

Best practice for funding the use of cloud in research

Neil Chue Hong¹, Mike Jackson¹, Jeremy Cohen²

¹The Software Sustainability Institute, EPCC, The University of Edinburgh, James Clerk Maxwell Building, Mayfield Road, Edinburgh, EH9 3JZ.
N.ChueHong@software.ac.uk, M.Jackson@software.ac.uk

²London e-Science Centre, Imperial College London, South Kensington Campus, London, SW7 2AZ.
jeremy.cohen@imperial.ac.uk

Contents

Contents	2
1 Introduction	2
1.1 Acknowledgements	3
2 What is a cloud?	3
2.1 Cloud types	4
2.2 Potential benefits of using clouds	5
3 Best practice for assessing project proposals within a programme	5
4 Best practice for managing provisioning of shared cloud computing infrastructure .	7
5 Recommendations to address current issues in the use of cloud computing infrastructure in research projects	8

1 Introduction

Cloud computing, with its potential to deliver computational, data and storage resources on-demand, is currently an area of great interest across academia, industry and commerce. It has prompted much debate as to its benefits and its costs, its strengths and its weaknesses. However, despite the hype, cloud is not a panacea that offers a guaranteed improvement over clusters, grids, or HPC infrastructures. Rather, it provides an alternative to or supplements these platforms, offering another means by which researchers can provision computational, data and storage resources for use in their research. For some researchers, a cloud might be ideal, for others an existing grid or a new cluster may be more appropriate while for a further group, a combination of these may be best. An understanding of the potential benefits of cloud computing and the issues arising from its use can contribute towards a more effective exploitation of cloud computing within research, particularly as the community seeks to clarify its strategic plan for the UK Research Computing Ecosystem.

This document provides best practice, guidance and recommendations to funders of research projects that potentially involve the use of cloud computing. It may also be of interest to senior managers and research directors. It is intended to help answer a range of questions, such as:

- If a researcher or institution requests money for cloud computing resources, what issues should I bear in mind as a programme manager?
- If a researcher or institution asks for funding for hardware, how should I assess this request to provision new infrastructure?
- What do I need to consider if investigating the provisioning of a shared cloud computing infrastructure that can be made available to researchers across a range of institutions?

This document is an outcome of the RAPPORT (Robust APplication PORTing for HPC in the cloud) project¹ which was funded under the JISC/EPSCRC Cloud Computing in Research programme². The

¹ <http://cloudresearch.jiscinvolve.org/wp/category/projects/rapport/>

project investigated the experiences of researchers in the areas of high-energy physics, optical character recognition in the digital humanities and bioinformatics in porting their applications to clouds. Best practices for both funders and researchers were developed from this work.

1.1 Acknowledgements

The authors acknowledge the support of EPSRC and JISC as part of the Cloud Computing in Research programme. The authors also thank the members of the RAPPORT project, based at Imperial College London for their input: John Darlington and Brian Fuchs of the London e-Science Centre; David Colling and Daniela Bauer of the High Energy Physics Group; Sarah Butcher, Mark R Woodbridge and Ioannis Filippis of the Bioinformatics Support Service and Matt J Harvey of the Information and Communications Technologies (ICT) department.

We also acknowledge the work of a previous study commissioned by JISC: Using Cloud Computing for Research³ in providing the framework for assessing drivers and barriers to the uptake of cloud computing in research.

2 What is a cloud?

There have been many definitions of cloud computing put forward. Cloud covers a wide range of areas of software and hardware which can be broadly categorised into the three-layer cloud stack:

- Infrastructure-as-a-Service (IaaS): Delivery of computational, data and storage infrastructure through some form of service interface. Users can generally access resources on-demand, paying only for what they use. Resources are normally provided using virtualised hardware and the user is responsible for managing the resources and loading their software onto them.
- Platform-as-a-Service (PaaS): Users access cloud resources through some higher-level platform interface. PaaS solutions provide services that simplify tasks that a user would have to handle themselves using IaaS, but often at the expense of the great flexibility of IaaS. They may build on top of an existing IaaS offering or use underlying resources that are accessible only via the PaaS interface. PaaS covers a wide range of service types from general value-added services on top of infrastructure clouds to highly advanced, domain-specific platforms.
- Software-as-a-Service (SaaS): The SaaS layer covers service-enabled remotely-accessible software packages. This provides an even more specialised offering than PaaS but where the software provides features suited to a user's requirements, it removes from users the complexities of accessing and managing their own resources and deploying and managing their software on these resources. The underlying infrastructure is hidden from the end-user and the SaaS provider manages capacity and scaling to ensure they can handle the number of service users. Examples of SaaS include web-based e-mail applications and a wide range of distributed software applications that use a client on a user's system to connect to a remote software service.

In this guide we focus on infrastructure clouds and we consider these clouds to have the following characteristics:

- Resources are provided to users "on-demand" and there is an expectation from users that the requested resources will be made available within a short time period (e.g. within a few minutes).

² <http://cloudresearch.jiscinvolve.org/>

³ <http://www.jisc.ac.uk/whatwedo/programmes/researchinfrastructure/usingcloudcomp.aspx>

- Supporting scalability is a fundamental goal of the infrastructure.
- Users have full control over their resources and are responsible for their management, starting them up when they need them, shutting them down when they are no longer needed and managing the deployment and operation of their software on these resources.

2.1 Cloud types

There are three main types of infrastructure cloud platform available, public, community and private.

- Public clouds are provided by commercial organisations who generally offer a range of cloud related services and resources and who bill users for their usage. Examples of public clouds are Amazon EC2⁴, Google Apps for Business⁵, FlexiScale⁶ and Rackspace⁷. Though they're termed public, this is in the sense that anyone can use them, it does not mean that any software and data you deploy is publicly-accessible. On public clouds, you retain full control over who can access any software or data you deploy. A major benefit of these platforms is that resources are usually available on-demand, with a short time required between a resource being requested and being made available to the user. Charging is also often relatively fine-grained (e.g. per hour) when compared to traditional remotely hosted servers that may have a minimum charging period of a month or more.
- Community clouds are created for specific communities. These could be organisations who have entered into strategic partnerships or who serve particular communities. Community clouds can be more expensive than public clouds to provision, even where existing hardware is to be used (since there is an initial set-up overhead), but may offer services tailored to the requirements of the community e.g. in terms of privacy or compliance with local or regional policies. Community clouds can be deployed using the same frameworks as private clouds but may require additional higher-level services to integrate users across multiple sites. Examples of community clouds are the UK National Grid Service's Cloud Prototype Service⁸ and NASA's Nebula⁹.
- Private clouds are set up within a specific private community. This may be a single institution or a group of individuals in a collaboration across multiple locations. This gives the institution or "cloud owner" complete control over the cloud deployment. Private clouds are often preferred for mission critical systems within companies or for cases when dealing with applications or data that cannot be deployed on third-party clouds (e.g. medical applications with sensitive data). Private clouds may make use of any standard computing hardware that is installed with suitable middleware software and this may include the re-tasking of existing hardware. Private clouds can be constructed using frameworks like Eucalyptus¹⁰, Nimbus¹¹, OpenNebula¹² or OpenStack¹³.

⁴ <http://aws.amazon.com/ec2/>

⁵ <http://www.google.com/apps/intl/en/business/cloud.html>

⁶ <http://www.flexiant.com/products/flexiscale/>

⁷ <http://www.rackspace.co.uk/rackspace-home/>

⁸ <http://www.ngs.ac.uk/accessing-the-ngs-cloud-service>

⁹ <http://nebula.nasa.gov/>

¹⁰ <http://open.eucalyptus.com/>

Combinations of the above classes are also possible. For example, a hybrid cloud may be formed by combining an internal private cloud with a public or community cloud platform. When the internal private cloud reaches a high load, the external cloud may be used to handle overflow in demand on the private cloud (“cloud-bursting”), or may be used for archiving data.

2.2 Potential benefits of using clouds

In many respects, the benefits to research from the use of cloud computing, and the approaches to analysing these benefits, are similar to those from other types of distributed computing e.g. clusters or grids. These benefits could include:

- Increasing innovation and productivity;
- Reducing inefficiency and duplication;
- Increasing collaboration, including cross-disciplinary collaboration;
- Helping HEIs to provide resources to researchers.

3 Best practice for assessing project proposals within a programme

In the guide, “Best practice for using cloud in research”, the RAPPORT project sets out best practice for researchers seeking to use cloud computing infrastructure. The best practice that we set out for funders builds on this to enumerate the areas in which the proposers of a research project should have set out their motivations and justifications for their intended use of cloud infrastructure. It also addresses aspects of project resourcing more generally.

Understanding motivation and requirements

- Do the proposers understand their motivations for the use of cloud computing, in particular with respect to improving research quality, improving research quantity, improving cost-effectiveness, or improving environmental impact?
- Do the proposers understand their application requirements, how the application utilises resources, and how to configure the application to optimise the use of these resources?
- Has the proposal described its resource requirements in such a way as to identify and justify the potential benefits of cloud computing? Does it indicate the limitations of the target cloud infrastructure? Is there justification for why these limitations would not prove detrimental to the proposed project?

Justifying choices

- If the desire to port to cloud is motivated by performance, are the proposers clear about how the expected performance improvements will be achieved? Is a significant performance increase expected?

¹¹ <http://www.nimbusproject.org/>

¹² <http://www.opennebula.org/>

¹³ <http://www.openstack.org/>

- If the proposed project has a short duration, is the use of existing infrastructure requested to reduce lead times on resource availability? Could cloud be a suitable alternative?
- If a proposal requests dedicated hardware resources, is there justification for this? Would cloud computing or use of other shared resources be more cost-effective? This is particularly pertinent for short duration projects and those where resource utilisation will be sporadic.
- If existing suitable infrastructure is available, do the proposers seek to use it?
- If requests are made for commercial cloud usage, has the expected resource usage (including on-going storage and data transfer costs) been considered, and the costs justified? Does the proposal explain how potential changes in pricing within the project's lifetime would be addressed?

Planning and management

- As part of its risk analysis, does the proposal address the risks relating to the use of cloud infrastructure, particularly dependability, longevity, the challenges that can arise in porting applications to a cloud environment, and, for commercial clouds, changes in pricing? Is there a contingency plan that outlines alternative options the project would explore in case of difficulties?
- Has the proposal addressed ethical and confidentiality concerns relating to transfer and deployment of software and data on third-party infrastructures, if relevant?
- Has the proposal addressed licensing restrictions relating to any software to be deployed? In the case of licensing it is particularly important to consider that even within a local private cloud infrastructure, every time a node is started, it will effectively be a different machine (e.g. with different MAC/IP addresses), even though the underlying physical hardware is the same.
- Has the proposal made provision for data transfer out of the cloud or on-going storage and usage at the end of the project's funding period?

Promoting reuse and sustainability

- Do the proposers indicate how they will feed back and disseminate lessons learned from porting to and use of cloud infrastructure and related technologies?
- If the development of cloud-specific tools is proposed as part of the project, will these conform to existing cloud standards, where available or under development, whether by OASIS¹⁴, OGF¹⁵ or others?
- Does the proposal address the issues of reproducible research, in particular the ability to provide long term access to services and data, and to aid peer review and reuse through encapsulated research environments?
- Does the proposal address how end users will use their software and/or access data once deployed, to promote update and usage? Does this seem appropriate? Do they include any provision for assessing its appropriateness and usability prior to the end of their project?

¹⁴ <http://www.oasis-open.org/>

¹⁵ <http://www.ogf.org/>

- Is it sensible for the project to make “virtual appliances” (virtual machine images with pre-configured software stacks) developed for the project more widely available?

It is also advised that projects consult with bodies such as the National Grid Service¹⁶, JISC Legal¹⁷, OSS-Watch¹⁸ and the Software Sustainability Institute¹⁹, both in preparing proposals that address the above and for on-going advice when their projects are underway.

4 Best practice for managing provisioning of shared cloud computing infrastructure

In addition to the individual choices that a project may make in determining the best infrastructure within which to run applications, there is also the potential for funders to provision shared computing infrastructure. This may be done for reasons of economic efficiency, to reduce environmental impact and to promote, or enforce, openness of research.

Here best practice helps to both support decision making, and ensure long term sustainability and reuse of resources. While for individual projects the use of a shared cloud computing infrastructure may incur a small degradation in performance or loss of individual resource ownership, such concerns can be addressed by promoting the improvements in support, sustainability, reuse, economic efficiency, and reduced environmental impact that can potentially be realised for a group of projects across a programme or community.

Addressing requirements

- Any provision should assess the specialist requirements coming from the academic research community. These may include the need for more stringent networking requirements, higher memory capacity, higher storage capacity, faster CPU speeds or specialist hardware such as GPUs.
- Where compute-intensive jobs are likely to represent a significant amount of the usage of a shared cloud infrastructure, virtual machines will need dedicated CPUs and there will be less scope for sharing hardware between virtual machines. Such infrastructures will require more physical hardware and result in higher costs to service a given number of users.
- Provision community clouds which provide researchers with dedicated cloud infrastructure that meets their requirements and which they can use for free or at a non-prohibitive price. Such clouds could support long running programmes and also correct current misconceptions that locally-hosted resources are the only option for compute or data intensive applications.

Cost and accounting models

- If a general capacity computing resource is provided it should be priced at a level which gives an incentive over purchasing, deploying, supporting and using a dedicated resource.
- Where an existing commercial service provides equivalent terms of service to an existing or requested private resource, outsourcing should be considered. This may be done through an

¹⁶ <http://www.ngs.ac.uk/>

¹⁷ <http://www.jisclegal.ac.uk/>

¹⁸ <http://www.oss-watch.ac.uk/>

¹⁹ <http://www.software.ac.uk>

agreement with a commercial provider and a more rigid set of SLAs on behalf of the research community (seeking to deliver a community cloud that is based on a public cloud).

- Proposals should be allowed to treat cloud computing usage as an eligible consumable cost.
- Be aware that some cloud providers may provide free access to their resources, up to certain limits, or offer discounts or grants to individuals and groups involved in education or research. For example, Amazon's AWS in Education programme²⁰.

Encouraging reuse

- Undertake short programmes of follow-on funding to allow successful cloud porting projects to make their outputs available for use by others.
- Encourage HEIs to re-task resources from completed projects as clouds (if appropriate) through provision of some additional specific funding for both the effort involved and the training of system administrators in the skills and expertise to deploy and manage a cloud.
- Discourage purchase of dedicated hardware for one-off projects unless it will be incorporated into existing shared infrastructure, either as part of the project, or once the project has finished.
- Provide a searchable, federated set of documentation on the use of the cloud infrastructure and encourage researchers and system administrators to contribute to its evolution and improvement.

5 Recommendations to address current issues in the use of cloud computing infrastructure in research projects

This section suggests recommendations to the UK Research Councils and JISC which address some of the standing issues present in trying to use cloud computing within research projects that have been identified as part of the RAPPORT project.

Lobby commercial cloud providers to offer alternative models for managing usage such as invoicing options and quotas

- Commercial clouds typically require the registration of a credit card which is then charged when resources are consumed. This can deter both individual researchers or their HEIs from using commercial clouds, since there is a risk of inadvertently running up large charges (e.g. if forgetting to shutdown running services). Researchers and HEIs should be consulted as to the most effective and acceptable ways, to them and funders, of paying for commercial cloud usage.
- Liaise with other funders, both within the UK and internationally, and discuss with public cloud providers some form of academic charging and costing model. This is both in terms of providing services (e.g. data transfer or storage) at a cost that is not prohibitive to researchers and also ways of paying for these services that are more acceptable to researchers and HEIs (e.g. receiving warnings when charges exceed a certain threshold or halting resource consumption when this threshold is exceeded).

Identify or establish bodies to provide legal advice

²⁰ <http://aws.amazon.com/education/>

- Set up an advisory body, or extend the remit of existing advisory bodies, e.g. JISC Legal, to provide advice on the implications for UK researchers of contracts and SLAs of public clouds in the US, which are defined in terms of US law.

Encouraging reuse

- Carry out a consultation exercise with HEIs, in particular system administrators, about the issues around the provisioning, use and support of private clouds within their institutions and how best to manage this.
- Set up workshops for researchers and system administrators who have experience in setting up and managing clouds to come together and share and exchange information with researchers and system administrators interested in this.

DRAFT