JISC Final Report

Project Information			
Project Acronym	TeciRes		
Project Title	Technical Review of Using Cloud for Research		
Start Date	16 November 2009	**End Date**	15 May 2010
Lead Institution	University of Southampton		
Project Director	Dr Gary Wills		
Project Manager & contact details	Dr Gary Wills Learning Societies Lab Electronics and Computer Science University of Southampton SO17 1BJ		
Partner Institutions	As consultants: University of Warwick.		
Project Web URL	www.techires.ecs.soton.ac.uk		
Programme Name (and number)	Research Infrastructure		
Programme Manager	James Farnhill		

Document Name			
Document Title	Final Report		
Reporting Period	*Not applicable*		
Author(s) & project role	Dr Gary Wills Lester Gilbert Xiaoyu Chen Dr David Bacigalupo		
Date	7 June 2010	**Filename**	
URL	www.techires.ecs.soton.ac.uk		
Access	☐ Project and JISC internal	☑ General dissemination	

Document History		
Version	**Date**	**Comments**
0.a	27/04/10	Initial draft
1	7/6/10	Final draft

Final Report

Technical Review of Using Cloud for Research (TeciRes)

Gary Wills, Lester Gilbert, Xiaoyu Chen and David Bacigalupo
Learning Societies Lab
Electronics and Computer Science
University of Southampton
SO17 1BJ

Table of Contents

Acknowledgements

The authors acknowledge the support and funding from the JISC's Research Infrastructure programme.

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
Cardiff University	Omer Rana, James Osborne
Curtis+Cartwright Consulting Ltd	Max Hammond, Rob Hawtin
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

Executive Summary

The purpose of this 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 project combined expert consultancy and desk research approaches. The project started with stakeholder analysis, use case gathering 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. Our project team attended research conferences and organised workshops to enrich our 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.

The primary purpose of the TeciRes project was to perform a technical review of using Cloud for research, and produce a report along with two guidance notes; one on cloud computing for research, the other on technical issues on cloud computing for research. These are available on the project website at http://tecires.ecs.soton.ac.uk/ along with other project outputs.

During the 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 academic in this field. The outputs from the project are a reflection of an extensive literature research but more importantly of the issues that practitioners in the field are facing. Where possible we have tried 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 be tried on a small scale.

Recommendations

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.
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 of review of the environmental and organisational implications of cloud computing in Higher and Further Education.
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.
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.
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.

Background

There is a rapidly increasing range of research activities which involve the outsourcing of computing and storage resources to public cloud providers, who provide managed and scalable resources virtualised as a single service. Amazon Elastic Computing Cloud (EC2) and Simple Storage Service (S3) are two widely adopted open cloud solutions, which aim at providing pooled computing and storage services and which charge users according to their weighted resource usage. Early cloud initiators focused on resource provision at infrastructure level, also known as Infrastructure as a Service (IaaS) cloud, and did not allow implementation of advanced features until the emergence of the platform cloud, which offered full software stacks and enabled multiple applications to share a single development and hosting environment. Platform cloud providers such as Google Application Engine and Microsoft Azure service normally provide a dedicated Software Development Kit (SDK) and hosting environment, making it hard to migrate applications developed on one cloud platform to the other, although attempts have been undertaken to extend generic programming models utilising cloud capabilities. Rather than using third-party cloud services, researchers in the UK have shown great interest in developing custom cloud solutions from existing IT infrastructure.

One of the first UK private clouds was set up by the St. Andrews Cloud Computing Initiative (StACC) in summer 2009, based upon 'Elastic Utility Computing Architecture Linking Your Programs To Useful Systems' (EUCALYPTUS), an open source software that enables the creation of private computing clouds on premises and hybrid clouds allowing these on-premises clouds to interact with public clouds through Amazon Web Service (AWS) interfaces. In order to ensure the interoperability of development tools and applications across cloud infrastructures, the Open Cloud Computing Interface (OCCI) working group was founded in April 2009 by Open Grid Forum (OGF) to deliver standard APIs for management of cloud computing infrastructure.

Despite all this activity on cloud computing the term itself and the technologies that under-pin it are still confusing to many. For instance, whether cloud computing is the same as grid computing or high performance computing; we say it is not and we also explain why it is not. In this project we provide some guidance on what the literature is calling cloud computing. We have also interviewed staff in HE institutions in England, Scotland, and Wales as to the technical issues they found in providing cloud resources and cloud computing.

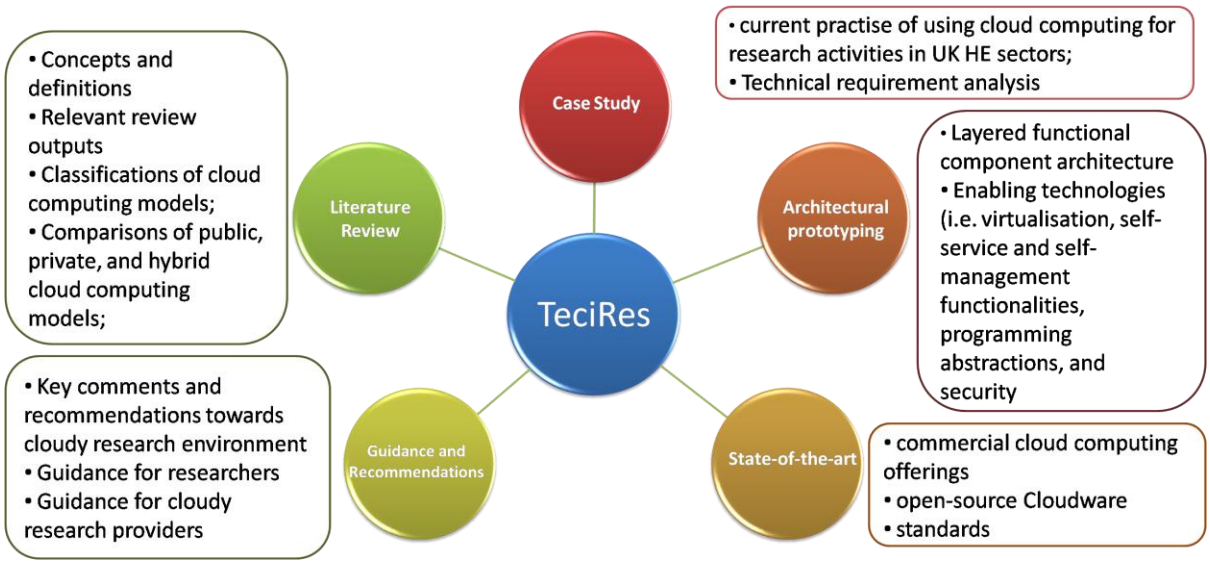

Figure 1: Project Aims and Objectives

Aims and Objectives

The purpose of this 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.

In order to achieve these aims, a sequence of target outcomes was classified into five categories as illustrated in Figure 1 and detailed as follows:

Aim One: Literature review

- Identify the concept and essences of cloud computing from a technical point of view
- Classify clouds and their technical aspects
- Compare the advantages and shortfalls of using public clouds, developing private clouds, and employing hybrid clouds

Aim Two: Case study and requirement analysis

- Identify and interview primary stakeholders
- Construct representative example case scenarios of using clouds for research in UK HE sectors
- Categorise the example case scenarios
- Ascertain technical requirements per category and per example case

Aim Three: Architecture prototype

- Construct an architectural view of and identify the key capabilities for IaaS, PaaS, and SaaS cloud models
- Identify enabling technologies for self-service provisioning and management, self-management and automated management (monitoring, accounting, billing, and reporting), security, and programming abstractions
- Review intermediate solutions being employed in HE and state-of-the-art cloud technologies from professional, commercial, and industrial domains to fulfil these key capabilities, with emphasis on service level management, virtualisation technologies, monitoring, logging and accounting, and interoperability
- Compare open-source and commercial solutions for cloud systems and evaluate their applicability for cloud-based research

Aim Four: State-of-the-art

- Identify and analyse current commercial cloud strategies and offerings
- Identify and analyse open-source Cloudware for building private clouds
- Identify and analyse standards and interoperability
- Review capacity planning and resource allocation mechanisms

Aim Five: Guidance and recommendations

- Suggest technical guidance and best practice for cloud providers who provide, or intend to provide, cloud facilities for UK HE, with rules of thumb and best practice
- Suggest guidance and best practice to researchers who are interested in using cloud system for scientific research
- Suggest recommendations to JISC for cloud research environments

Methodology

The project combined expert consultancy and desk research approaches. The project started with a stakeholder analysis, example case scenario gathering from HE institutions and research departments, and close collaboration with the “using cloud for research” group, while paying particular attention to technical requirements and issues. Our project team attended research conferences and

organised workshops to enrich our 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. Paid expert consultancy was commissioned to take a detailed look at two Russell group universities (the University of Warwick, the University of Southampton) to identify technical requirements and issues. This included input from researchers, information service and research support staff, and heads of department with academic and research responsibilities. Finally we provided guidance and best practice reports on the adoption of existing cloud technologies for future HE researches, as well as recommendations to JISC for open technical issues and standards tasks where applicable.

Implementation

The literature survey provided the known issues with cloud computing, and we specifically focused on the issues involved in supporting research. Existing literature was surveyed mainly through searching the major research databases, for instance IEEE Xplorer, SpringerLink, ACM digital library, and Google scholar. Interviewees also recommended specific publications and conferences to survey. Coupled with this the team attended a number of national and international events around cloud related issues.

Of note are the EU report on the future of cloud computing [1] and the US Cloud computer use case white paper [2].

A stakeholder review was conducted through visits, teleconferences, email, and online questionnaires. Feedback was received from over twenty people, from various stakeholder groups including individual researchers, HE information service departments, and HE research computing service providers. We were careful to include the research communities in England, Scotland, and Wales.

The issues raised in the literature review were the starting points for the questions asked at the interview stage. The interviews were grouped to represent the different roles of those responsible for providing the technical solutions. There were

- HE general information service departments
- HE research computing services
- Early adopter researchers
- Commercial vendors

In addition to the interviews and invited speakers from universities and commercial organisations, we also attended Workshops and conferences:

- All Hands Meeting 2009, UK
- EU Future of Cloud Computing, Brussels January 2010
- Cloud Camps: January and March 2010
- JISC workshop March
- OGF 28 March 2010, where we also ran two workshop sessions
- Cloud Computing - Sharing the Experience, 13 May, 2010 e-Science Institute, Edinburgh, where we provided a presentation
- IEEE International Parallel & Distributed Processing Symposium, where we provided a presentation
- IBM T.J. Watson Research Centre, where we provided a presentation

Outputs and Results

The primary purpose of the TeciRes project was to perform a technical review of using cloud for research, and produce a report which included the following outputs:

- A review of the concept, taxonomy, capabilities, and technical requirements of infrastructure cloud for research activities

- A review of stakeholders and example case scenarios from individual researchers, HE Information Service providers, and Resource Computing Service providers
- An Architectural prototype, and comments on the key issues to be addressed
- An assessment of state-of-the-art solutions from both commercial and open-source providers which would address identified issues
- Guidance to researchers and cloud infrastructure providers including both HE Information Services and Research Computing Services
- Recommendations on further work to JISC on open issues

These are available on the project website at http://tecires.ecs.soton.ac.uk/

Outcomes

The project targeted the following outcomes:

- Understanding technical requirements and how solutions can be used in infrastructure clouds for research
- Guidance on use, adoption, and migration to cloud computing by applying existing cloud technologies for researches in UK HE
- Recommendations to JISC for future technical work on the use of cloud computing in UK HE

These impact a number of stakeholders in the following way:

- JISC: Funder/recipient of recommendations for future work
- CETIS: Guidance to influence advice provided by CETIS
- Janet UK: Guidance to influence Janet UK investigation into providing a UK HE cloud
- Researchers: Guidance to help researchers to carry out high performance and cost-effective research by using clouds
- HE Information System/research computing services: Guidance to facilitate provision of cloud facilities for researchers
- Commercial cloud vendors: Guidance to help them effectively target the HE market

Conclusions

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

- **Cloud computing for research is in its infancy**

  Cloud computing for research is still in a very early stage of development. Its concepts, characteristics, underlying technologies, and applicability for research are not clear to many. Although early cloud adopters have shared some useful information, including example case scenarios demonstrated in this report, in the “Using Cloud for Research” project, and in those published by commercial public CSPs, most of this information is constrained by the interests involved and only show that cloud computing can run specific research applications. We need more information (e.g. performance benchmarks, economic savings, etc) before proceeding to define a global strategy for using cloud computing in research.

- **Commercial cloud or “research cloud”, there is no right answer yet**

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

- **Current offerings are not research friendly**

  The current cloud computing offerings comes 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 applications. Although public CSPs provide powerful self-management facilities and well-

defined programming APIs, these enhanced features also "lock" applications in to a specific cloud infrastructure. On the other side, researchers are starting to think of building specific cloud environments by using open-source Cloudware. These open-source offerings only provide low-level programming APIs and service interfaces, however, leaving resource management tasks to application developers.

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

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

Implications

During the 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 academic, in this field. The outputs from the project are a reflection of an extensive literature research but more importantly of the issues that practitioners in the field are facing. We have shown some ways of overcoming these problems, yet there are still many questions unanswered, in part because cloud computing for research is still only a concept to many and because it has only been tried on a small scale.

Figure 2: Roadmap towards the vision of cloudy research environments

Recommendations

According to the roadmap, we list five prioritised recommendations to JISC to stimulate actions leading towards UK cloud research environments. We use the following terms to distinguish three different levels of priorities.

- "SHOULD" implies top priority and need to act immediately
- "COULD" implies medium priority that is to be planned later
- "MAY" implies lowest priority

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.
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 of review of the environmental and organisational implications of cloud computing in Higher and Further Education.
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.
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.
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.

References

[1] Export Group, "The Future of Cloud Computing", Jan. 2010. Online available at: http://cordis.europa.eu/fp7/ict/ssai/docs/cloud-report-final.pdf

[2] Open Cloud Discussion Group, "Cloud Computing Use Case White Paper (version 3.0)", Feb. 2010 Online available at: http://opencloudmanifesto.org/Cloud_Computing_Use_Cases_Whitepaper-3_0.pdf