by customers, third-party CTPs, or CSPs themselves. An IaaS cloud is also referred as a resource cloud.

Examples: Amazon Web Services (AWS), GoGrid, VMware vCloud, Canonical

According to the different types of resource offered, IaaS cloud can be further divided into three sub-categories: Computing as a Service (CaaS), Storage as a Service, and Database as a Service (DaaS).

2.2.2 Computing as a Service

Computing as a Service (CaaS) offers customers access to raw computing power on virtual servers or virtual machine instances. CaaS provides self-service interfaces for on-demand provisioning and management (i.e. start, stop, reboot, destroy) of virtual machine instances. A CaaS provider may also provide self-management interfaces for auto-scaling and other automatable management facilities.

Examples: Amazon Elastic Compute Cloud (EC2), ElasticHost

2.2.3 Storage as a Service

Storage as a Service offers online storage services allowing on-demand storing and access to data on third-party storage spaces.

Examples: Amazon Simple Storage Service (S3), Amazon Simple Queue Service (SQS), Amazon Simple Notification Service (SNS)

2.2.4 Database as a Service

Database-as-a-service (DaaS) includes standardized processes for accessing and manipulating (i.e. write, update, delete) data through database management systems (DBMS) that are hosted in the cloud.

Examples: Amazon SimpleDB, Azure SQL

2.2.5 Platform as a Service

A Platform as a Service (PaaS) cloud lies directly upon an IaaS layer with a solution stack encapsulating everything required for the whole software engineering lifecycle (i.e. design, development, debugging, testing, and deployment). The potential consumers of a PaaS cloud service are therefore software developers and testers. Most PaaS vendors lock developers into particular development platforms and debugging tools, however, and do not allow direct communication with lower computing infrastructures, although certain programming APIs might be provided with limited functionalities of infrastructure control and management.

Examples: Windows Azure, Google Application Engine, IBM Websphere Cloudburst, Force.com

2.2.6 Software as a Service

Software as a Service (SaaS) is a cloud delivery model that has existed for a long time. A SaaS is an implementation of a business application or process that is developed on a cloud platform and hosted in a cloud infrastructure. SaaS providers deliver domain-specific applications or services over the Internet and charge end users on a pay-per-usage basis.

Examples: Saleforce CRM, Google Mail

2.3 Deployment models

A cloud system (IaaS, PaaS, and SaaS) can be deployed using the following three main models.

2.3.1 Private cloud

Private cloud services are owned and delivered only within an enterprise or organisation. Private clouds employ virtualisation and automated management technologies to enhance the scalability and utilisation of a local data centre while reducing administrative and management tasks. A Private cloud can be built on existing on-premises computing infrastructure using open-source cloud software (Cloudware) or third-party commercial offerings to meet the needs of an organisation.

2.3.2 Public cloud

A public cloud provides computing services that are publicly accessible through standard self-service APIs over the Internet. Public cloud services may be free for development and test purposes, or may charge on a per-usage basis in a production cloud environment. A private cloud can be easily turned into a public cloud by opening self-service APIs for public access. Because most public cloud services are paid for, SLA enforcement is of great importance. A public cloud may offer a certain degree of control over virtualised resources (e.g. virtual images, virtual machine instances) but reserve full control over local computing infrastructure.

2.3.3 Hybrid cloud

A Hybrid cloud is a mixed deployment model, employing both private and public infrastructures. A hybrid cloud is mostly used for offloading processes and/or data to a public cloud while maintaining the desired degree of control inside a private cloud. The hybrid cloud model is mostly adopted for maintaining sensitive data inside a local private cloud. In this report, we distinguish the cloud bursting model (see section 3.2) from the hybrid cloud model because cloud bursting does not necessary involve a local private cloud infrastructure. The hybrid cloud model may involve many cloud capabilities provided by more than one CSP, therefore standardisation is of more importance than for other cloud deployment models.

2.3.4 Community cloud

Multiple organisations with common concerns, such as security requirements, policy, interests, and/or missions, may share cloud infrastructures across administrative domains to form a community cloud. The infrastructure constituents of a community cloud can be managed by a partner organisation or by a third party.

2.4 Tower architecture and enabling technologies

The tower architecture, illustrated in Figure 2-2, is a cloud computing system consisting of layers of services. The virtualisation layer sits directly on top of hardware resources and sustains high-level cloud services.

The IaaS layer provides an infrastructural abstraction with a set of standard self-service interfaces for self-provisioning, controlling, and management of virtualised resources. These self-service interfaces allow stakeholders to directly communicate with virtualised resource instances. Platform CSPs provide another layer upon the infrastructure layer with full support for cloud application development lifecycles.

PaaS cloud consumers may leverage the development platform to design, develop, build, and deploy cloud applications without requiring other complex procedures, in particular capacity planning.

The SaaS layer is the top of the cloud architectural tower and delivers specific applications as a service to end users. At the right wing of the tower, a stack of services collectively provides self-management capabilities of the cloud system. The key feature of a self-managing cloud system is dynamic capacity planning, which is underpinned by monitoring and accounting services. Capacity planning hides complex infrastructural management tasks from users by automatically scaling in and out virtualised resource instances in order to enforce established SLA commitments.

Finally, security should apply at each of the service delivery layers to ensure authenticated and authorised cloud services are delivered. Cloud security services also include user account management and Single Sign On (SSO). This towered architecture also implies that the more self-management facilities a cloud system supplies, the less self-service functionalities the cloud system enables. Therefore the common enabling technologies of the cloud computing model include virtualisation, monitoring, metering and accounting, capacity planning SLA management, and programming APIs and abstractions.

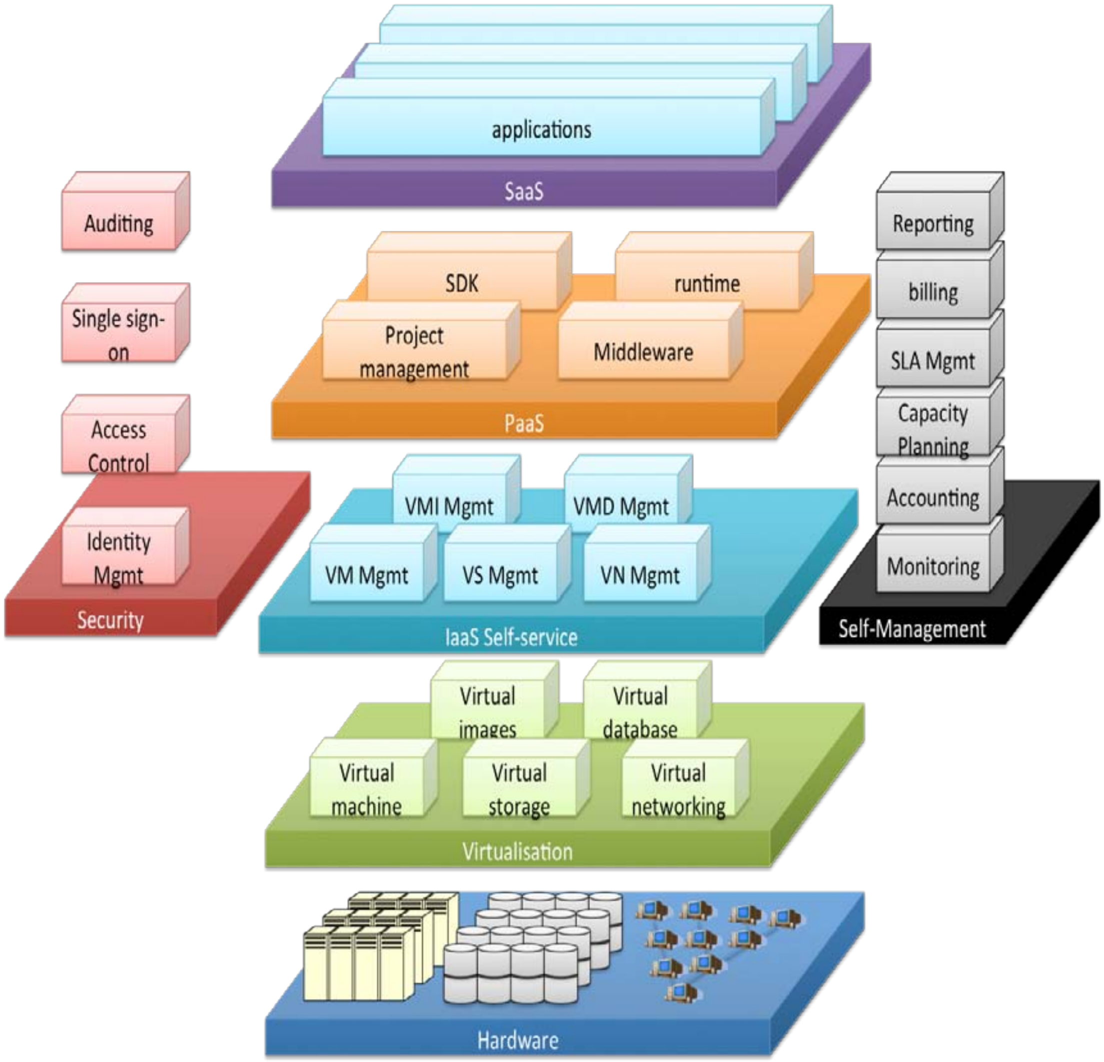

FIGURE 2-2: CLOUD SYSTEM ARCHITECTURE TOWER

2.4.1 Virtualisation

Virtualisation is a key enabling technology of cloud computing. Basic to the understanding of virtualisation is the emulation of multiple computers on a single physical machine. This type of virtualisation is known as full virtualisation since it allows an operating system to be installed in an isolated virtual environment. Multiple emulated computers are called virtual machines.

There are many virtualisation softwares available. Some can be installed directly on a bare metal environment, while others require hosting operating systems. The bare-metal virtualisation gives higher and near-native efficiency than hosted virtualisation in that it directly accesses hardware rather than going through an operating system. The bare-metal virtualisation solution places an additional virtualisation layer, the Virtual Machine Manager (VMM) or Hypervisor layer, between guest operating systems and underlying hardware. This solution was challenged in the virtualisation of x86 hardware, because of the fact that x86 operating systems are designed to run directly on hardware. There are three main alternative techniques available for handling this challenge: binary translation, para-virtualisation, and hardware-assisted virtualisation. The binary translation was firstly proposed and implemented by VMware [11] which translates non-virtualisable system calls from guest operating systems into a sequence of instructions that are suitable to be run on virtualised resources. Para-virtualisation, also known as Operating System-assisted virtualisation, modifies the kernel code of an operating system so that it can more closely resemble physical machines. However, para-virtualisation can only support a few operating systems. Finally, hardware vendors are rapidly embracing virtualisation features (e.g. Intel VT-x and AMD-V) through which a VMM can accept sensitive system calls from operating systems directly. Both Kernel Virtual Machine (KVM) [12] and Xen [13] use para-virtualisation and can do hardware-assisted full virtualisation.

Virtualisation is not limited to the emulation of a complete computer machine. There are many different virtualisation technologies related to other hardware resource types, such as storage virtualisation, database virtualisation, memory virtualisation, and networking virtualisation, each of which provides an abstraction layer between virtualised resources and physical resources. These virtualisation technologies are not detailed in this report.

Although virtualisation is the key enabling technology of cloud computing, it is not a cloud. A cloud computing system may employ more than one virtualisation technology, each of which has custom interfaces defined for virtualised resource management. The cloud infrastructure layer (the IaaS) as defined in the architecture tower (Figure 2-2) aims at providing a higher-level abstraction and a set of standard self-service interfaces for provisioning, control, and management of virtualised resources hosted by various virtualisation technologies. These well-defined interfaces can be used by CTPs and CSPs to provide advanced self-management capabilities such as auto-scaling and SLA enforcement.

2.4.2 Monitoring

Monitoring is another import technological aspect in cloud computing. Monitoring is a lower-level service that aims at providing real-time system performance information. With a monitoring service, CSPs can measure performance and availability of the overall cloud system. The monitoring service can also be used by users to measure the performance of provisioned virtual machines, and to determine whether new virtual machines need to be started. Monitoring processes can take place at the service level and provide information for SLA management.

2.4.3 Metering and accounting

In a cloud system, everything is metered and accounted. Metered usage information provides for flexible pricing and charging in a public cloud environment. Historic usage information together with monitoring information is critical to automate capacity planning and ensured agreed QoS.

2.4.4 Security

According to the analysis [14] from Ganter, cloud computing is fraught with security risks. However, these risks are more related to trust, governance, and political issues than to technical issues. Technically speaking, a cloud system is responsible for user data privacy, authenticated and authorised service access. SSO also helps users to enjoy different types of cloud services without signing in each time.

2.4.5 Capacity planning

Cloud users can use self-service interfaces for capacity planning, or self-management facilities provided by either CSPs or third-party CTPs for dynamic resource provisioning by specifying high-level policies, rules, or SLAs. In the latter case, the burden of capacity planning shifts to the Cloud providers' side. For CSPs, well-defined strategies of capacity planning are critical to successful business. Nevertheless, users may use utility functions to combine different QoS metrics and maximise their usage of paid Cloud services.

2.4.6 Programming APIs and platform

Programmability is essential to exploit cloud features. There are two types of programming APIs a cloud computing system is required to provide: self-service APIs allowing developers take care of scalability and autonomic capabilities themselves, and self-management APIs that allows developers to call auto-scaling and autonomic capabilities provided by CSPs. These self-service and self-management APIs can be packaged as a programming platform with enhanced support for features such as design, debugging, and testing.

2.5 Clarifications

So, what is cloud computing? How does it differ from other computing models?

In order to better understand the essence of cloud computing, we use an analogy to Web 2.0, which caused many arguments when it was firstly proposed and when it was in the same situation as the cloud is in. First of all, cloud or cloud computing, as Web 2.0, is a business term which refers to the way computing services are delivered and used, rather than to a new technology. Both Web 2.0 and cloud are underpinned by existing computing technologies and communication protocols. The emergence of Web 2.0 brought a new RESTful design pattern. Similarly, cloud applications need to be rethought in order to get the full benefits of cloud computing, in particular high scalability.

Regarding its relations to other pre-existing computing models, cloud computing can be used in principle to deliver computing services, including grid computing services, high-performance computing services, etc. This does not imply that cloud computing is a replacement for grid computing, high-performance computing, or any other pre-existing computing model. Instead, cloud computing should be thought of as a new business model that aims at service delivery in a highly scalable and highly flexible manner. The difference between cloud computing and other pre-existing computing models can be better demonstrated by using the towered architecture and its key enabling technology of virtualisation. Although other enabling technologies may also employed by pre-existing computing models, they are not fundamental in differentiating cloud computing.

3 Towards Cloudy Research

Due to its strong commercial nature, the concept of cloud computing and its use for research is still confusing research communities. In order to analyse technical requirements of using cloud for research, a major part of this project reviewed current practices of researchers and HE institutions (HEIs) as early adopters of using cloud for research. The project focussed particularly on those HEIs that provide cloud computing services and use public cloud capabilities for research.

The following is the summary requirements analysis based on responses received from our example cases. Further details of these cases are given in this report as Appendix A.

These example cases have also been informed by the "Using Cloud for Research" report. Further details from the user's perspective can be found in the "Using Cloud for Research" report, along with its separate annexes.

3.1 Cloudy research environments

Based on interview results from researchers, HEIs, and standards bodies, it became clear that the research usage of cloud computing is different from commercial usage.

The differences from commercial cloud usage are illustrated in Figure 3-1. In contrast to the roles defined in a typical commercial cloud environment as discussed in section 2.1, a typical cloudy research environment involves four broad active roles, the HEIs, researchers, public CSPs, and standards bodies.

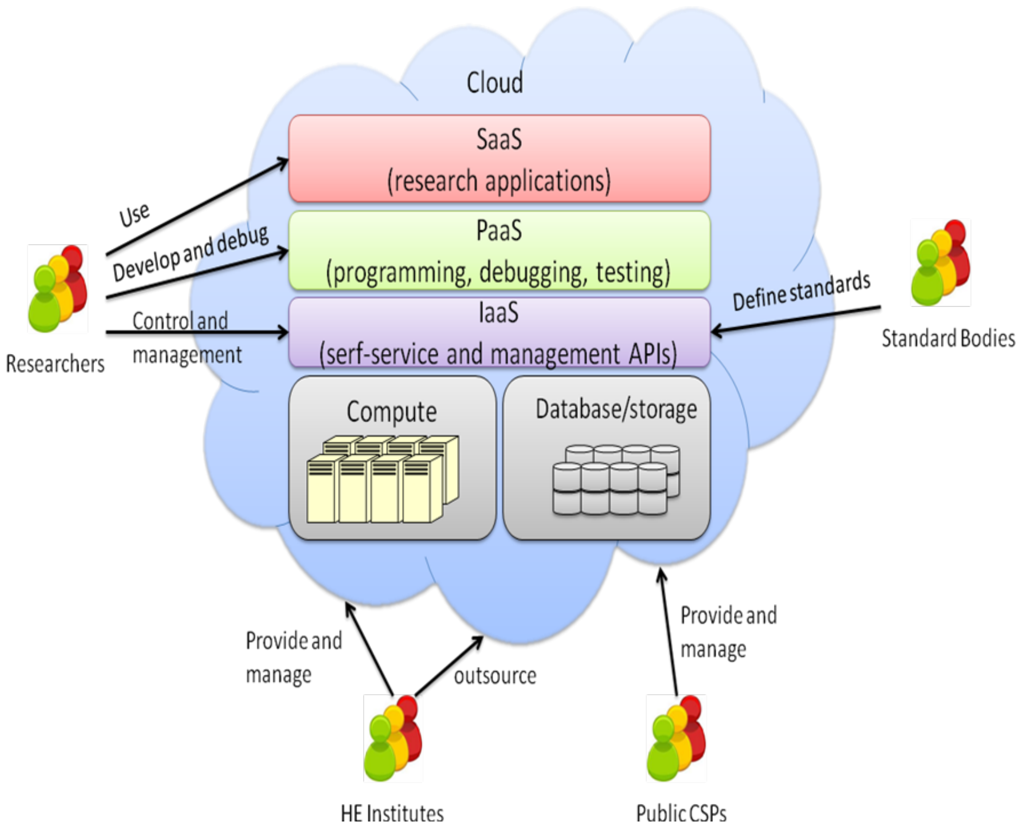

Figure 3-1: Cloudy research environment and roles

Rather than explicitly separating end users, developers, and tool providers, researchers are able to take on the responsibilities of all these roles, and have full control throughout the cloud application development lifecycle. HEIs can build private cloud services upon existing IT infrastructure, while also being end users of public CSPs, for example to burst and offload to public cloud services to meet fluctuating demands. Therefore, the HEI is defined as a separate role from public CSPs. Interoperability is of utmost importance in such a cloudy research environment so that shared research services can be deployed in a consistent manner across multiple HEI private cloud infrastructures.

3.2 Categorisation

Case studies in this section are based on six categorised Research Case Scenarios (RCS) as listed in Table 3-1. Such categorisation helps the identification and abstraction of common technical requirements from example cases.

Table 3-1: Categorised list of research case scenarios

Category ID	Category Name	Short Description	Applied Roles
RCS-1	Cloud outsourcing	Researchers using cloud capabilities (compute, storage, platform) provided by public CSPs to develop, test, or run research applications.	Researcher, public CSP, Standards body
RCS-2	Cloud bursting	HEIs own research computing services while bursting and offloading to public cloud services due to fluctuating demands.	Researcher, HEI, public CSP, Standards body,
RCS-3	Private cloud	HEIs own research cloud computing services shared inside an institution only.	Researcher, HEI
RCS-4	Hybrid cloud	Cases involving both private cloud and public cloud.	Researcher, HEI, CSP, Standards body
RCS-5	Community cloud	Multiple private clouds with shared requirements and interfaces. A federation of multiple private clouds.	Researcher, HEI, CSP, Standards body
RCS-6	Cloud tool/service provisioning	Provisioning of self-management facilities, programming abstraction tools, debugging tools, and other platform services to public and/or private clouds	Researcher, HEI, CSP, Standards body

(NOTE: These usage scenarios are categorised based on the example cases reviewed so far.)

3.3 Case studies

The following cases analyse technical requirements according to the categories of Table 3-1. Further details of example usages can be found in Appendix A and the final report of the "Using Cloud for Research" group. It should be noted that the key requirements discussed in each example case scenario are from cloud providers' perspectives, including both HEIs and public CSPs.

3.3.1 Cloud outsourcing

In this scenario (Figure 3-2), researchers develop and run research application by using public cloud capabilities delivered by one or more public CSPs. Researches can either use IaaS self-service to upload and deploy a virtual machine image, which packages the runtime environment, application component, and configuration details as a single deployment object, or leverage a development platform to develop research application directly upon a cloud infrastructure. Although a researcher might use cloud capabilities across multiple CSPs, there is no inter-communication between two public cloud services at the infrastructural level. For example, a research application is developed and run on Windows Azure, while saving experimental results to the Amazon S3 service. It is the researcher's responsibility to call service interfaces inside application codes.

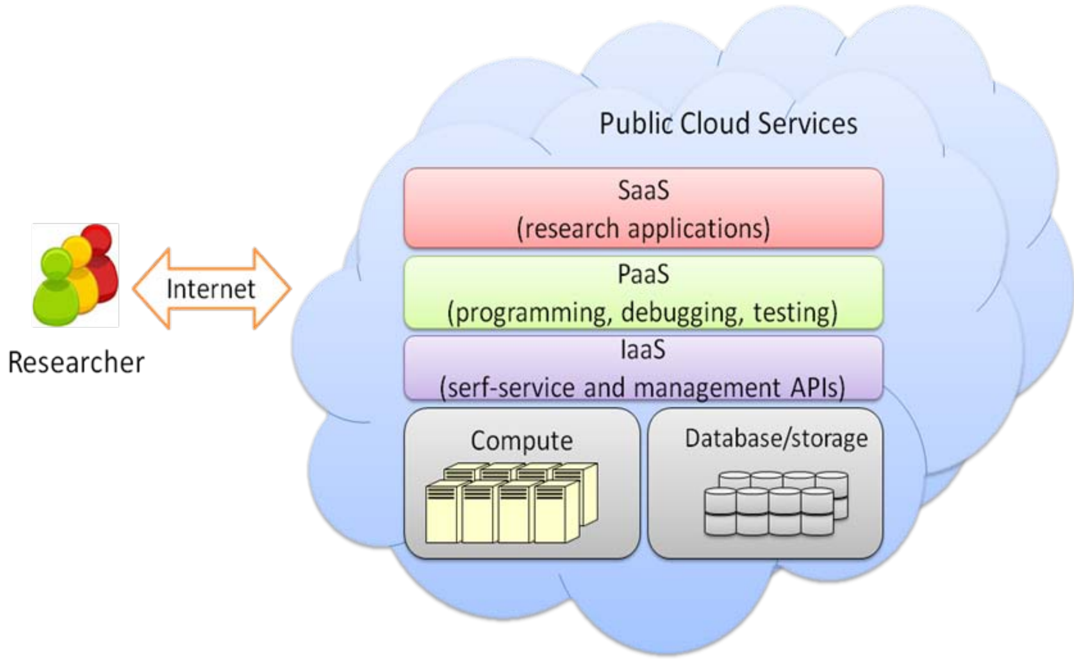

Figure 3-2: Research case scenario – Cloud outsourcing

Example cases – A.1, A.2

Technical requirements

- Security – The CSPs must provide security mechanism to authenticate and authorise a researcher's access to cloud infrastructures. Data integrity is also required to secure data transfer over the Internet.
- Programming APIs – CSPs need to provide either virtual machine image management APIs allowing researchers to upload and deploy research applications or a Software Development Kit (SDK) platform for cloud research application development.

- Monitoring – CSPs need to provide monitoring interfaces for researchers to monitor execution performance in real-time. PaaS CSPs are also required to provide real-time usage information to adjust system performance so that the SLA is achieved.
- Metering and Accounting – CSPs are required to meter service usage information for charging and billing purposes, as well as providing historical usage information in support of capacity planning.
- Interoperability: it is preferable to have a single self-service and management API to control and manage cloud runtime environments across multiple public CSPs.
- SLA management – CSPs must ensure the guaranteed services they agreed to provide.

3.3.2 Cloud bursting

In this scenario (Figure 3-3), a HEI provides research computing services to researchers, while bursting and offloading to public cloud services to meet fluctuating requirements. Public cloud services are invisible to researchers. This scenario is different from the hybrid cloud scenario since no private middleware is deployed on research computing facilities inside a HEI. A HEI may also outsource to multiple CSPs. A Grid-Cloud bridge is the mostly interesting example case. Most UK HEIs have a campus cloud deployment, and are interested in delegating computational tasks to public cloud services when their local system is overloaded.

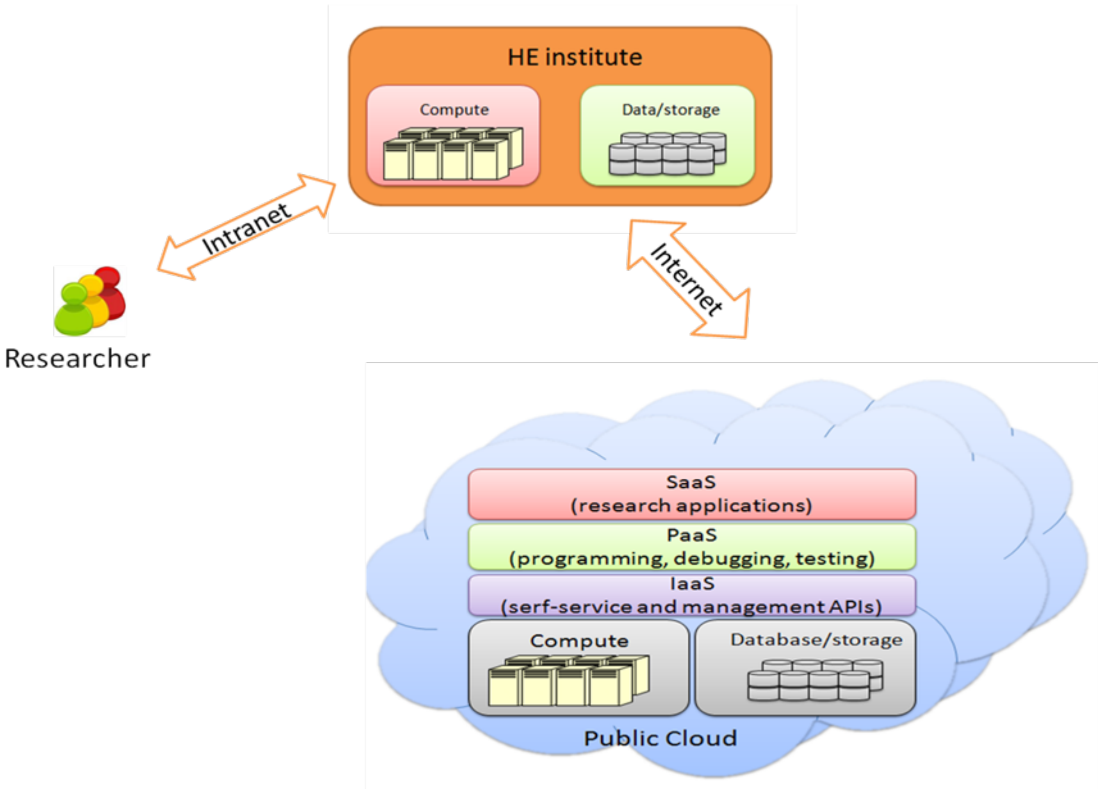

FIGURE 3-3: RESEARCH USAGE SCENARIO – CLOUD BURSTING

Example Cases – A.3

Technical requirements

- Technical requirements for public CSPs are same as USC-1. Additional requirements from HEIs' perspectives are listed as follows.
- Security – HEIs need to consider to how to integrate local security policies and mechanisms to the public cloud security framework. For example, the Grid-cloud bridge usage requires the establishment of mutual authentication and enforced single sign-on (SSO) across local grid and remote cloud infrastructures.
- Capacity planning – HEIs should employ appropriate resource scheduling algorithms and polices based on historical service usage of both HEI data centre and public cloud service usage.

3.3.3 Private cloud

In this scenarios (Figure 3-4), HEIs leverage virtualization technologies and open-source Cloudware to build a private research environment to researchers inside an organization only. This scenario is widely adopted by HEIs to increase local resource utilisation.

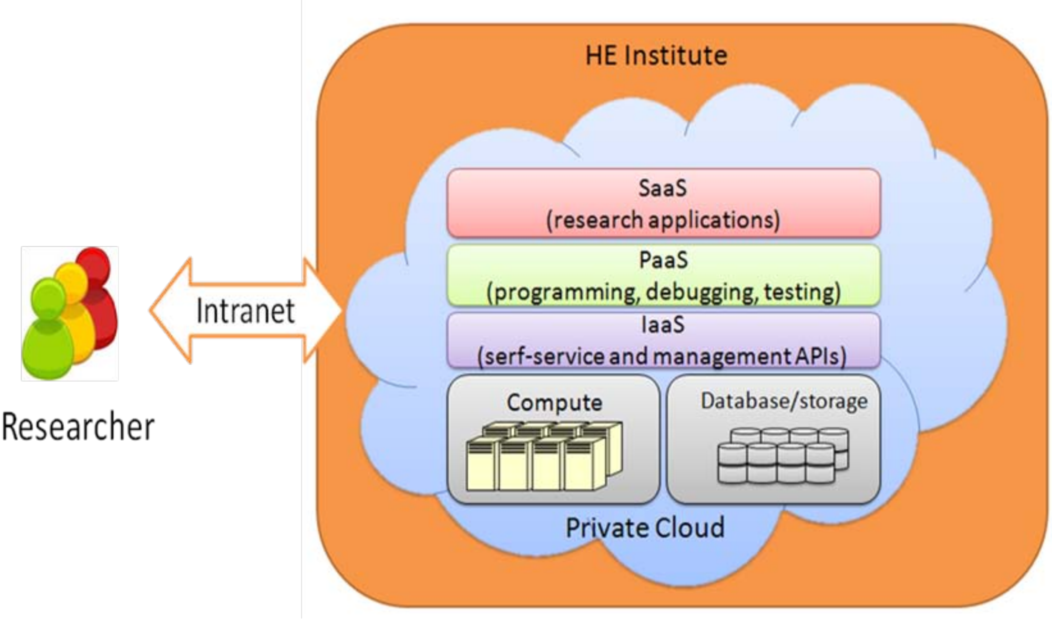

FIGURE 3-4: EXAMPLE USAGE SCENARIO – PRIVATE CLOUD

Example cases – A.4, A.5, A.6, A.7, A.10

Technical requirements

- As RCS-1 and RCS-2
- Programming abstractions – HEIs should provide common programming abstractions to avoid re-engineering existing research applications in the private cloud.

3.3.4 Hybrid cloud

This scenario (Figure 3-5) involves multiple clouds working together including, in particular, cloud services provided by public CSPs and HEIs. This scenario is ideal for HEIs to maintain a certain degree of control over the cloud infrastructure. For example, by keeping research data in local storage cloud services, HEIs maintain full control over data privacy.

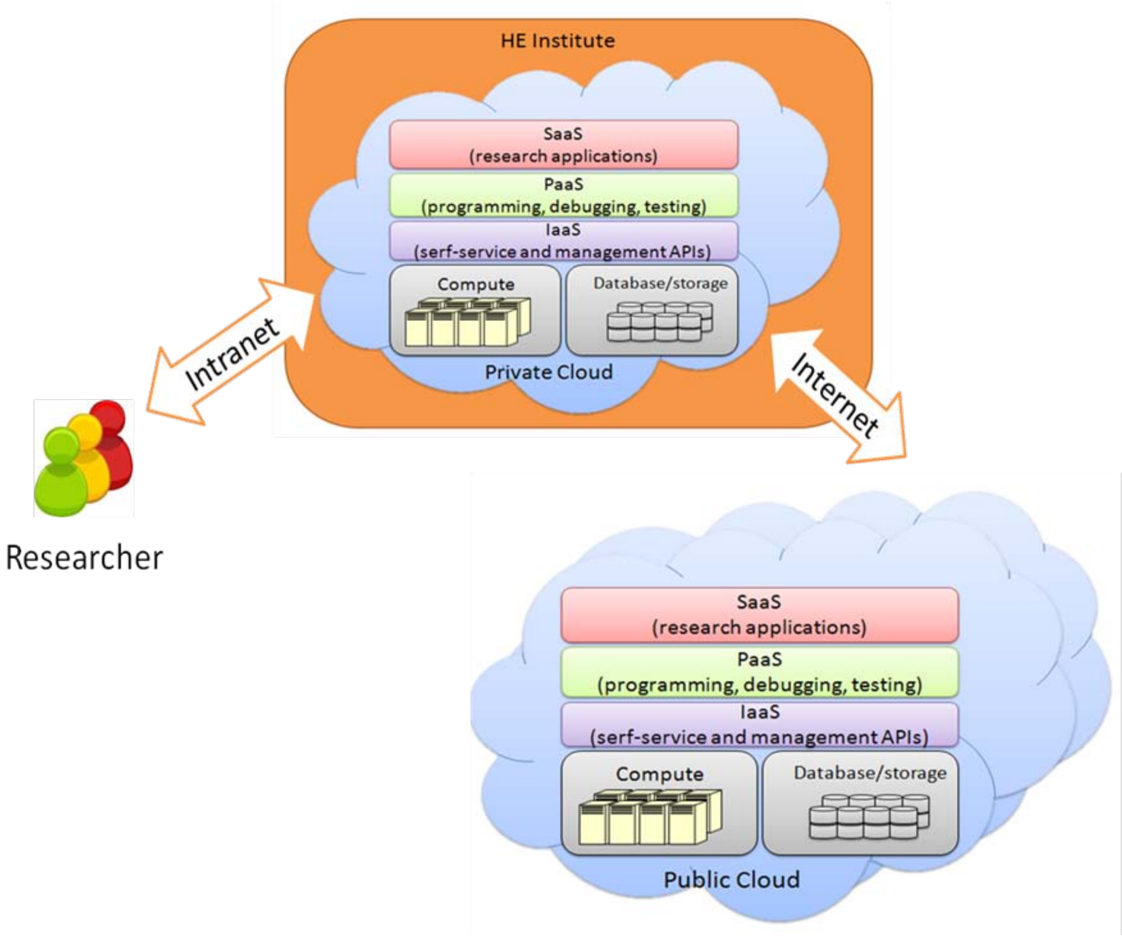

FIGURE 3-5: EXAMPLE USAGE SCENARIO – HYBRID CLOUD

Example cases – No example cases available for this scenario yet.

Technical Requirements

- As RCS-3

3.3.5 Community cloud

This scenario (Figure 3-6) is a research-specific vision that allows HEIs with the same research scopes or common requirements to share cloud services and management tasks.

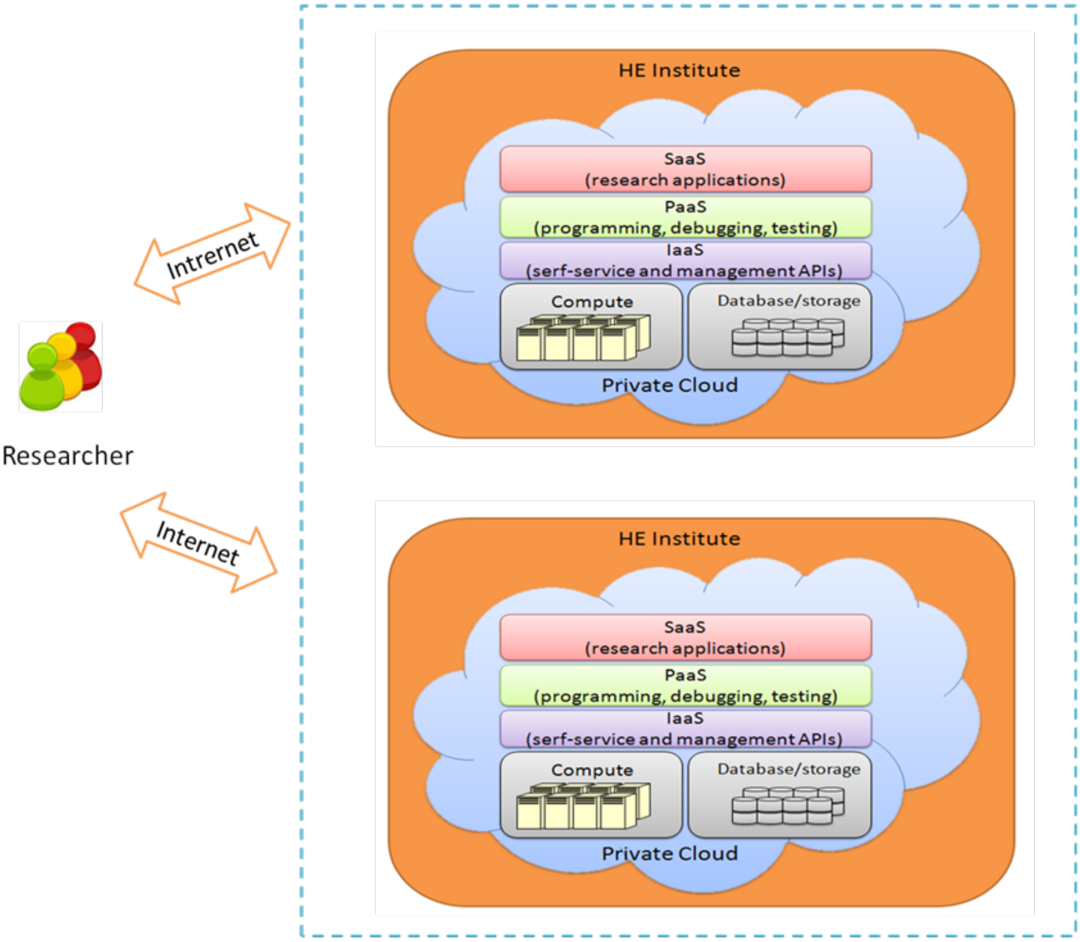

FIGURE 3-6: RESEARCH CASE SCENARIO – COMMUNITY CLOUD

Example cases – A.8

Technical requirements

- As RCS-3
- Federated Identity Management – Federated identity and management is required in this scenario to enable researchers from one domain to access cloud services provided by another domain and act as a community.

3.3.6 Cloud tool/service provisioning

This scenario does not provide any infrastructural cloud services, but focuses on particular unified self-management services and development platforms to ease management tasks and software engineering. Such tools or services are normally not tied to any specific cloud providers. For example, a general-purpose federated identity management system can be used in a community cloud environment.

Example cases –, A.9, A.11, A.12

Technical requirements

- As RCS-3

4 State of the Art

This section reviews current offerings of public cloud services, open-source Cloudware, third-party services, and standards. The content of this section is based on interviews, invited seminars, online virtual events, attended conferences, organised workshops, and literature review.

The following discussion focuses on the technical aspects of current offerings. Further details from the user's perspective can be found in the "Using Cloud for Research" report.

4.1 Commercial cloud offerings

This section reviews cloud capabilities offered by commercial CSPs, mainly IaaS and PaaS cloud providers. SaaS cloud providers are intentionally omitted because there are no research-specific cloud application services observed. The review of commercial cloud offerings is based on the list of public CSPs provided in Table 4-1.

Table 4-1: List of public commercial cloud services and offerings

Vendor	Cloud Service	Description	Target Delivery Model	Target Deploy-ment Model
Amazon	Amazon Web Services	Amazon Web Services offers a package of public cloud capabilities that delivers a set of highly-profiled infrastructure services.	IaaS	Public
IBM	Smart Business Develop-ment and Test Cloud	IBM's smart business test cloud provides a solution for enterprises to build a flexible and cost-efficient cloud-based development and test environment.	PaaS	Private/ Hybrid
Micro-soft	Azure	Microsoft Azure is a PaaS cloud that offers a development environment for developers to create cloud applications and services in a highly flexible and scalable manner.	PaaS	Public

4.1.1 Amazon Web Service

Amazon Web Services (AWS) is currently the public CSP with the highest profile that delivers comprehensive infrastructural cloud services. As illustrated in Figure 4-1, the cloud capabilities implemented by AWS are well matched to the tower architecture (Figure 2-2).

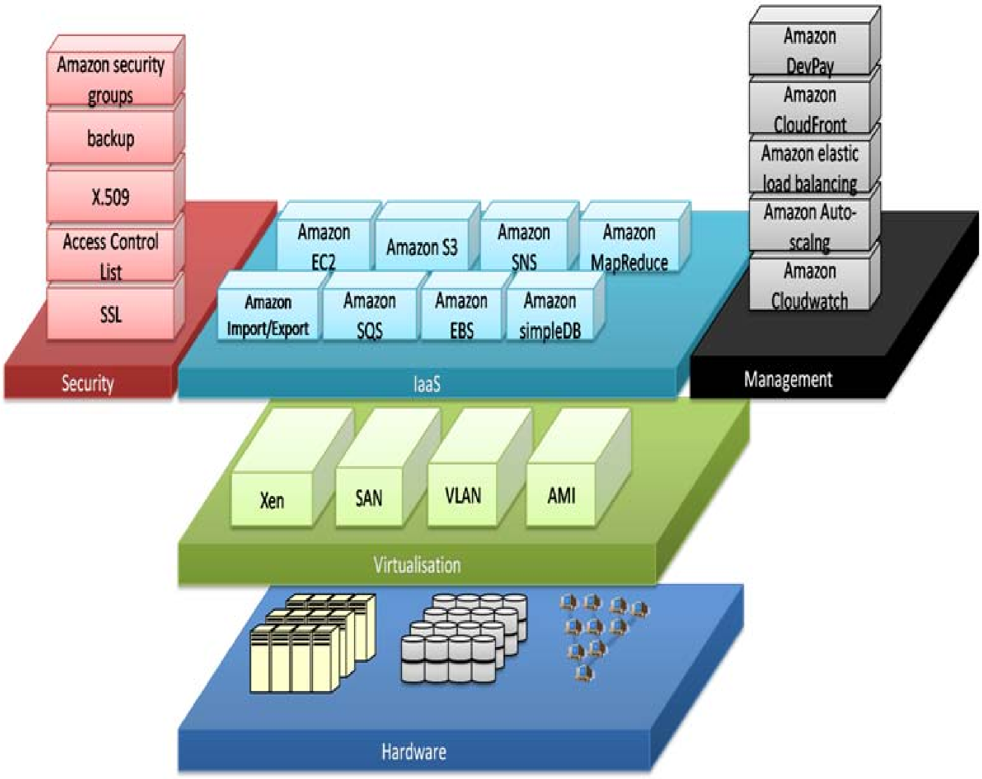

FIGURE 4-1: AMAZON WEB SERVICE CLOUD CAPABILITIES AND LAYERED ARCHITECTURE

AWS cloud services consist of a set of IaaS service and management services as follows:

- Amazon Elastic Compute Cloud (EC2). This is a Web service that provides CaaS capabilities. EC2 supports the Xen hypervisor as a virtual machine manager, which supports operating systems including Linux, Sun Microsystems' OpenSolaris, and Microsoft Windows Server 2003. EC2 interfaces provide complete control over virtual machine instances and on-demand scaling. Amazon auto scaling enables automatic EC2 capacity scaling up or down according to predefined conditions.
- Security. Within EC2 this is provided on multiple levels, by the operating system (OS) of the host system, the virtual instance operating system or guest OS, a stateful firewall, and signed API calls. With Amazon Elastic Load Balancing (ELB), incoming traffic can be automatically distributed across EC2 instances in response to traffic load. Amazon CloudWatch is used for monitoring EC2 instances and Elastic Load Balancers in real-time.
- Amazon Simple Storage Service (S3). This delivers Storage as a Service in the cloud. Amazon S3 is based on an existing data storage infrastructure used to run Amazon's global network of Web sites, and provides interfaces for storing and retrieving data. The content delivery service, Amazon CloudFront, works with

Amazon S3 for delivering static and streaming content using a global network of edge locations giving optimal performance. For moving large amounts of data, the AWS Import/Export service can be employed to use Amazon's high-speed internal network, bypassing the Internet. Amazon S3 APIs only permit authenticated access by the bucket and/or object creator, and provides both bucket- and object-level access controls. Amazon S3 access can be granted based on an AWS User Account, a DevPay Product ID, or an open-to-public access. Data stored in Amazon S3, Amazon SimpleDB, or Amazon EBS are automatically backed up in multiple physical locations.
- Amazon Elastic MapReduce. This is a web service that enables businesses, researchers, data analysts, and developers to easily and cost-effectively process large amounts of data on a hosted Hadoop [35] framework running on the web-scale infrastructure of Amazon EC2 and Amazon S3.
- Amazon Simple Database. This is a web service providing the core database functions of data indexing and querying in the cloud. Unlike traditional database management systems, Amazon SimpleDB is much simpler and requires no schema definition.
- Amazon Virtual Private Cloud (VPC). This enables cloudburst by dynamic recruitment of EC2 instances to existing infrastructure inside an organization. Amazon VPC is a secure bridge between on-premises infrastructure and a set of isolated Amazon EC2 instances via a Virtual Private Network (VPN).
- Amazon Relational Database Service (RDS). This provides simple interfaces to set up, operate, and scale a relational database in the cloud. RDS is fully compatible with MySQL databases, making it easy to transfer existing MySQL databases into Amazon RDS. Amazon RDS automatically patches the database software and backs up the database for a user-defined retention period. Amazon RDS flexibly scales EC2 or S3 capacity associated with relational database instances via a single API call.
- Amazon Simple Queue Service (Amazon SQS). This offers a special data management facility for storing and queuing messages between virtual machines. Using Amazon SQS allows automated workflows, working in close conjunction with the Amazon EC2 and the other AWS infrastructure web services.
- Amazon Simple Notification Service (Amazon SNS). This implements another messaging model, the subscription model, which can be employed by cloud applications for event handling, event-driven workflow, and application monitoring.

4.1.2 Microsoft Azure platform

Microsoft Azure Platform provides PaaS cloud capabilities with developer-accessible services for cloud application development. As illustrated in Figure 4-2, Windows Azure platform uses the Azure hypervisor, based on Hyper-V [15].

PaaS is a layer on top of IaaS, making it easier to start development without being concerned about servers, networking, storage, etc. IaaS works at the infrastructure layer, providing an abstracted underlying virtualisation layer, allowing developers to control infrastructure capabilities. For example, when building cloud applications upon Amazon EC2, a developer is required to create and deploy an appropriate virtual image containing an operating system, software stacks, and other runtime dependencies. With PaaS, these tasks are automatically done by CSPs. Compared to IaaS, however, PaaS gives less control over the cloud infrastructure.

The core components of Azure platform contain three parts:

- Microsoft Azure compute service and storage service. These support Windows applications and the storage of application data. Rather than creating VMs or supplying custom VM images manually, Windows Azure creates two 'roles' of VM instances for developers, the work role and the Web role. Each VM instance has an agent deployed to contact underlying Azure fabric, which provides programming APIs to allow applications to scale out with more VM instances. A built-in load balancer spreads incoming requests across Web role instances of the same application. Azure also provides data storage, allowing cloud applications to save data in three different types, the Azure blob, Azure queue, and Azure table. These three storage services have well-defined programming APIs on a dot-Net framework.

- SQL Azure. This is a database service enabling Azure applications to use relational database features such as those provided in a SQL server database. Developers may access databases through Tabular Data Stream (TDS) or existing SQL server client libraries, including ADO.NET, ODBC, and PHP.
- Azure Platform AppFabric. This provides two services, the Azure service bus and access control service, for publishing applications as discoverable services with custom access control policies.

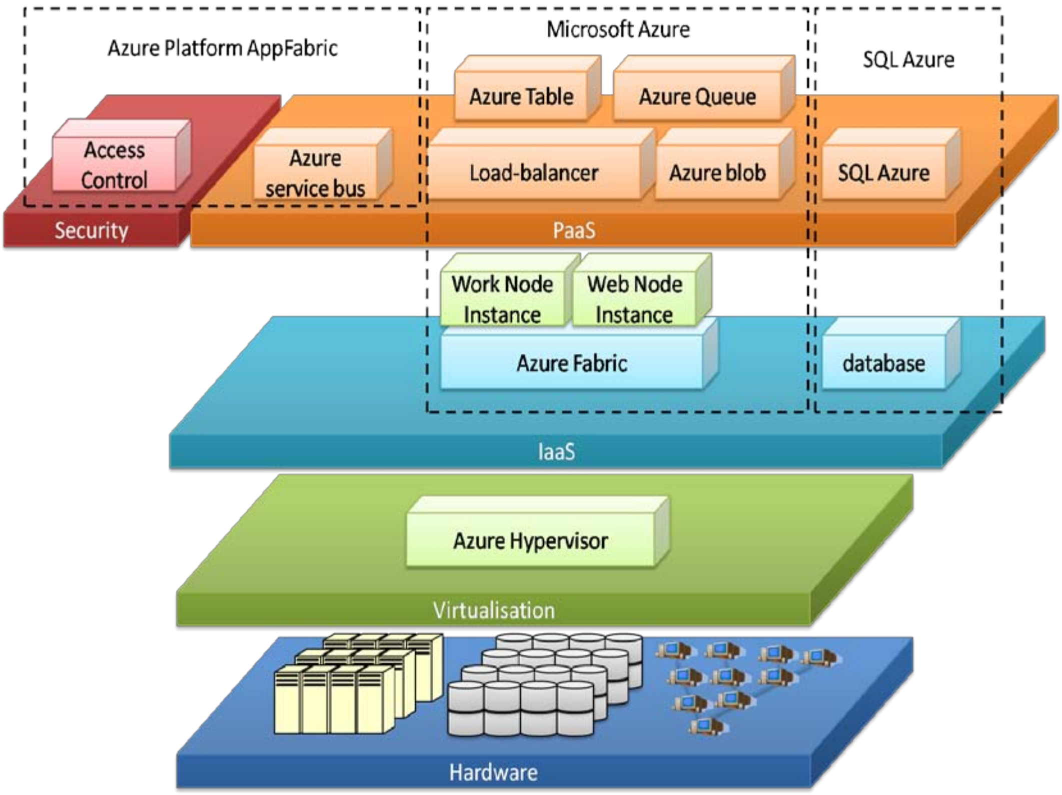

FIGURE 4-2: MICROSOFT AZURE PLATFORM ARCHITECTURE

4.1.3 IBM Smart Business Cloud

The cloud strategy of IBM focuses on specialised software and purpose-built appliances to package private and hybrid cloud solutions for enterprise data centres with a flexible and cost-efficient, cloud-based development and testing environment. This strategy is delivered as four main solutions:

- Smart Business Development and Test on IBM Cloud. This is a public cloud environment to help developers to quickly set up a test environment with a set of standard services delivered on the IBM cloud.

- Smart Business Development and Test Cloud. This offers services to implement a private or hybrid cloud using a set of IBM automated management solutions and cloud strategies, including Tivoli Service Automation Manager (TSAM), Tivoli Provisioning manager, IBM Tivoli Composite application manager (ITCAM), Websphere Cloudburst, and IBM Cloudburst.
- Websphere Cloudburst Appliance. This is purpose-built hardware and software for speedy application deployment to cloud and on-premises virtualisation environments. It consists of two main components: the IBM hypervisor edition Images and the WebSphere CloudBurst Appliance. Hypervisor Edition Virtual Images are virtual images that are packaged and contain an operating system, a WebSphere Application server, an IBM HTTP server, and an activation engine. The activation engine is able to turn a single image into multiple personalised images by applying custom configurations input by users. These virtual images are packaged for three different hypervisor platforms, VMware, IBM PowerVM, and IBM z/VM. WebSphere CloudBurst Appliance is hardware that sits in a data centre and provides secure, self-service cloud management services both internally and externally. The CloudBurst appliance is mainly used to dispense hardened WebSphere patterns into a pool of virtualised resources running a supported hypervisor (i.e. VMware, PowerVM, z/VM).
- IBM CloudBurst. This is a line of pre-integrated service platforms including hardware, storage, networking, virtualisation, and service management software, which enables an enterprise to create private cloud environments. It is a purpose-built infrastructure upon proven technologies, including pre-packaged software and hardware, self-service and featured automated management facilities, automated operation for rapid deployment, and a set of pre-packaged virtual images.

Technical details of individual IBM cloud solutions are intentionally omitted here, because of their commercial nature and “lock-in” to branded IBM platforms. Further technical details can be found in Table 4-2.

4.1.4 Technical features of commercial offerings

Table 4-2 gives a comprehensive comparison of technical features of the three representative CSPs offered.

TABLE 4-2: COMPARISON OF TECHNICAL FEATURES OF COMMERCIAL CLOUD OFFERINGS

Vendor	Virtualisation solution supported	Operating system supported	Monitoring	Programmability	Security	Metering & accounting	Auto-scaling	Standards compatibility	Virtual image management
Amazon	Xen, KVM, VMware, SAN, VLAN	Linux. Windows, OpenSolaris	CloudWatch	Java API, bash CLI	VM isolation	NO	Auto-scaling interface (custom impl. Required)	EC2 APIs	ESB
IBM	PowerVM, z/VM, VMware	Windows, Linux	Tivoli Monitoring	NO development framework	Tivoli Federated Identity Manager	Tivoli Usage and Accounting Manager	Tivoli Provisioning Manager	NO	NO
Microsoft	Azure Hypervisor	Windows	NO	.Net, Java, PHP	Azure Platform AppFabric	NO	NO (configuration +Azure Fabric API)	NO	NO

4.2 Open cloudware

Only a few research projects on developing open-source cloud middleware software (Cloudware) have been initiated recently. This section reviews the three most prominent Cloudware (as listed in Table 4-3) that can be used for building private cloud infrastructures.

Table 4-3: List of reviewed open cloudware

Organisation	Product	Description	Target Delivery Model	Target Deployment Model
Eucalyptus	Eucalyptus Systems	Open-source private cloud software enabling an organisation to establish an in-house cloud computing environment.	IaaS	Private
OpenNebula Community	OpenNebula	Standards-compatible open-source toolkit for building public, private, and hybrid IaaS clouds based on on-premises IT infrastructures.	PaaS	Private
Globus	Nimbus	Open-source cloud toolkit that allows turning a computer cluster into an IaaS cloud by employing virtualisation technologies.	PaaS	Private

4.2.1 Eucalyptus

Elastic Utility Computing Architecture Linking Your Programs to Useful Systems (Eucalyptus) [16] is an open-source Cloudware for implementing private clouds. As illustrated in Figure 4-3, Eucalyptus supports KVM and Xen hypervisors, therefore supporting Linux operating systems only. It also provides a list of self-services, each of which has a well-defined Web service interface allowing users to control and manage virtual machine instances and other hardware resources.

These self-services include:

- Node Controller (NC) – executes on each compute node, and controls the lifecycle on VM instances running on the machine. The NC component provides interfaces to start, inspect, query, and destroy VM instances. It is also responsible for managing virtual network endpoints.

- Storage Controller (SC) – implements block-accessed network storage as Amazon EBS and is able to be adapted to various storage systems including Network File System (NFS) and Internet Small Computer System Interface (iSCSI).
- Cluster Controller (CC) – manages a cluster of NC machines within a broadcast domain, and is able to schedule VM instances to compute nodes of the cluster.
- Walrus – provides EC2 S3 and AMI compatible interfaces allowing users to interface with underlying storage capabilities, set access control policies, and manage virtual machine images.
- Cloud Controller (CLC) – the entry point for system administrators, developers, and end users to manage the cloud environment.

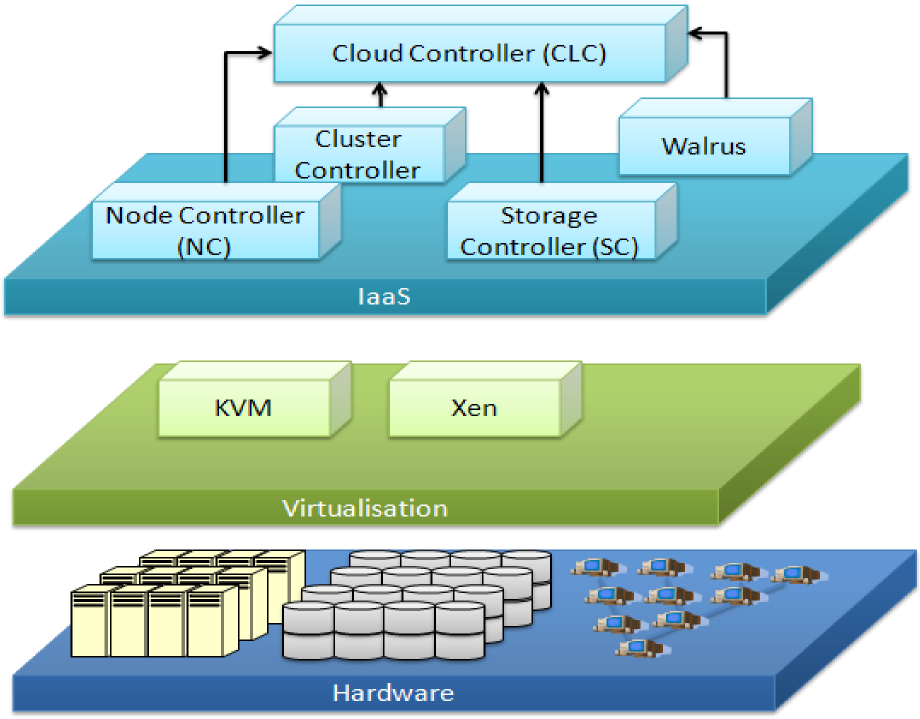

FIGURE 4-3: EUCALYPTUS LAYERED ARCHITECTURE

4.2.2 OpenNebula

OpenNebula [17] is an open-source toolkit which provides self-service interfaces compatible with the Open Cloud Computing Interface (OCCI) specification [18], EC2, and ElasticHost interfaces. It allows an organisation to build a private cloud infrastructure, publish it as a public cloud service, and interconnect it to other public cloud services as a hybrid cloud.

The layered architecture is illustrated in Figure 4-4. OpenNebula currently supports KVM, Xen, and VMware hypervisors. Its internal infrastructure components can be divided into two layers:

- Driver layer – consist of a set of abstract driver interfaces that are customized to support various virtualisation plug-ins.
- Core layer – defines a set of core and OCCI compatible APIs for management and monitoring of VM instances, physical resources, VM images, and virtual networks through appropriate driver plug-ins. A persistent generic pool based on a SQLite3 backend is used by OpenNebula as its core internal data structure.

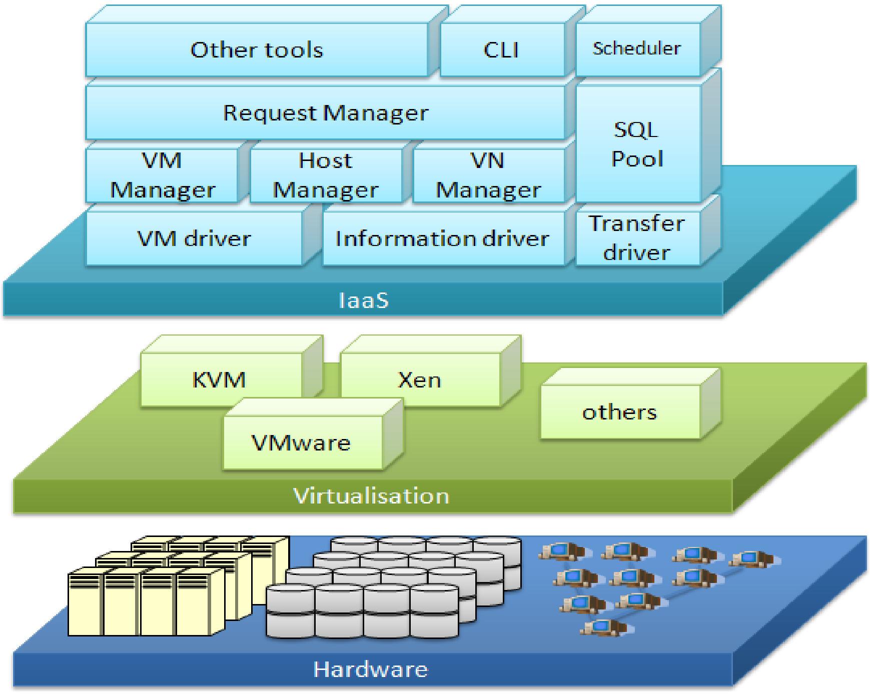

FIGURE 4-4: OPENNEBULA LAYERED ARCHITECTURE

4.2.3 Nimbus

Nimbus [19] is developed by Globus development team making it easy to turn a local cluster into an IaaS cloud and bridging it to high-level grid middleware, in particular Globus Toolkit version 4 (GT4) [20]. With Nimbus, a participant site of a grid system can be easily turned into a cloudy site. Nimbus also provides EC2 interface implementations therefore allowing users to employ EC2 virtual nodes as an elastic grid site.

As described in Figure 4-5, Nimbus supports KVM and Xen hypervisors and consists of the following core components:

- Workspace Control – an agent acts as VMM installed on each virtual machine and provides networking facilities between hypervisors.
- Workspace pilot – enables integration of VMs into resources managed by the Local Resource Manager System (LRMS) (i.e. PBS, LSF, etc.)
- Workspace resource manager – provides LRMS-like capabilities to a cluster of VMs.
- Workspace service – the elastic site manager.
- Resource manager API – abstract resource manager interfaces (WSRF, and EC2 APIs) allowing custom implementations for a specific site manager.
- Context broker – a separate service allowing a large virtual cluster to launch automatically.

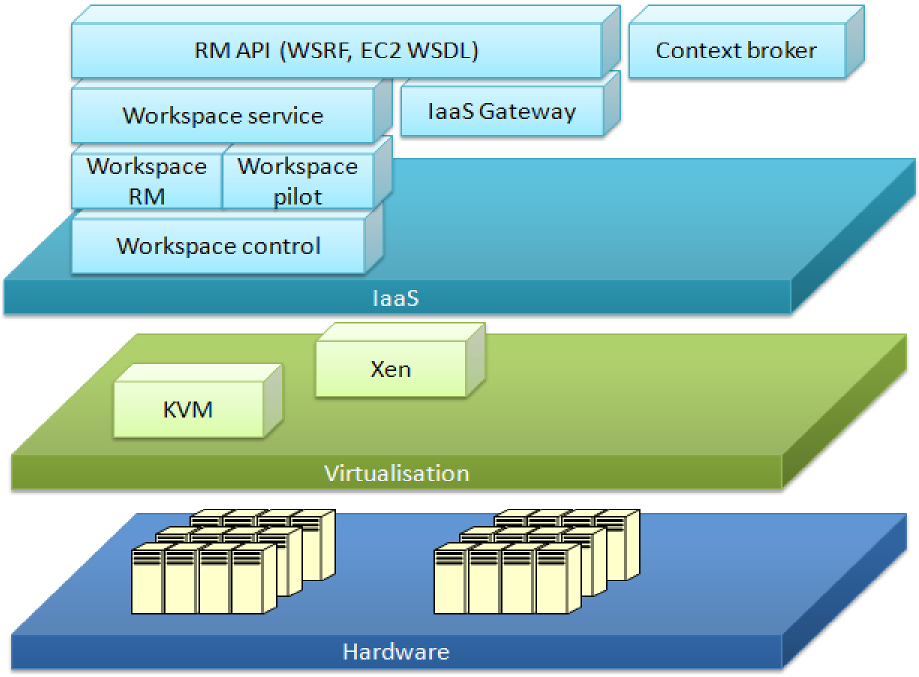

FIGURE 4-5: NIMBUS COMPONENT ARCHITECTURE

4.2.4 Open cloudware technical features

Table 4-4 gives a comparison of the technical features of the three open source Cloudware offerings.

TABLE 4-4: COMPARISON OF TECHNICAL FEATURES OF OPEN CLOUDWARE

Vendor	Virtualisation solution supported	Operating system supported	Monitoring	Programmability	Security	Metering & accounting	Auto-scaling	Standards compatibility	Virtual image management
Nimbus	Xen	Linux.	NO	WSRF	VM isolation	NO	RM API	NO	NO
OpenNebula Community	Xen	Linux	Limited LiL Limited VM monitoring	CLI	VM isolation	NO	Scheduler API only	OCCI, EC2	YES
Eucalyptus Systems	KVM, Xen, VMware (commercial version only)	Windows	Limited VM monitoring	Java API and CLI	X.509, authentication only	NO	CLC API Only	EC2	YES

4.3 THIRD-PARTY SERVICE OFFERINGS

In addition to commercial CSPs and open-source Cloudware providers, there is another group of companies that plays an important role in supplying cloud-based management services. These services normally provide unified management capabilities and can be adapted for custom management purposes.

This section reviews three cloud-based services, which deliver three important features that support the adoption of cloud computing. As shown in Table 4-5, these third-party services offer unified monitoring, auto-scaling, and security solutions across cloud environments.

TABLE 4-5: LIST OF CTPS AND CLOUD-BASED SERVICES

Vendor	Functional Domain	Description
NimSoft	Monitoring	NimSoft is the fastest growing provider of performance and availability monitoring software, and provides pubic and private monitoring solutions based on a unified monitoring architecture.
RightScale	Auto-scaling and Capacity planning	RightScale provides a management platform that enables the creation and management of scalable cloud applications across cloud environments.
CohesiveFT	Security	CohesiveFT's VPN-Cubed is claimed as the first commercial solution that enables security control in a cloud and cross clouds.

4.3.1 Nimsoft

Nimsoft [21] provides a unified monitoring architecture, upon which a set of monitoring solutions, known as Nimsoft Monitoring Solution (NMS), were developed for monitoring and management of performance in various environments including data centres and clouds.

As illustrated in Figure 4-6, the Nimsoft monitoring architecture is comprised of functional components and interfaces at three bottom-up layers: the data collection layer, application layer, and presentation layer. Each layer provides well-defined APIs and allows custom extensions.

- Data collection layer – consists of various probes that collect performance and QoS data and statistics from data centres, clouds, virtual machines, and other runtime component instances. Custom probes can be provided by implementing appropriate unified monitoring APIs.
- Application layer – includes an integrated set of solutions to which collected performance and QoS data might apply (e.g. SLA management, event reporting, etc.). Advanced solutions can be extended through unified monitoring APIs.
- Presentation layer – provides visualisation tools, portals, and dashboards to view and analyse performance and QoS data. This layer is also extensible for custom representation and display.

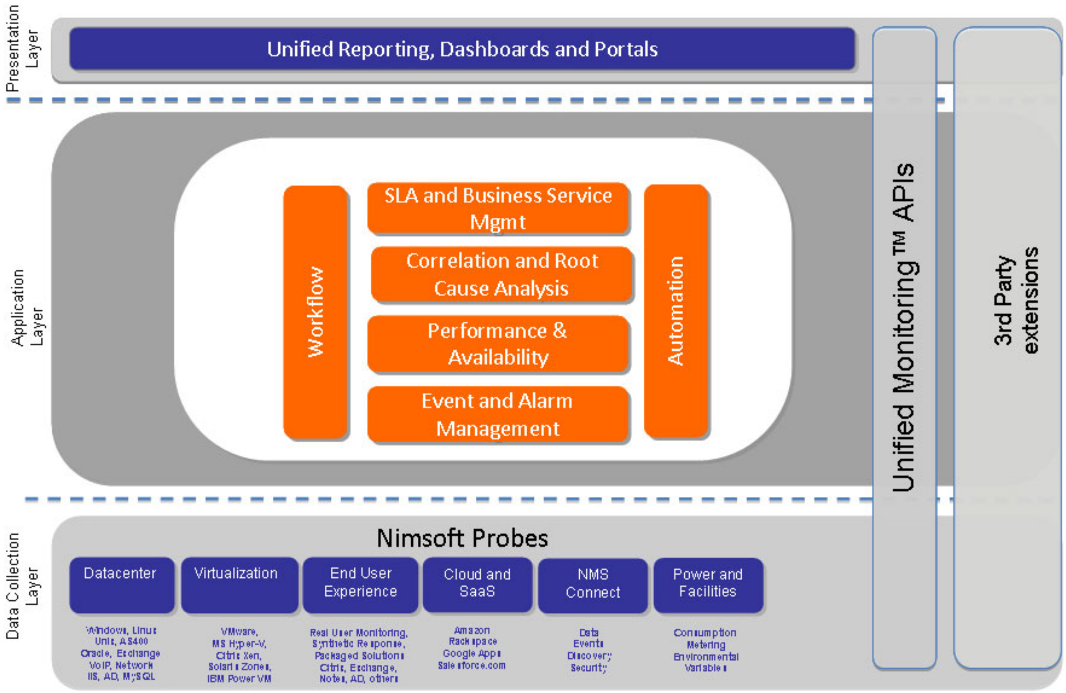

FIGURE 4-6: NIMSOFT UNIFIED MONITORING ARCHITECTURE (FROM [21])

At present, Nimsoft provides monitoring solutions for public cloud services such as Amazon EC2, S3, Microsoft Azure, Google application engine, and Saleforce CRM. It is also possible to extend Nimsoft's unified monitoring architecture to monitoring the performance and QoS of private clouds.

4.3.2 RightScale

As shown by this review, limited management services are provided by both commercial CSPs and open-source Cloudware. Some commercial CSPs may provide automated management facilities, such as IBM's smart business test cloud, but these are usually locked into specific cloud environments.

RightScale [22] provides a general-purpose cloud management platform that offers significant advantages for the cloud application lifecycle (Figure 4-7).

- At the plan stage – RightScale provides a template for planning resource capacity for a cloud application. This template allows pre-configuring servers by starting from a base machine image and then adding scripts that run during the boot, operation, and shutdown phases. This "all-in-one" template deployment solution enables users with limited knowledge to employ cloud capabilities while concentrating on application requirements.
- At the deployment stage - RightScale allows coherent management of a group of servers in a multiple-server environment for production, test, and staging. It also allows the configuration of monitoring data metrics and usage costs at the same time, as well as the configuration of software upgrading across multiple servers.
- At the management stage – Once a cloud application launches, the RightScale platform provides a management dashboard enabling real-time management of the application deployment across multiple cloud environments. It allows monitoring key metrics, setting automatic responses, and specifying thresholds.

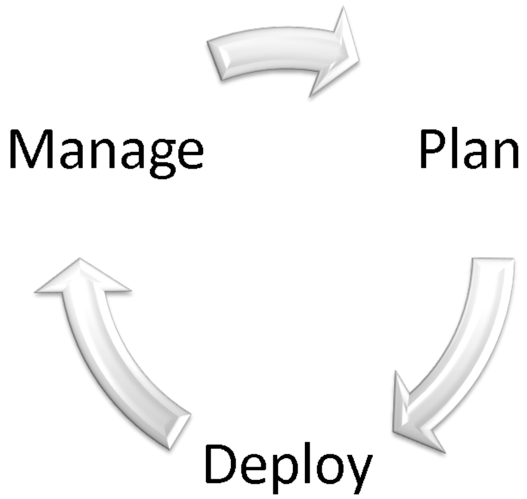

FIGURE 4-7: CLOUD APPLICATION LIFECYCLE

4.3.3 CohesiveFT

Security has been identified as a concern that slows uptake of cloud computing solutions. As discussed before, security in cloud computing is not so much a technical challenge as a trust issue, i.e. not trusting third-party management.

CohesiveFT's VPN-Cubed [23] solutions offer alternative ways to ensure security by giving customers control in public cloud environments.

- VPN-Cubed SSL – provides an overlay network that allows cloud consumers control over addressing, topology, protocols, and encrypted communications for deployed resources (i.e. VM instances) in the public cloud.
- IPsec to Cloud – a version of the VPN-Cubed overlay network, packaged to work between a data centre using IPSec "extranet" connectivity and a public cloud.

4.4 STANDARDS

Although cloud computing has been widely accepted as the next-generation computing model that provides significant value-added for both research communities and commercial domains, there are still many concerns affecting uptake.

One of the most common concerns is vendor lock-in. Most current commercial cloud capabilities offered by public CSPs were evolved from their purpose-built data centres. For example, the massive infrastructure of EC2 was built by Amazon to support its retail business. Therefore, it is difficult to have the same cloud interface for different public cloud services. With the increasing number of customers, public CSPs now do want to "lock-in" their customers and win this "cloudy" market.

According to the case studies contained in this report, concerns of interoperability and standardisation are of utmost importance to research communities. Research applications across multiple cloud services are currently limited by proprietary interfaces. HEIs bursting to public cloud offerings need to redesign their code by using cloud-specific APIs. Adopters of hybrid cloud solutions must carefully chose one or more public cloud services because migrating applications from one cloud to the other is difficult. Besides, current cloud offerings, both public commercial clouds and open Cloudware, provide limited management functions for customers. One option would be use third-party management frameworks or tools, though are also challenged by interoperability issues. The answer to these issues is interoperability through standards-based control and management protocols.

There are two cloud-related standards available at present, the Open Virtual Format (OVF) [24] defined by the Distributed Management Task Force (DMTF), and the Open Cloud Computing Interface (OCCI) specification stacks proposed by the Open Grid Forum (OGF) OCCI working group.

The OVF specification defines an open virtual machine format for packaging and distribution of virtual appliances composed of one or more VMs. OCCI standards aim at developing open APIs for control and management IaaS cloud resources. OpenNebula has claimed its compatibility to OCCI interfaces. For commercial clouds, the EC2 APIs have become the de-facto IaaS standard. Recently, leading IT companies have initiated industrial efforts on standardisation and interoperability. Multiple companies have joined the Open Cloud Manifesto [25] and announced that “the cloud should be open”.

5 Key comments and Recommendations

In this section we cover the recommendations from this report. These recommendations should be read in conjunction with the recommendations from the reports on the Review of Using Cloud Computing for Research and the Review of Environmental and Organisational implication of Could Computing in HE and Further Education.

5.1 Background

In order to understand how cloud computing can be used for research, there are a few facts which need to be appreciated.

5.1.1 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 researches are not clear to many. Although early cloud adopters have shared some useful information, including example cases analysed in this report, in the cases of the "Using Cloud for Research" project, and in those published by commercial public CSPs, most of the 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.

5.1.2 Commercial cloud or "research cloud", there is no right answer yet

There is still no right answer to the way in which researchers should use cloud. Whether public or commercial, interoperability is a key issue for cloud computing in research.

5.1.3 Current offerings are not research friendly

The current cloud computing offerings come from two main sources, public CSPs and open-source Cloudware. 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, MPI applications, and parallel computing. Although public CSPs provide powerful self-management facilities and well-defined programming APIs, these enhanced features also "lock" applications into a specific cloud infrastructure. On the other hand, researchers are starting to think of building specific cloud environments by using open-source Cloudware. These open-source offerings only provide low-level programming APIs and service interfaces, however, leaving resource management tasks to application developers.

5.2 Things to learn

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

5.2.1 From ourselves

Researchers need to learn from themselves about the reasons they need cloud computing, how they can benefit from it over other computing models being employed in HEIs, and what key technologies enable the migration to cloud computing.

5.2.2 From commercial CSPs

There are many commercial CSPs that are successful in different contexts. Amazon is the highest-profile IaaS CSP at present, while IBM has 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 into cloud service delivery.

5.2.3 From technologies

Researchers have started using open-source Cloudware to develop their own custom private environments. Considering the 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 the self-management capabilities of private cloud infrastructure.

5.3 Action to take

Figure 5-1: Roadmap towards the vision of cloudy research environment

As illustrated in the roadmap of Figure 5-1, we list five prioritised recommendations to JISC to stimulate actions leading towards UK cloudy research environments. We use the following terms to distinguish three different levels of priority.

- “SHOULD” implies top priority and the need to act immediately
- “COULD” implies medium priority that is to be planned later
- “MAY” implies least priority

Recommendation 1: JISC SHOULD continue support for proofs of concept of using cloud for research. Cloud computing is in its infancy; hence a number of proof of concept application are need in order to give a baseline understanding/measure from which to assess the usefulness of cloud computing for research.

Recommendation 2: JISC SHOULD stimulate cloud-based shared service delivery. Various JISC committees have invested considerable time and resources in developing shared services in a non-cloud context. Hence there are a number of existing shared services in the Grid and VRE space that can be migrated to the cloud for research, and there are new services to be developed, including the need to provide management software for virtual machine images. This will allow Universities to explore the appropriateness of various cloud types and also their 'green' credentials. More detailed recommendations related to environmental issues can be found in the final report[3] of review of the environmental and organisational implications of cloud computing in Higher and Further Education.

Recommendation 3: JISC COULD work with international standard bodies, for instance Open Cloud Computing Interface, a working group in OGF, or similar non-grid organisations. This will require funding and efforts, however, on standards adoption through reference implementations.

Recommendation 4: JISC COULD fund various cloud service facilities, particularly storage and data cloud facilities, to explore data security and privacy as highlighted areas of concern from practitioners and stakeholders. Funding projects in this area will allow researchers to provide alternative methods of sharing research data inside UK research communities, with enhanced security and data privacy.

Recommendation 5: JISC MAY fund large scale institutional and cross institutional clouds for research. This could be informed by the work of the NGS cloud pilot project currently underway, with a view to providing cloud services on the NGS structure.

[3] http://www.jisc.ac.uk/whatwedo/programmes/greeningict/environmentalreviewcloudcomp.aspx

6 CONCLUSIONS

Cloud computing is an emerging business model that delivers computing capabilities over the Internet in a highly flexible and elastic manner. There are many different interpretations of cloud computing. Some think cloud computing is a new computing technology. We clarify in this report the view that cloud computing or the cloud is a business term which does not bring a new technology but instead a new way to design, deliver, and use computing capabilities.

It is hard to give a unified definition for cloud computing, as its meaning varies for different stakeholders. From end users' perspectives, cloud computing provides a highly available and reliable platform enabling service delivery over the Internet with guaranteed QoS. For cloud application developers, it means a new software design and delivery platform. In order to provide a consolidated understanding of cloud computing, we have characterised a cloud computing system as comprising four basic features: elasticity, self-servicing, self-managing, and cost-effectiveness, which applies to all stakeholders.

Cloud computing is a well-established term in the commercial domain, with proven application use cases. It is relatively new to the research domain of HEIs, and its applicability for research usage is still not clear. In this report, we reviewed current practices of early uptakes of cloud computing for research activities and can report that the concept of using cloud computing for research still needs time to be proven. We also observe that technical requirements of these example cases are much more demanding than current cloud offerings can provide. There is an increasing number of projects being funded which carry out experimental work on cloud computing. It is expected these projects will provide further details on the practicability of cloud-based research, and will provide meaningful technical details (e.g. performance benchmarks).

We recommend JISC supports: proofs of concept of using cloud for research; migration of existing shared services in the Grid and VRE space to the research cloud; development of standards through reference implementations; storage clouds and data clouds to develop enhanced security and data privacy; and the building of large-scale research cloud infrastructures.

7 REFERENCES

[1] J. Varia, “Cloud Architectures”, Amazon Web Service white paper, Available online at: http://jineshvaria.s3.amazonaws.com/public/cloudarchitectures-varia.pdf

[2] IBM Software Group, “The value of cloud computing to outsourcers and their clients”, IBM cloud computing white paper, June 2009, Available online at: ftp://ftp.software.ibm.com/software/uk/rational/cloud_computing_white_paper.pdf.

[3] Sun Microsystems, “Introduction to cloud computing architecture white paper”, June 2009, Available online at: http://webobjects.cdw.com/webobjects/media/pdf/Sun_CloudComputing.pdf

[4] D. Chappell, “Introducing the Windows Azure platform”, Dec. 2009. Available online at: http://go.microsoft.com/fwlink/?LinkId=158011

[5] Ubuntu Ltd. “An introduction to cloud computing”, Available online at: http://www.ubuntu.com/system/files/introduction-to-cloud-computing_server_cloud.pdf

[6] VMware, “The VMware vCloud Solution Enterprise-Class Cloud Computing for Any Application, Any Customer”, Available online at: http://www.vmware.com/files/pdf/VMware-VCLOUD-SB-EN.pdf.

[7] Ganter, “Cloud computing confusion leads to the opportunity”. Available online at http://www.gartner.com/DisplayDocument?doc_cd=159034

[8] P. Mell and T. Grance, “The NIST Definition of Cloud Computing”, Jul. 2009. Available online at http://csrc.nist.gov/groups/SNS/cloud-computing/

[9] Export Group, “The Future of Cloud Computing”, Jan. 2010. Available online at http://cordis.europa.eu/fp7/ict/ssai/docs/cloud-report-final.pdf

[10] M. Armbrust, A. fox, R. Griffith, et. al, “Above the Clouds: A Berkeley View of Cloud Computing”, Feb. 2009. Available online at: http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf.

[11] VMWare, http://www.vmware.com/

[12] Kernel Virtual Machine, http://www.linux-kvm.org/

[13] Xen, http://www.xen.org/

[14] J. Brodkin, “Gartner: Seven cloud-computing security risks”, Available online at: http://www.infoworld.com/d/security-central/gartner-seven-cloud-computing-security-risks-853

[15] Hyper-V, Available online at: http://www.microsoft.com/hyper-v-server/

[16] D.Nurmi, R. Wolski, C. Grzegorczyk, et. al, “The Eucalyptus Open-source Cloud-computing System”, *in Proceedings of 9th IEEE International Symposium on Cluster Computing and the Grid*, Shanghai, China.

[17] OpenNebula, Available online at: http://www.opennebula.org.

[18] OCCI-wg, “Open Cloud Computing Interface –Core & Models”, OGF, Jan. 2010. Available online at: http://www.ogf.org/Public_Comment_Docs/Documents/2010-01/occi-core.pdf.

[19] P. Marshall, K. Keahey, and T. Freeman, “Elastic Site: Using Clouds to Elastically Extend Site Resources”, *IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid 2010)*, Melbourne, Australia. May 2010.

[20] I. Foster, “Globus Toolkit Version 4: Software for Service-Oriented Systems”, *IFIP International Conference on Network and Parallel Computing*, Springer-Verlag LNCS 3779, pp 2-13, 2006.

[21] Nimsoft, http://www.nimsoft.com.

[22] RightScale, http://www.rightscale.com.

[23] VPN-Cubed, CohesiveFT, http://www.cohesiveFT.com.

[24] DMTF, “Open Virtualisation Format Specification”, Available online at: http://www.dmtf.org/standards/published_documents/DSP0243_1.0.0.pdf

[25] “Open Cloud Manifesto”, Available online at:http://www.opencloudmanifesto.org/Open%20Cloud%20Manifesto.pdf.

[26] D. Britton et al, ”GridPP: the UK grid for particle physics“, *UK e-Science All Hands Conference*, Phil. Trans. R. Soc. A June 28, 2009 367:2447-2457.

[27] T. Dimitrakos, Common Capabilities for Service Oriented Infrastructures - Grid and Cloud Computing”, Chapter 8 of Grid and Cloud Computing A Business erspective on Technology and Applications, pp. 123-145, Springer Berlin Heidelberg, ISBN: 978-3-642-05192-0, 2010

[28] A. Merzky, K. Stamou, and S. Jha, “Application level interoperability between Clouds and Grids”, *Procs. Of Workshops at the Grid and Pervasive Computing*, 2009, Pg. 143-150.

[29] K. Chine, “Biocep, an e-Science Computational Platform for the Cloud'', *IEEE international conference on eScience*, 2008.

[30] K. Chine, “Scientific Computing Environments in the age of virtualization'', *Proc. of IEEE International Workshop on Open-source software for scientific computation (OSSC 2009)*, 2009.

[31] Arjuna Agility™, “Removing the Barriers to Business Agility white paper”, June 2008, Available online at: http://www.arjuna.com/files/Agility_White_Paper.pdf

[32] D. A. Bacigalupo, J. van Hemert, A. Usmani, D. Dillenberger, G. B. Wills, and S. A. Jarvis, “Resource Management of Enterprise Cloud Systems Using Layered Queuing and Historical Performance Models”, *Proc. of 24th IEEE International Parallel and Distributed Processing Symposium (IEEE Press)*, Atlanta, USA, 19-23 April 2010

8 BIBLIOGRAPHY

Allspaw, T., 2008. *The Art of Capacity Planning: Scaling Web Resource.* O'Reilly.

Velte, A. T., Velte, T. J., V. & Elsenpeter, R., 2009. *Cloud Computing a Practical Approach,* McGraw-Hill Osborne.

Hurwitz, J., Bloor, R., Kaufman, M., & Halper, F., 2009. *Cloud Computing for Dummies,* John Wiley & Sons.

Mather, T., Kumaraswamy, S., & Latif, S., 2009. *Cloud Security and Privacy,* O'Reilly.

Jennings, R., 2009. *Cloud computing with the Windows Azure,* Wrox.

Katarina, S. S, Thomas, W., & Santi, R., 2010. *Grid and Cloud Computing,* Springer.

Katz, R. N. (ed), 2008. *The Tower and the Cloud: Higher Education in the Age of Cloud Computing*, EDUCAUSE.

Appendix A: Example Cases

The following lists a selection of example cases collected via face-to-face interviews, paid consultancy, tele-conferences, and online questionnaires. These example cases provide extensions to those collected from the "Using Cloud for Research" project, in particular those that provide research cloud facilities using open-source Cloudware.

A.1 Cloud Computing for Planetary Defence

Project Description

This project is partially funded by Microsoft Ltd. to utilise its Azure platform in the Cloud analysis of planetary defence and space situational awareness (SSA). Propagating trajectories and calculating conjunctions becomes increasingly time critical, thus requiring an architecture which can scale with demand such as the Cloud.

Current Status

- Use Azure as a cloud provider
- Application developed and debugged using Azure Desktop, which simulates Azure Cloud environment
- Application is programmed based on .NET framework only.
- Using Azure Blob storage for data storage
- A workflow of jobs are queued and scheduled to

Key Concerns/comments

- Resource allocation does not deal with poise message
- No proper monitoring and logging mechanism for fault detection (custom extensions developed)
- No security issue, because data are obtained in public repository

A.2 Cloud-Enabled Neuroscience Research

Project Description

The project is run by the North East Regional e-Science Centre, and the School of Computer Science and the Informatics Research Institute at Newcastle University. It is also partially funded by Microsoft.

Current Status

- Using Amazon EC2, at least three reserved VM instances for database storage, processing, and delegating work tasks and data to application servers at NEeSC;
- 2 application servers at Newcastle scheduling tasks to Azure;
- Tasks are queued and scaled to Azure computing resources.

Key Concerns/comments

- Different patterns of programming, especially for Azure, .Net framework only
- It is hard to migrate existing application to the Cloud directly. Extra development and debugging may be undertaken.
- When coming across more than one Cloud provider, different APIs must be employed on a per Cloud basis.
- Consideration must be given for different paying models as well.

A.3 Cloud-based Monte Carlo Simulation

Project Description

GridPP [26] is a collaboration of particle physicists and computer scientists from the UK and CERN, with distributed compute resources spanning 17 UK institutions. GridPP has a number of key stakeholders – it is the UK's contribution to the worldwide Large Hadron Collider (LHC) Grid (WLCG), overseeing the Tier 1 facility at RAL and the Tier 2 organisations of ScotGrid, NorthGrid, London, and SouthGrid, and also contributes to the interdisciplinary project EGEE - Enabling Grids for E-sciencE.

Current Status

- Planning to address the issues of cloud computing in the context of GridPP and its future development.
- Identified cost of data storage on the cloud and the cost of getting data in/out of the cloud compromises their usefulness.
- Experimental work is collaborated with StACC to run Monte Carlo jobs using private cloud employed by Eucalyptus as a simulation of Amazon EC2 environment.
- Initial pilot survey in association with one or more of the Large LHC experiments by enabling Monte Carlo production system to be submitted to commercial vendors such as the Amazon Cloud EC2;
- Further work will be considered in outsourcing peak demand of experiments to public Cloud in an automatic manner, i.e. integrated in gLite middleware

Key Concerns/Comments

- Current surveys from experiments presenting at the CHEP09 Conference indicate that such a solution is technically viable for some limited applications (e.g. "Belle on the Amazon EC2 Cloud") however the full costs and the overall ability to anticipate and meet peak demand have not been established in the (UK) market. In particular, the economics dictates that great care must be taken to minimise the amount of time data resides on the cloud, yet the vendor control of bandwidth into and out of the cloud creates a form of vendor lock-in that makes the cost unpredictable. Thus, the lack of transparency in the overall vendor pricing/business model is a key concern.
- Concerns also come from how to integrate job submission and authorisation mechanisms of existing WLCG infrastructure to public Cloud infrastructure.
- Interoperability is another issue, legacy solution might be considered by using Cloud as an alternative computing element.

Link

- http://www.gridpp.ac.uk

A.4 Campus cloud using virtualisation

Project Description

The IT services of Cardiff University offer high-end computing facilities using Grid technologies to provide on-demand support for research projects. The Campus Grid facilities at Cardiff consists of up to 1400 student and staff windows workstations as a Condor pool. In order to maximise computing efficiency, the IT service department is planning to adopt virtualisation and Cloud technologies and migrate the existing Condor pool as a private Cloud.

Current Status

- Workstation virtualisation using CoLinux

- Server virtualisation – a central manager VM for test
- More cost-effective than Amazon EC2, approximately 2p per CPU per hour at Condor pool compared to 6p per CPU per hour on Amazon EC2 Cloud
- Key Concerns/Comments
- Would a central manager VM be a right option?
- Connection to NGS, lacking of a gateway VM

A.5 Open-source private cloud for enterprise systems

Description

StACC is a research collaboration, launched in April 2009, focusing on research in the area of cloud computing. Unique in the UK, StACC aims to become an international centre of excellence for research and teaching in cloud computing and will provide advice and information to businesses interested in using cloud-based services.

Current Status

- Ten servers, one for Cloud manager, one for storage, and eight for compute as a private cloud;
- Eucalyptus migrated from version 1.5 to 1.6 recently (many incompatibility issues);

Key Concerns/Comments

- Researches the monitoring and measurement of energy saving;
- Migration of existing applications to the Cloud

A.6 UK academic cloud

Description

JANET is the network dedicated to the needs of education and research in the UK. It connects the UK's education and research organisations to each other, as well as to the rest of the world through links to the global Internet. In addition, JANET includes a separate network that is available to the community for experimental activities in network development.

Current Status

- Planning a centralised virtual data centre and tools to support UK research communities;
- At an early stage and requires guidance on technical barriers and solutions;

Key Concerns/Comments

- Lacking common programming, debugging and profiling models;
- Security is a big concern;
- Virtualisation is not interoperable for heterogeneous resource management

A.7 Private cloud as a development and test platform for OGSA-DAI

Description

The IS department of EPCC provides support for student projects with long-term commitments.

Current Status

- A Cluster of ESX servers provides an on-demand, flexible configuration of a research environment;
- Pilot test using Eucalyptus
- Example application: OGSA-DAI [27] performance test

Key Concerns/Comments

- Still in early stage, not sure technical barriers

A.8 National Community Cloud for Research

National Grid Services (NGS) launched a cloud pilot project, an Oxford-Edinburgh R&D NGS cloud study with Matteo Turilli, Steve Thorn, David Wallom and David Fergusson. They are installing Eucalyptus and looking at how it can be integrated with NGS systems. They are particularly interested in using this for users from biology; and to quickly deploy software for their internal training effort.

Current Status

- In the process of configuring Eucalyptus for both Oxford and Edinburgh NGS nodes;
- Investigating how to use the two together in combination with existing NGS infrastructures and services (e.g. using the NGS Security infrastructure for Eucalyptus)
- VM images
- Hardware: 8 Cores (16 virtual cores) at Edinburgh, 64 cores at oxford.

Key Concerns/Comments

- Security (Authorisation and authentication) - Need agreement on policies e.g. from institutions in the NGS. Need to decide what privileges to give people and how to monitor, in particular how to integrate cloud in existing security infrastructures and auditing services?
- Monitoring and logging - Accounting, for security, resource usage etc, including reporting to stakeholders and day to day monitoring of the system status. How to integrate existing accounting infrastructure? Note: there is an existing NGS accounting system.
- User Support
- Financial Management e.g. is FEC possible?
- Ethics/sensitivity of data/legal issues

A.9 Provisioning cloud services and tools for research

OMII-UK provides and supports free, open-source software to enable a sustained future for the UK e-Research community.

Current Status

- Cloud middleware solutions or support tools for researchers
- SAGA APIs [28]: application-level interoperability for Grid and Cloud;
- Biocep [29][30]: a statistical analysis tools on the Cloud
- SimulationBox: using virtualisation technologies for Cloud-based simulation software deployment

Key Concerns/Comments

- Virtualisation technologies will be used extensively for migrating existing OMII-UK middleware/software to Cloud tools;
- For SAGA project, HPC programming abstraction and interoperability would be of high importance

A.10 Backup of Cancer Data

Description: This example case describes work on the Cancer Cloud Initiative that forms part of a UK National Health Service (NHS) project.

Current Status

- Breast cancer is the most common cancer in women and has a worldwide annual incidence of over 1 million cases
- This work looks at how medical cancer data can be backed up
- Extremely large amounts of storage are required (> 10 Terabyte)
- The work involves integrating software and cloud technologies from commercial vendors including Oracle, VMWare, EMC and HP.
- It is essential that the system is easy to use and that the data can be automatically and reliably backed up. It is also helpful to be able to migrate the data between resources.
- The data is backed up internally using raid storage resources
- The most important data is then backed up to a multi-terabyte data center storage resource
- A second multi-terabyte data centre Raid storage resource is planned

Key Concerns/Comments

- Security of public clouds.
- Disaster recovery and the time this takes (e.g. two weeks). Have considered for proprietary solutions to look after.
- How to move from IaaS to PaaS and SaaS.

A.11 Arjuna Agility

Description

Arjuna [31] Technologies Ltd is a technology consultancy and product development business which specializes in Cloud Computing. Rozmic Wireless, a global provider of email spam and virus protection, has cut its operational costs and increased the efficiency and flexibility of its email cloud service by deploying Arjuna Agility using Amazon EC2.

Current Status

[NOTE: The following are the cloud computing issues flagged by the Arjuna management team as having a high priority, during an interview on 8/2/10. These issues do not refer specifically to the Rozmic case study.]

- Arjuna agility creates an internal cloud of computing resources within the enterprise.
- This allows the dynamic provisioning of IT resources to meet changing business demands.
- Rozmic's software-as-a-service offering, emailcloud, is a fully managed email service that scans and removes unwanted email before it reaches the customer network, saving bandwidth, time and eliminating the irritation of personal spam management.
- Agility dynamically manages Rozmic's infrastructure, creating a single private cloud of in-house resources, from over 50 servers in two data centres.
- Agility allows seamless integration with EC2, to deliver scalable capacity in line with customer demand.
- Rozmic supports the base load using the in-house infrastructure and uses EC2 to accommodate demand spikes.
- Arjuna maintains links with the University of Newcastle, and Arjuna's offices are located on the University of Newcastle campus.

Key Concerns/Comments

- Service level agreement management
- Security
- Standards, protocols and interoperability
- Managing organisational policies (e.g. that restrict where the company can put its data)

Links

- Arjuna Agility: http://www.arjuna.com/agility
- Arjuna Rozmic Wireless case study: http://www.arjuna.com/node/84

A.12 Using Performance Evaluation to Enhance the Scheduling of Scientific Jobs

Current Status

- The University of Warwick's High Performance Systems Group (HPSG) is in the Department of Computer Science and is led by Prof. Stephen Jarvis. One research area is "Business Systems and Cloud Computing". This work [32] is investigating the performance and scalability of multi-tier architectures. It is the group's intention that the research should be both academically interesting and commercially relevant, with much of the recent work focusing upon significant industrial issues such as policies for dynamic server switching, performance prediction for distributed enterprise applications and job allocation in heterogeneous environments.
- For example, being able to rapidly evaluate high quality performance models of a system can be used to help facilitate service level agreement-based cloud scheduling (i.e. workload and resource allocation management). See e.g. recent work involving the University of Southampton, the University of Warwick, the Schools of Informatics and Engineering at the University of Edinburgh, the IBM T.J. Watson Research Centre and the TSB FireGrid project.
- HPSG is working on the analysis and performance evaluation of software/systems combinations. Much of the work has focused on computationally intense applications of interest to the science and engineering community. In a joint undertaking with IBM's T.J. Watson Research Laboratories in New York these techniques have also been applied to distributed enterprise systems.
- Prof. Jarvis comments "this [capacity planning, resource allocation and fault-tolerance mechanisms] is very important for sustainable research. Previous research efforts were trying to focus purely on technical issues and not enough on how to provide a predictable and reliable service. For example, how do you provide a sustainable service at a particular cost?"

Key Concerns/Comments

- Security – as without this people are much less likely to use cloud computing.
- If the service is charged, and hence qualities of service are expected, the management of service level agreements are required.
- Capacity planning, resource allocation and fault-tolerance mechanisms.
- Computing/storage solutions being available for the long-term
- Costs – issues that need to be addressed include: i.) is there a cost? ii.) what is the cost? iii.) how to make the cost predictable?
- Ease of use i.e. "it should just work".
- Virtualisation – as without this, cloud computing is much less likely to succeed.
- Storage, information preservation and back-ups.

Links

- http://www2.warwick.ac.uk/fac/sci/dcs/research/hpsg/
- http://www2.warwick.ac.uk/fac/sci/dcs/research/hpsg/contact/faculty/saj