

Best practice for using 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	3
1.1 Acknowledgements	4
2 What is a cloud?	4
2.1 Cloud types.....	5
3 What are the potential benefits of clouds for researchers?.....	6
3.1 Improve the quality of your research	6
3.2 Improve the quantity of your research.....	7
3.3 Improve your cost-effectiveness	7
3.4 Reduce your environmental impact.....	7
4 How can I assess whether I should use a cloud?	8
4.1 Understand your application.....	8
4.2 Assess the dependability of the cloud	9
4.3 Consider legal and ethical issues	10
4.4 Consider financial issues.....	10
5 How do I port my application to my chosen cloud?	11
5.1 Consider how you'll handle unforeseen but inevitable delays	11
5.2 Port your application	12
5.2.1 Exploit the knowledge of others.....	12
5.2.2 Don't try to do everything at once.....	13
5.2.3 Write down what you try and share your experiences as you go	13
5.2.4 Let suppliers know of bugs.....	13
5.2.5 Keep an eye on your resource usage.....	14
5.2.6 Assess the stability and maintainability of other software	14
5.2.7 Think of your users	14
5.3 Share your outputs and experiences	14
6 Migrating between cloud providers	15

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. This guide is intended to help you understand what cloud computing is, the benefits it may be able to offer you as a researcher and the different options available for gaining access to cloud computing resources. The guide is intended to help you make informed decisions about whether or not use of a cloud computing platform could contribute to your research and associated costs and benefits of its use.

We briefly describe what a cloud is, list the types of cloud available, and summarise the potential benefits of clouds for researchers. We then address two fundamental questions which researchers may raise when considering the cloud:

- How can I assess whether I should use a cloud?
- How do I port my application to my chosen cloud?

We do not attempt to answer these questions for you in this guide, as every researcher differs in terms of their requirements, applications and funding. Likewise, clouds are constantly evolving in terms of the range of cloud providers available, the computational, memory and storage resources they offer, how these are priced, and the tools and technologies available to build and deploy applications on a cloud. Rather, we outline questions that you need to ask yourself and provide advice, suggestions and hints to allow you to answer these questions based on your specific research domain and requirements. Some of these questions are specific to cloud computing and some relate to good practices to follow during any software development project. We also consider provisioning of resources, irrespective of whether these resources are made available through a cloud, grid, HPC facility or cluster.

Figure 1 summarises the key activities you should undertake when considering use of a cloud within your research and when porting applications to a cloud. These activities help you answer the two questions above and so understand the technical, legal and economic challenges you may face and devise means of overcoming these.

It is hoped that this guide will assist you when applying for funding by helping you to justify any requests for use of cloud resources, or in considering whether to use cloud on projects that you may be currently involved in.

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

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

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

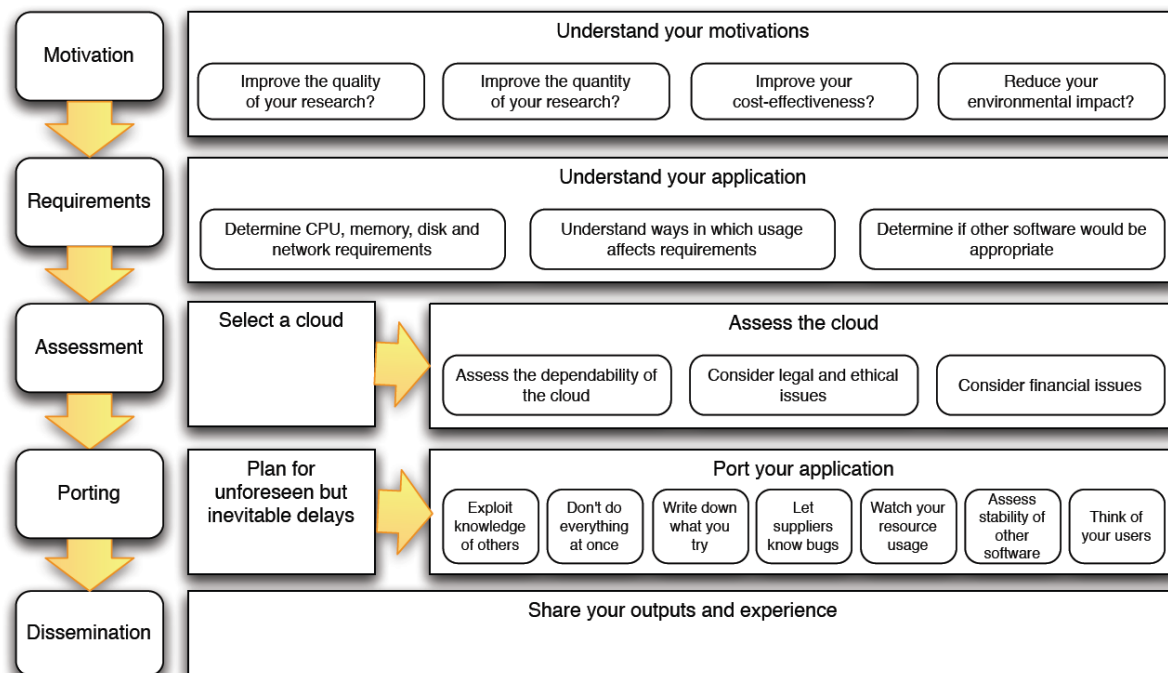


Figure 1: Best practice for researchers in porting applications to the cloud

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. Thanks also to Alastair Hume of EPCC, The University of Edinburgh for reviewing an early draft.

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).
- 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 EC2³, Google Apps for Business⁴, FlexiScale⁵ and Rackspace⁶. 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 Service⁷ and

³ <http://aws.amazon.com/ec2/>

⁴ <http://www.google.com/apps/intl/en/business/cloud.html>

⁵ <http://www.flexiant.com/products/flexiscale/>

⁶ <http://www.rackspace.co.uk/rackspace-home/>

⁷ <http://www.ngs.ac.uk/accessing-the-ngs-cloud-service>

NASA's Nebula⁸. Your funders may be able to suggest suitable community clouds that could support your work.

- 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. For example, the RAPPORT project built a project cloud from machines that were formerly part of a now decommissioned cluster. Private clouds can be constructed using frameworks like Eucalyptus⁹, Nimbus¹⁰, OpenNebula¹¹ or OpenStack¹². Ask around, your colleagues or institution may be operating a private cloud already. Or, your systems administrators may be prepared to set one up for you (especially if you have an old, unused cluster lying around).

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 (“cloudbursting”), or may be used for archiving data.

3 What are the potential benefits of clouds for researchers?

Cloud has many potential benefits for researchers. We say “potential” because whether or not these benefits arise in practice can be dependent upon a myriad of variables including the cloud platform adopted, the types of resources used, the nature of the application, and the time, money and effort available. In addition, each of the benefits may require trade-offs, for example you might gain the ability to run a far larger number of jobs in parallel, but at the expense of a small increase in execution time per job due to different resource specifications; you may gain access to increased computational power but at the cost of no longer owning the underlying hardware; or, you no longer need to purchase and deploy your own cluster, reducing your capital expenditure, but increasing your operational expenditure in the long term. Understanding the trade-offs you need to make is a key aspect of selecting any computing infrastructure, whether it be cloud, grid, HPC facility or cluster. The following sections discuss in more detail some of the potential benefits to researchers of using cloud computing.

3.1 Improve the quality of your research

A cloud could provide you with faster, more advanced or more scalable resources to enable you to run tasks that you cannot manage with your existing resources whether these be within your department,

⁸ <http://nebula.nasa.gov/>

⁹ <http://open.eucalyptus.com/>

¹⁰ <http://www.nimbusproject.org/>

¹¹ <http://www.opennebula.org/>

¹² <http://www.openstack.org/>

institution or the wider academic community. Similarly, cloud resources may allow you to derive more accurate results than you can achieve at present.

A cloud can provide the means to deliver your software and data as a service or to sustain it beyond the current lifetime of your project. Making your software and data accessible to other researchers, either through a service, or directly, allows them to both validate your research and allow it to contribute to theirs. Being able to demonstrate impact in this way may enhance future funding bids you make.

3.2 Improve the quantity of your research

Clouds can improve the quantity of your research, that is, the amount achievable within a given time. This can be enabled by access to more powerful resources that allow you to run your analyses faster. Likewise, a cloud can provide the means to augment your existing resources to handle peaks in demand and reduce contention for local resources e.g. when a group of you are working towards a paper deadline.

A cloud can provide you with resources over which you have greater control. You may have the freedom to install software and use it as soon as you need it, rather than having to rely on system administrators at your institution to do this for you. You can provision resources quickly, rather than waiting for hardware to be ordered, delivered and installed at your institution. You can quickly run tests over a variety of platforms and on virtual machines with different specifications, which can aid rapid prototyping, inter-operability testing and performance testing. Being able to provision resources quickly could be especially valuable for short-duration projects.

A cloud can enable you to deliver your software and data as a service or to sustain it beyond the current lifetime of your project. This can save other researchers time, allowing them to reuse your outputs, rather than having to reinvent them. Again, demonstrating such impact may enhance your future funding bids.

3.3 Improve your cost-effectiveness

Cloud has the potential to improve your cost-effectiveness in a number of ways, allowing you to make the most of your funding. It may be cheaper to use a cloud than to order, install and support new hardware within your institution. Within your institution, unused resources, for example old clusters, could be retasked as clouds making them more flexible and easier to access and removing the need to purchase new hardware where ultimate CPU performance is not critical.

If you can demonstrate that a cloud can reduce your capital expenditure allowing you to dedicate more of your funding to research and less to hardware purchase, installation and support, then this may make your funding bids more attractive to funders.

Delivering long-term access to your services and data resources, by promoting reuse of these resources, you may contribute to the cost-effectiveness of other researchers, saving them from having to expend effort repeating work you have already done.

3.4 Reduce your environmental impact

Cloud can offer the potential for you to reduce your environmental impact. Physical resources in a cloud are generally virtualised and can host a number of independent virtual hardware instances that can be used by a number of different projects, or institutions to optimise available CPU capacity. If these projects or institutions each had separate physical hardware resources, they might sit idle for a significant amount of the time, consuming power and space and requiring cooling.

Public cloud platforms, where resources are likely to be housed in an environment designed for efficient energy consumption and owners can amortise usage across a large number of users with different usage requirements, may offer particular benefits in reducing environmental impact when compared to locally hosted resources.

Retasking existing hardware, in effect recycling it, can also contribute to reducing environmental impact although the lower efficiency of old hardware when compared to more modern resources may mean this is limited.

4 How can I assess whether I should use a cloud?

You may have a particular cloud platform in mind, or have been recommended to use a cloud by your funders. The next step is to determine whether this cloud is suitable for you. Not only should you consider the technical aspects of using the cloud but also issues of the cloud's dependability, legal and ethical issues relating to deploying your software or data and financial issues.

4.1 Understand your application

- Do you understand the CPU, memory, disk and network requirements of your application?
- Do you know the specifications of the CPU, memory, disk and network resources offered by the cloud?
- Do you know how to configure and use your application such that its requirements can be optimised with respect to the available resources? e.g. using it in parallel or partitioning its inputs in a different way?
- Will the cloud support your application's requirements?
- If the cloud doesn't support your applications, are there any other applications that do the same tasks as your current application but which could be supported by a cloud platform?

If we want to deploy and run an application we need to understand whether our computer has sufficient resources, for example, CPU, memory and disk space, to handle the application. A cloud is no different, so in determining the suitability of a cloud you need to address the question "is this cloud suitable for my application?" Of course, you can't actually answer that definitively until you've moved your application into a cloud environment but there are a number of activities you can do to at least determine if it's theoretically suitable.

A cloud will offer a set of virtual computational and data resources. The nature of these virtual resources is constrained by the underlying physical hardware used to implement the cloud. You should determine whether the virtual resources offered by the cloud you have access to can run your application(s). You can find out this information from the cloud provider.

You also need to understand the demands, in terms of CPU, memory, storage, and network bandwidth of your application. You may already know this. If your application was written by a third-party, then they may have this information. Otherwise, you could run some performance tests to identify your application's demands. It's important to consider the performance of the application at its extremes e.g. running it on the largest, or most complex data sets that users are likely to use. You should also determine your input and output demands, for example, the size of any inputs and outputs and where these need to be located for your application. With this information you can then compare your application's demands to the resources offered by your cloud and determine if the cloud could run your application.

If it seems that the cloud will not be able to support your application then you should consider whether there are alternative ways in which your application could be configured and used to circumvent any limitations. You should identify if there any alternative programming models or techniques you could adopt to best exploit the cloud. If your application can be parallelized, or is already parallelized then consider if this is a possible solution. For example, in RAPPORT bioinformaticians experimented and consulted with the authors of an application, GenomeThreader¹³, to identify possible ways to partition input files to reduce the application's memory requirements.

If the cloud's resources still prove insufficient then you might want to consider whether there are other applications that do the same task but are more efficient and so can be run on your cloud. You may want to consider this approach if, for example, you were more concerned with the ability to scale to large numbers of parallel jobs or to conduct analyses of larger data sets than with the use of a specific application.

4.2 Assess the dependability of the cloud

- Will the cloud you choose be there for as long as you need it?
- For commercial clouds, what happens if you don't pay your bill? Is your content deleted? Are you warned first?
- Does the cloud provider manage backups and, if so, how often? If not, then is there a way for you to easily do backups?
- Is the help and support offered by the cloud providers adequate for you?
- Is there an SLA defining resource availability, downtime, networking bandwidth, etc.
- Do you have a contingency plan for if your cloud were to become unavailable? Is there another infrastructure you could use? Would you have the time, money and effort to migrate your content? What are the consequences if there is no alternative available?

If using a cloud then, like any infrastructure or service, you'll want to clarify:

- That the cloud you choose will be there for as long as you need it. For example, if you intend to use the cloud to deliver a service or preserve your data, check that the cloud will be available for your intended lifetime.
- For commercial clouds, what happens if you no longer pay your bills. You need to find out if your content is just deleted and, if so, whether or not you will be warned beforehand.
- What happens to your content if there are any problems in the cloud.
- Are backups done and, if so, how frequently. If not, then is there a means of backing your content up yourself.
- What you can expect in terms of help if you run into problems and whether this is acceptable given the expertise you have.

Your potential cloud provider should be able to help you answer all of these questions.

¹³ <http://www.genomethreader.org/>

You should also consider how you would handle the situation of your cloud becoming unavailable. You cannot assume that you could just move your content back in-house since you may be using resources on a cloud for which you have no equivalents available locally. You may want to identify other infrastructures you could port your content to, and whether you would have the time, money, and effort available to migrate your content. You should also consider what the consequences are if there is no alternative infrastructure currently available onto which you could deploy your content.

4.3 Consider legal and ethical issues

- Are you allowed to put your data on the cloud?
- Are there any community procedures, institutional policies or legal frameworks you have to comply with in both hosting data on the cloud and transferring applications and data to and from it?
- Is the use of a public or community cloud acceptable to your stakeholders?
- Does the licencing of your application allow you to deploy and use it on the cloud?
- Do you understand the licencing of your application or will you need to consult with advisory bodies e.g. OSS-Watch or JISCLegal?

There are a number of legal and ethical issues you must consider. Some of these will be complex and difficult to resolve.

Whether there are ethical issues depends on your research. If you are dealing with sensitive data e.g. medical data, especially patient data, there may be severe restrictions on how and where and in what form this data can be transferred and hosted. You may be able to host the data on a cloud but only if in an anonymised format, or you may have to ensure that software or data is adequately encrypted in transit, for example. You should understand any policies relating to the security and privacy of your data and ensure that these won't be violated if deploying this data onto a cloud. For public or community clouds you may also need to explicitly discuss these with your stakeholders (e.g. funders, data providers or users) to ensure that their use will cause no issues.

There are also issues relating to copyright and, particularly, licensing of any software you need. Whether the software you use is proprietary or open source, was free or paid for, there may be terms and conditions attached that affect whether you can use it on a cloud. For example, deploying the software on a cloud could be considered as supplying it to a third-party and violate the terms of a licence that prohibits distribution to third-parties. 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. OSS-Watch¹⁴ provides consultancy in the use of open source software and may be able to help you with understanding any restrictions on the use of your software on a cloud. JISCLegal¹⁵ can provide advice on legal issues relating to data protection and storing data on third-party resources.

4.4 Consider financial issues

- Is the cloud you use free to use or will you have to pay for it?

¹⁴ <http://www.oss-watch.ac.uk/>

¹⁵ <http://www.jisclegal.ac.uk/>

- Can you estimate how much it will cost you? Is this within your financial abilities? Is this acceptable to your funders?
- How will you pay for usage? Are you happy to use your own credit card? Does your institution have a credit card you can use and would be happy for you to do so?
- How, and how often, will you monitor your resource usage to ensure you don't incur excessive charges?

If using a commercial cloud e.g. a public cloud, or certain community or private clouds that you have to pay for, you'll have to consider whether this is within your financial abilities to do so. You need to be aware of your application's demands (in terms of CPU, memory, disk and data transfer) and the period for which you'll need to use the cloud e.g. if deploying a long-running service or data. These can help you to estimate any costs and so determine if these are acceptable. You'll also need to ensure that such expenditure is acceptable to your funders.

You also have to consider how you'll actually pay for your usage. Typically, public clouds require a credit card to be registered. Are you prepared to use your own credit card or does your department or institution have a credit card that you could use, and would they be happy for you to use it? Community and private clouds might offer other pricing structures e.g. billing your department or organisation.

Some commercial clouds may offer pricing structures based upon levels of resource consumption (whether this be CPU, memory or storage) agreed in advance. If you exceed these they may just bill you at a higher rate, for example, you may be billed for a virtual server you are running, even when you are not actively using it. It's important to understand any pricing terms and conditions carefully so that when you come to use the cloud you are not hit by unforeseen and excessive bills.

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 programme¹⁶. You may wish to investigate these as possible options for your use of cloud. You should make sure you understand any terms and conditions carefully.

5 How do I port my application to my chosen cloud?

Now, you're ready to port your application to your cloud. How you do this really depends upon your preferred working practices – you may want to scope out the work in advance in some detail or you may just want to get on and experiment. We use the term 'port' to describe moving your application to the cloud. In reality, this process may be as simple as copying an application executable to a remote resource and running but it may equally require significant re-working of the application's configuration or code. We offer some suggestions below which can help with the porting process, but firstly, there is one thing that it will be very valuable to do at the outset which is...

5.1 Consider how you'll handle unforeseen but inevitable delays

- How will you manage problems in porting or using software on the cloud?
- How much time will you spend trying to get one piece of software working with another?
- Do you have a contingency plan in place with alternative options to explore?

¹⁶ <http://aws.amazon.com/education/>

- When will you decide that you've spent too long and either quit or explore alternative options?

Cloud are a relatively new technology which is only just beginning to be used in research. As a consequence, there is an absence of many of the tools and technologies that make the use of other infrastructures e.g. grids, straightforward. You need to accept that you may end up having to do a number of activities that are not directly relevant to your research, but which are necessary when porting your application to a cloud. This can include developing virtual images, and creating, managing and releasing resources on a cloud, activities comparable to those you would have to undertake if managing your own server.

Furthermore, cloud software that makes resources available and supports communication with end-users consists of a number of components involved in complex interactions. Some of these components may be evolving or prototypical and may not have complete or up-to-date documentation. Components that are designed to inter-operate e.g. cloud components and virtualisation components may differ in the APIs they support for inter-operability or not exactly conform to published APIs. Consequently, when porting your application, undertaking tasks such as developing a machine image, or when deploying your own private cloud system, you may run into problems with these components, both in isolation and when trying to use them in unison. These may incur time-consuming web searches and e-mail exchanges and trying different fixes to resolve the problems.

It is vital for you to consider how you will react to such problems, in particular how much time and effort you are prepared to expend in addressing issues and when you'll decide that you cannot afford to spend any more time on it. Complementing this, it's useful to have a contingency plan so that if you do run into problems you have alternative options to try, and the time to try them in. These are very much specific to your own circumstances but could include identifying alternative software you could use or trying to deploy your application in a different way. Ultimately, the time you choose to spend moving your application into a cloud environment is likely to depend on the expected future use of the application. If there is a case for regular use of the application over some reasonable time period, the effort and time that may be expended moving the application into a cloud environment and configuring it to provide simple user access could be time well spent. If you have only a one-off use case, it may be more efficient in terms of time, effort or expenditure to use an alternative infrastructure e.g. an existing cluster or a grid.

5.2 Port your application

How you actually port your application depends, of course, upon your application, your cloud and your personal preferences. However, there are a number of things that have the potential to make the porting go smoothly and that will also help others who wish to undertake similar work in the future. Many of these are common techniques applicable across a whole range of software development activities.

5.2.1 Exploit the knowledge of others

You may want to check the web to see if others have ported applications similar to yours. How they did it may help you on your way or provide ideas if you are stuck.

Though clouds allow you to provision resources without waiting for a system administrator to do this for you, administering these resources is no easier than running a server, so do not neglect your local system administrators. They have a wealth of expertise in system configuration and deployment and may be able to help you if you run into problems. Indeed, they may be prepared to work with you, or manage these for you, which may prove more efficient than doing everything yourself.

5.2.2 Don't try to do everything at once

Don't do everything in one go as when things go wrong you'll be less sure where the problem might lie. So, for example, in RAPPORT, researchers in physics wanted to run their analysis software, CMSSW¹⁷, on CernVM¹⁸, a virtual machine made available by CERN for Large Hadron Collider Experiments, and then deploy these onto a cloud. They worked on deploying CernVM onto the cloud. Once this was done they worked on deploying CMSSW onto CernVM. Once this succeeded they then deployed CMSSW onto a CernVM running on the cloud. Breaking down the task in this way allowed issues to be identified more rapidly since the potential sources of problems caused by interactions with the various components was reduced.

5.2.3 Write down what you try and share your experiences as you go

Note down what you do. If you don't already, it's useful to record what you try. This can include:

- Software used, including its version.
- Commands run, arguments provided, data files used.
- Errors encountered.
- Errors you ignored that turned out to be OK to ignore
- Fixes or workarounds you applied, how you applied them and where you found these.
- Links to sites where you found hints and tips.

Once you get things running this will give you a record of how you got there. This can be useful if you have to take a break for a few days e.g. you're away at a conference, as it allows you to reproduce your steps.

Such a log also provides raw material that can be converted into a HOW-TO or tutorial for both yourself and other researchers, so they too can deploy your application. It may also help others as you are likely to have solved deployment or configuration problems that they have encountered too and sharing this information can save them from having to spend time solving the same issues again.

You may want to consider publishing your experiences as you go, rather than waiting until you have finished. You could do this as a set of web pages or as a blog. Others working on similar porting activities may then benefit from your experiences which could save them encountering the same problems you did, or allow them to solve these more rapidly. Alternatively, those that have successfully ported their applications may offer advice and guidance. We return to this in section 5.3 on sharing your outputs and experiences. If you're reluctant to share your experiences and work-in-progress, in case a critical reader takes these as hard facts or passes comments on their quality, you can always add a disclaimer stating that these are *your* notes which you're publishing in the hope that they may prove useful to others.

5.2.4 Let suppliers know of bugs

It's useful to pass back bugs, comments and suggestions to software providers. If you encounter problems in user documentation, things that are wrong or are not clear, or you detect bugs, notify the

¹⁷ <https://savannah.cern.ch/projects/cmssw/>

¹⁸ <http://cernvm.cern.ch/>

software providers. This will contribute to the improvement of the software products and make their use easier for future users.

If suppliers provide e-mail lists, ticketing systems or forums then use these so that bugs are recorded and discoverable by others. If not then you may want to highlight bugs via your web site or blog. These can ensure that others are aware of the bugs and don't spend the same time you did in trying to detect, or solve, them.

5.2.5 Keep an eye on your resource usage

For clouds that you have to pay for, keep an eye on your resource usage. For example, you may want to shut-down your virtual instances if you're not using them. This can help you avoid any nasty surprises when you are billed for your usage.

5.2.6 Assess the stability and maintainability of other software

You will likely find that you need to use other software when deploying your application to the cloud, whether this related libraries, cloud configuration software or virtualisation tools. There are a number of issues to consider when selecting what software to use, many of these apply to selecting software for any purpose.

Consider the status of the software. Is it stable or a prototype? Is it developed by a team or a single developer? Does it have a user community, or is there evidence of users? Is there any support available? Is there documentation? Evidence of a large number of developers, high-traffic on e-mail lists, bug and issue trackers, lots of documentation and prompt responses to queries can indicate that the software has an active user community and there'll be help available if you get stuck. Unanswered questions, a solo developer, limited references to the software on the web could imply that the software has stagnated and it won't be so easy to get support if you run into problems.

You should also consider whether the software can be run on the cloud both from a technical and licensing perspective and whether you have the expertise, or the expertise is available (either in your project, institution or community, on the web, or from the software's developers) to help you overcome any deployment problems. If not, then see if there are suitable alternatives available.

5.2.7 Think of your users

You may be the primary user of your application once it's deployed, or it may be your colleagues. It can be useful to consider how you could make your application easier to use once deployed. If the application takes dozens of steps to set up and configure, could you do this via a script? Or via a web interface.

Thinking of how easy it is to use your, now cloud-enabled, application and coming up with an improved approach could benefit you and your colleagues. For example, an investment of a day or two writing a simple web interface, asides, from being a fun diversion, may save you time in the longer term. It may make it easier to demonstrate your application at conferences, or yield a more visually engaging experience. It could also improve the accessibility of your application for others in your community, which could increase its uptake and consequently allow to demonstrate demand if seeking future funding for your work.

5.3 Share your outputs and experiences

When you're done, let people know about what you've achieved and how you did it. It may be of interest to others who wish to exploit and use your application to contribute to their research. You may want to consider making any components you've developed or machine images you've produced available to others (if the licencing of the software you use permits) so they can use your application in

their research. Or, you may want to consider exposing your application as a service to others, if applicable.

Your work in porting may be of interest to others who are porting similar applications and want to see how you did it. Sharing your experiences about how you ported your application, the software you used, bugs you encountered and patches you tried can save others going through the same trials-and-errors and searching for solutions that you applied in your porting.

If deploying data or porting applications to be exposed as Software-as-a-Service you should ensure that the availability of the software or data is publicised to your research community (e.g. via e-mail lists, blogs, Twitter, or posters and presentations at conferences). This will encourage other researchers to reuse your software and data in their research, which has the potential to improve the quality of your research and the quantity and cost-effectiveness of theirs, to your mutual benefit.

6 Migrating between cloud providers

After a period of time you may find you need to migrate your application from one cloud provider to another. There are many reasons why you might need to move. For example, it may be because you are using a commercial cloud and their pricing becomes prohibitively expensive, or you may be using a private cloud that the provider no longer wishes to maintain. The activities involved in migrating an application from one cloud provider to another are no different to those we have presented above for porting an application to the cloud. You need to consider your application and its requirements, the dependability of your target cloud provider, and the legal, financial and technical issues the migration may incur.