Research Councils Workshop on:

Cloud Computing for Research

The *Window* Conference Centre, London, Tuesday 20 July 2010

INTRODUCTION

This is a report of a workshop hosted by the Research Councils, in July 2010, to discuss the opportunities and challenges surrounding the use of cloud computing for research in the UK. At the event, a number of recommendations were made, aimed at the Research Councils, providers of e-Infrastructure, commercial cloud vendors and higher education institutions. Recommendations made by four breakout groups can be found, highlighted in text boxes, in the relevant section of the document below. A number of these recommendations were also raised during a final plenary Q&A session. These can be found in the final section, "Final Panel Discussion".

The Research Councils consider this workshop the beginning of an ongoing consultation on their role in facilitating research use of cloud. While we assess our future strategy in supporting cloud for research, in the light of the outputs of this workshop, we continue to welcome feedback on the subject. Please send any comments to matthew.davis@epsrc.ac.uk or michael.ball@bbsrc.ac.uk.

On behalf of the Research Councils,

Matthew Davis, EPSRC
Michael Ball, BBSRC
Andrew Lawrence, EPSRC
Richard Bailey, EPSRC

July 2010

WORKSHOP AIMS

On 20 July 2010, the Research Councils held a workshop in London to discuss the opportunities and challenges in the use of cloud computing for research. The event was co-chaired by Professors Malcolm Atkinson and David de Roure. It was attended by around 50 people, including UK academics from across the Research Council disciplines, providers of university e-Infrastructure, representatives of JISC, commercial cloud providers (Microsoft, Amazon and IBM), and RC staff. An attendee list is given in Appendix I.

The aims of the meeting were:

- To enable networking and information exchange between key potential research users of cloud, cloud computing vendors, and the Research Councils;
- For the cloud vendors to brief attendees on the cloud computing services they can offer;
- To highlight key opportunities in cloud computing for UK research;
- To discuss the issues around the use of cloud computing for academic research, and how they can be addressed.

WORKSHOP STRUCTURE

The workshop agenda can be found in Appendix II. The workshop began with invited presentations, followed by group working and a plenary panel discussion in the afternoon.

In the morning session, invited presentations were given by Professor Paul Watson of Newcastle University and Dr Glenn Proctor of EBI on their experiences of using cloud computing. These were followed by presentations from three commercial providers of cloud: IBM, Microsoft and Amazon Web Services. The morning concluded with a panel question and answer session in which all five speakers participated. The afternoon session began with a final presentation, by James Farnhill from JISC, on three recent JISC-supported projects on Cloud Computing for Research[1].

In the afternoon, delegates were divided into breakout groups to discuss major issues around the use of cloud computing for research, and how they might be addressed. Before the event, based on a pre-event delegate survey (see Appendix III), the event organisers and chairs had identified 5 potential breakout topics for this session. These were: *Tools, software and standards*; *Supporting the research data life cycle*; *Financial models and adoption criteria*; *Training, capacity building and methods*; and, *Use of cloud for cross-Council strategic themes*. Over lunch, delegates were asked to vote on which

[1] http://www.jisc.ac.uk/whatwedo/programmes/researchinfrastructure/cloudcomptechreview.aspx; http://www.jisc.ac.uk/whatwedo/programmes/researchinfrastructure/usingcloudcomp.aspx

topic they preferred to discuss. As a result of the delegate vote, four topics were taken forward for discussion (membership of each of the breakout groups is given in Appendix IV). Each group was chaired by a nominated rapporteur. The breakout discussion topics and nominated rapporteurs were as follows:

1. *Tools, software and standards* (David Snelling, Fujitsu Labs Europe)

2. *Supporting the research data life cycle* (Richard Baldock, MRC)

3. *Financial models and adoption criteria* (Neil Chue Hong, University of Edinburgh)

4. *Use of cloud for cross-Council strategic themes* (Sheila Anderson, Kings College London)

BREAKOUT SESSIONS

Within these topics, breakout groups were asked to brainstorm the following questions:

- *In this area, what opportunities does cloud computing present to UK researchers?*
- *What advantages does cloud computing offer over existing technologies?*
- *What are the major challenges in the use of cloud computing by researchers in the UK?*
- *What should be done to address these challenges?*
- *Who is in a position to address these challenges*?

Notes of the outputs of each of these sessions, captured on flipcharts, are given below. The recommendations made each of the breakout groups are highlighted in text boxes.

1. Tools, Software and Standards

In relation to tools, software and standards, what opportunities does cloud computing present to UK researchers?

Cloud computing acts as a *catalyst* for:

- Tool developers for better delivery (Software as a Service).
- Data as a service.
- Creation of workflow standards.
- Metadata services/standards.

In relation to tools, software and standards, what advantages does cloud computing offer over existing technologies?

- Cloud offers better development delivery (virtual Machines).
- Cloud offers commercial leverage.

What are the major challenges in the use of cloud computing by researchers in the UK?

- Heterogeneity of cloud platforms.
- Money is needed for development of tools, software and standards.
- The learning/awareness curve for users of cloud.
- Software licensing issues.
- Optimisation of cloud technologies

How might these challenges be addressed? (And, who is in a position to address them?)

- In terms of addressing the heterogeneity of cloud platforms, we should wait for the market to respond first.
- To develop tools, software and standards, spot grants for research, deals with vendors, and development environments are required (JISC, RCs, Vendors).
- To improve awareness of cloud computing, case studies and/or awareness programmes are needed (RCUK, JISC, peer support).
- The UK Research Councils, with international Research organisations, should lobby around licensing issues (RCUK and NSF etc.)
- Benchmarking of cloud technologies is needed for optimisation (by individuals, research funders).

What might be the effect of such action on the other topic areas discussed?

- Waiting for standards could lead to vendor lock-in.

2. Supporting the research data life cycle

In relation to the research data life cycle, what are the opportunities enabled by cloud computing?

- Cloud computing allows data to be accessed globally.
- It can be seen to be open.
- Enables both preservation and archiving of data.
- Allows access of computation to the 'long tail' of researchers who may have extremely strong computing skills.
- Accountability – allows the value of a given resource to be established.
- Disintermediation.

What advantages does the cloud offer in relation to research data?

- Moves the focus from hardware to service – i.e. computer scientists can focus on developing software/services rather than maintaining the hardware.
- Moves the focus onto data management.
- Reliability.
- Scale out deals with peak loads, in terms of bandwidth and computing.
- Cloud allows management of energy costs and opens the way to 'green' energy.
- Access mechanisms can be built on top of the cloud, e.g. *Dropbox*.
- Allows co-location of processing and data.

In relation to the research data life cycle, what are the challenges associated with cloud computing?

- Permissions for international/global access to data.
- Legal barriers for international access e.g. data protection act mainly/particularly for individuals.
- Portability/mobility/lock-in of data commercial policy changes
- Widening the user base.
- Gaining trust of organisations with a commercial aim.
- There is a perceived need to provide more sophisticated methods of virtualisation allowing transfer of data across clouds.
- Who will port the tools and client codes needed to enable access?
- The diversity of users, including skilled and unskilled users, allows misuse.

What should be done to address these challenges, and who is in a position to address them?

- Ethical questions around the balance of access and privacy should be addressed by lawyers.
- Issues around portability should be addressed by computational scientists.
- Cloud providers should agree and adopt standards.

- Research funders should coordinate globally to adopt SLAs that provide access.
- Computer scientists and developers can help with the portability of tools.
- An agency should broker deals with cloud providers.

3. Financial models and adoption criteria

In relation to financial models and adoption criteria, what are the advantages of cloud computing?

- Scalability.
- Cloud resources can be used in predictable bursts.
- Cloud is advantageous to multi-institutional working, providing a single point of log-in.
- Low access barriers to using cloud (you just need a credit card)
- Cloud billing is easy (for example, as a user you are not billed in a lump sum at the end of financial year).
- Although algorithm development is non-trivial, there are low training barriers to using existing applications (like *Matlab*) on the cloud. Integration of existing tools to the cloud is key to wider adoption.

What are the financial barriers to using cloud?

- Moving capital funding to recurrent spending (e.g. at Research Council level).
- The major financial barriers to the use of cloud are in cost prediction and planning. Transparency and prediction of the cost of computing, data storage and transactions, and the links between them, are barriers. Not knowing up front what a given activity will cost and being able to project this: to know the cost, you need to run the activity first.
- It is difficult to identify the right financial model for cloud computing when the real cost of cloud computing for research is not yet clear.
- As an extension to this, (full economic) costing and justification of cloud resources on Research Council grant applications. For example, costing of (long term) storage of data. However, the true, full economic cost of using e-infrastructure supplied at an institutional, central service level (in terms of power costs, upgrades, man hours) is likewise often hidden.
- Although cloud computing can be paid for by personal credit cards, if data storage is paid for this way it may be tied to individual researchers rather than the institution. What happens when the research leaves that institution? There are IP issues around the personal use of cloud within academia.
- By way of a solution, cloud providers do offer consolidated billing for multiple-users. However, they do not micro-manage group accounts and there is the danger of individual users (via “runaway experiments”) running up large bills.
- Although not a financial barrier, the lack of wider support for scientists using cloud is a barrier to uptake. There are opportunities here for enterprising scientists. Training in cloud methods should be the role of the university.
- Trust is a barrier in terms of choosing cloud providers (especially for data storage).

What should be done to address these barriers, and who is in a position to address them?

- Cost prediction should be investigated further (e.g. through test cases) as inability to predict costs restricts financial planning at a number of levels. Transparency of pricing and costing will be a learning process for the community.
- To enable cost prediction, cloud service benchmarking will be needed.
- Promotion of the use of cloud computing by UK academics will require a financial incentive: for example, a £multimillion open call (similar to one by the US National Science Foundation) to promote the use of cloud computing and to investigate its potential for research.

What potential role might a private (UK academic) cloud have, and what are the financial advantages?

- Such a cloud would not benefit from the same economies of scale as commercial clouds.
- What would be the price performance?
- Potential advantages of a private cloud are clearer billing, running of software that is too expensive to run on commercial cloud, and better band width for data transfer.

4. Use of cloud for cross-Council strategic themes

Discussions in this breakout group centred around the additional requirements faced by researchers working in cross-disciplinary areas, such as climate change and the ageing society.

In relation to the cross-Council research themes, what opportunities does cloud computing present to UK researchers?

- For many researchers, cloud computing would offer access to these kinds of facilities for the first time. There is a bottom-up need for this sharing of facilities (for example, in digitising humanities).
- Use of the cloud frees you from maintaining hardware, to concentrate on developing software.
- Take up of cloud computing is most likely to be by those who don't have an alternative. There is a long tail of researchers who don't have university resources; cloud will help them collaborate.
- Enables translation of data between research communities, decoupling data and applications.
- Cloud offers a scalable foundation: it can support bursts of demand.

In relation to research in the cross-Council themes, what advantages does cloud computing offer over existing technologies?

- It will be quicker to trial stuff (e.g. skunk-works). It will enable new collaborations to start quickly and fail quickly (and quietly). Researchers can try out other tools without re-inventing them.
- The sharing of approaches can be pushed forward by cloud.
- Easier to produce commercial versions of your software. Cloud offers better control of IP and licences.
- For central computing providers, cloud lowers the risk (recurrent not capital spend).

What are the major challenges in the use of cloud computing by researchers in the UK?

- Cloud is not yet mature in its offerings, although this could be because the cloud is not offering what researchers want. There are different levels of user need: Software developers who are IT literate, workflow creators, and users who just want to click the run button.
- There are issues around 'cultural' control: for example, ownership of data and ownership of hardware ("hugging your cluster").
- There are many activities going on in parallel in cloud computing. There is a need to avoid redundancy.
- There are issues around longevity of data. There is no such thing as zero risk.
- There are barriers in terms of licence conditions for access, internal strategies and public policies.

- Are people using cloud computing without realising? If there is a perception issue, should we move people to thinking more about 'cloud'?
- The perception of what cloud costs, especially in current financial climate.
- There is a trend for the centralisation of computing service, whereas cloud is effectively out-sourcing.
- There are issues around the storage of public data: how public is it?
- Not all researchers have grants to be able to fund cloud use. How should other researchers be funded to use cloud computing? Would use of micro-payments make it more attractive?

What should be done to address these challenges, and who is in a position to address them?

- Involve Centres for Doctoral Training and PhD students: change the culture towards cloud.
- Development should concentrate on solving researchers problems: it should not be a technology 'push' but should be about making it easy.
- Host a competition in cloud. Have open source data available and give a prize for cloud sourced research solutions. Have cloud access as the prize.
- Make sharing of research data on cloud mandatory. (Should we have a 'British Library' for data? What about its governance? Do we just need a repository or do we need regular access? There is a spectrum of needs.)
- It is not clear whether there is a need for a dedicated academic cloud, or whether use of commercial clouds should be supported. Would a JISC supported academic cloud be a true cloud?

FINAL PANEL DISCUSSION

In the final session, the rapporteurs from each breakout group reported back to the rest of the delegates. Following this, to close the workshop, the four rapporteurs took part in a plenary panel session to discuss the outputs of the breakout sessions. The following is a summary of the panel discussion:

The suggestion was made that the Research Councils hold a substantial (£multimillion) call in cloud computing linked with vendors, analogous to a recent activity by NSF in the States, as an incentive to UK researchers to exploit cloud computing services. Such a major promotional activity would catalyse action and leverage money from the cloud providers. Projects, for example exploratory studies, would lead to science outputs as well as cloud learning.

The point was made that the optimised research use of cloud computing will require a sophisticated set of software tools to allow researchers to exploit commercial clouds, analogous to what the UK e-Science Programme provided for the research community to exploit the grid. Much work is needed on prototyping and testing of research tools and software on the cloud.

There is incomplete evidence to show the (financial) efficiency savings of using cloud services against existing e-infrastructure provided by institutions and JISC. Although the case is made for compute aspects, the case is weaker for certain aspects.

There was a general discussion about the case for a private, academic cloud. The case for an academic cloud is still out. Commercial clouds are appropriate unless you really need local infrastructure, or have legacy issues. Economies of scale are smaller for smaller cloud clusters. However, if it cuts down barriers to the use of cloud computing, then there may be a case for a UK academic cloud.

The case for operating a private cloud then 'bursting out' into a commercial service, for example when extra data storage is required, was discussed. It was not clear to what extent, or where, economies of scale would apply for such a mixed operating mode.

Finally, a commercial cloud service provider asked what single thing the rapporteurs would like cloud vendors to do to facilitate research use of cloud. The responses were: better and more informed cost prediction for the use of cloud services; predictability and security of research data on the cloud; support for digitalised humanities data; and, the standardisation of cloud Infrastructure as a Service (IaaS) provision to avoid vendor lock-in and non-interoperability.

Appendix I. Workshop attendees

John	Ainsworth	Manchester
Sheila	Anderson	KCL
Malcolm	Atkinson	Edinburgh
Jim	Austin	York
Richard	Bailey	EPSRC
Richard	Baldock	MRC
Michael	Ball	BBSRC
Mark	Birkin	Leeds
Jon	Blower	Reading
Dan	Bretherton	
Mario	Caccamo	TGAC
Neil	Chue Hong	Edinburgh
Amanda	Collis	BBSRC
Peter	Coveney	UCL
Simon	Cox	Southampton
Vanessa	Cuthill	ESRC
Denise	Dabbs	EPSRC
John	Darlington	ICL
Matt	Davis	EPSRC
David	de Roure	Oxford
Margarete	Donovang-Kuhlisch	IBM
Martin	Dove	Cambridge
James	Farnhill	JISC
Donal	Fellows	Manchester
Paul	Flicek	EBI
Jeremy	Frey	Southampton
Iain	Gavin	Amazon
Neil	Geddes	STFC, NGS
Alastair	Gemmell	
Carole	Goble	Manchester
John	Gordon	STFC RAL
Mark	Hayes	Cambridge
Alex	Heneveld	Cloudsoft Corp
Andrew	Herbert	Microsoft
Andrew	Lawrence	EPSRC
Steve	Lloyd	QMUL
Sean	May	Nottingham
Stephen	McGough	Newcastle
Jane	Nicholson	EPSRC
Christine	Orengo	UCL
Ian	Osborne	Intellect
Glenn	Proctor	EBI
Andrew	Richards	STFC&NGS
Adrian	Rowland	UKTI
David	Snelling	Fujitsu Labs Europe
David	Thomas	NERC
Lesley	Thompson	EPSRC

Anne	Trefethen	Oxford
Jano	van Hemert	Edinburgh
Paul	Watson	Newcastle
Matt	Wood	Amazon
Jie	Xu	Leeds
Stefan	Zasada	UCL

Appendix II. Workshop agenda

10.00	Registration; tea/coffee
10.30	Welcome by Prof Malcolm Atkinson, Co-Chair
Invited presentations from academic users about their experiences of using cloud:	
10.40	Professor Paul Watson, Director of Informatics Research Institute, University of Newcastle
10.55	Dr Paul Flicek, Team Leader, Vertebrate Genomics, EBI Glenn Proctor, Ensembl Software Coordinator, EBI
What can cloud provide? Presentations by commercial cloud vendors:	
11.10	Matt Wood, European Technology Evangelist, Amazon Web Services
11.30	Margarete Donovang-Kuhlisch, European Government Industry Technical Leader, IBM
11.50	Andrew Herbert, Managing Director, Microsoft Research, Cambridge
12.10pm	Speakers Q+A session.
12.45pm	Lunch and networking
1.30pm	Introduction to session by Prof Dave de Roure, Co-Chair
1.35pm	James Farnhill, Programme Manager for e-Research, JISC, Findings of a JISC-supported study of Cloud Computing for Research
1.50pm	Group working: • What are the opportunities and advantages in the use of cloud computing? • What are the challenges and how can they be addressed?
3.10pm	Tea/coffee
3.30pm	Groups report back
4.00pm	Plenary panel discussion and wrap-up: • What are the major challenges? • How can these challenges be addressed? • Is there a role for the Research Councils in facilitating research access to cloud?
4.30pm	Closure by Chairs

Appendix III. Results of the Delegate Survey

In advance of the meeting, delegates were asked to answer the following question:

> "*What do you think are the three most important issues regarding the use of cloud computing for research?*"

The responses received are shown below:

Data transfer costs to/from cloud providers
Cost per cycle, and means to fund cloud use (e.g. via EPSRC grant applications)
Security/data ownership
In the short term, the major challenge has got to be financial (e.g., proving that the greater transparency of costs associated with a utility cloud infrastructure is a benefit)
Over a slightly longer term, the big challenge is going to managing data volumes that will be coming from research, ensuring that they get delivered to cloud computation units in a timely fashion and with appropriate levels of security
Further ahead, curation is the big challenge as it will be necessary to retain checkpointed large data sets and applications to work with them, and in a form whereby in silico experiments may be repeated in the future, even if supporting infrastructures change
What the real cost of clouds will be and whether they will really be competitive.
Whether clouds can handle sufficient data and data rates.
Whether individual providers will be able to provide stable long term solutions.
Opaque business models, similar to the mobile telephone market, will become more prevalent in cloud computing as everyone is trying to protect increasingly small margins of profit in a market with a low entry barrier. How will a researcher know what is the best, in terms of value for money, cloud vendor to use for their particular tasks?
Data volumes will keep increasing. Already Amazon has started a scheme to use physical drives (as opposed to the Internet) to upload data into their infrastructure. Some groups already have data larger than can be shipped via this scheme. This leads to a major problem with sharing. Even if the data can be put into one cloud vendor, this will create a lock-in mechanism as it will be difficult to get the data out and into another infrastructure (or even cloud vendor) again.
User interfaces are the key. Researchers do not invest their valuable time into learning tools or infrastructure, they actually learn the interface to tools and interfaces. Tools and infrastructure can easily be changed if the user interfaces, i.e., the modus operandi, can remain the same. It is vital to ensure the right interfaces are produced first for the data/compute intensive tasks researchers want to achieve. Then we can find and adapt the appropriate underlying software (think analysis packages, simulators, modelling, etc.) and hardware (whether virtual or real).
Cloud vs HPC computational models. We are not ready to replace HPC supercomputers with clouds. Supers are designed to give peak performance for single highly parallel computational simulations. Clouds are designed to deliver scalable reliable performance to collections of on-line services and massive data analysis capabilities on hosted data. Clouds use automatic replication and restart to provide 24x7 application reliability.
Long term data archive costs, especially for data that needs public on-line interfaces. As research in science becomes more multi-institutional and geographically distributed, the cloud is a very attractive way to host shared data collections. Cloud data is replicated so data loss is not a problem. Cloud providers upgrade hardware themselves and it is not the burden of the institutions to replace data servers and migrate the data. Cloud downsides: getting the data there in the first place. Best cloud data collections are grown from birth in the cloud or accrued incrementally (such as a web crawl).

Linking desktop applications to the cloud. Scholars want to solve bigger problems and study bigger shared data collections. But they don't want to have to learn about clouds and Grids and HPC. If we provide them with seamless ways to access cloud based research tools and data such as maps, web search etc they win big. How do we support users who want a way to extend the power of their desktop/pad/phone and not just the tradition consumers of big iron?
Provider lock in - one must make an early and somewhat arbitrary choice of which cloud provider to use, and it is not straightforward to switch providers later on
Data volumes - particularly in bioinformatics, much of the work requires large volumes of data (on the terabyte scale), and getting this into and out of cloud systems is challenging
Security - there is a perception that if it's not under direct management control from within an organisation, it is somehow insecure, even though this is generally not the case
Ease of use of the systems by non computer experts.
Compatibility between all systems, so that use of any platform can be achieved.
Security of data and services.
Standardisation of access methods to avoid locking researchers into one provider and to prevent barriers to collaboration across providers.
Accounting and auditability.
Data in all its forms – access to, publishing of, sharing.
We must characterize those categories of research applications that will benefit from cloud based deployment. The benefits can be in terms of performance (unlikely), cost (likely), easy of use, environmental impact, etc.
Data integrity and privacy assurance framework for cloud suppliers to adhere to.
Transparency to the environmental impact of a cloud based application execution. We don't need to see the real cost or performance, but the environmental impact will be important in time.
understanding why researchers cant/don't use (buy) cloud services today ? (ie untangling "more funded computing please" from real barriers
what, if any, is the role of private clouds
Are data clouds really any use to anyone
Reproducible research: using the cloud to provide long term access to services and data, and in aiding peer review and reuse through encapsulated research environments.
Scale: making use of elastic infrastructure and storage to process increasingly large, complex data.
Funding models: the move from initial upfront investment to operational spending and long term hosting costs
Security of data - certainty that the data is secure, only accessible with user controlled permission, all transmission encrypted.
Security of computation - certainty that the calculation requested is executed as expected e.g. the compute environment must be completely defined and reproducible.
Security of access - resource available when required, e.g. some guarantee of service at what ever level of agreement.
Of course with these constraints satisfied the cost then becomes the deciding factor.
Few researchers are actively using the cloud as a primary tool to aid their research (as opposed to those who are using the cloud as a research area);
Perception and awareness of available cloud tools amongst researchers
It involves significant effort to port existing ways of working and existing work to the cloud.
What do the cloud providers do to facilitate researchers moving all three of the above easily onto their platforms?
What standards and software frameworks should researchers (and their data/software applications developers) use to minimise the cost of moving platforms?
How should researchers, research communities and research funding bodies establish criteria which accurately guide them as they decide to move their research computations, onto, between and off cloud computing platforms?

Information Assurance - confidentiality, etc
Contracts and SLAs - management overheads associated with
The impact of pay as you go on innovation - may curtail the freedom to experiment
Most current codes do not target cloud infrastructures efficiently - particularly those involving complex data.
It is difficult to compare pricing / SLAs between commercial clouds and academic HPC and cluster resources
There aren't clear statements from an authoritative body regarding the use of commercial clouds for processing sensitive data, or for licensing issues.
Not every type of research problem, computational or data problem, is suitable for deployment in the cloud. Ensure that services provided in the cloud aimed at research are specific, well identified/described, and focussed on providing a specific type of solution rather than trying to solve everything for everyone. Cloud computing is not a panacea.
Ensure technology utilised is based on an industry standard where possible. Enable institutes and researchers to build and provide clouds that do not become bespoke and ultimately restrict utilisation by researchers.
Not every researcher can invest time and effort in solving technological problems. The barriers to entry in using cloud based resources need to be low. Simple to use interfaces
How to make it easier for those without strong IT skills to take advantage of clouds in their research
Accounting, payment and quotas.
Vendor lock-in
Internet-based virtual computing environments for large-scale research and problem solving
Efficient and effective aggregation and utilization of resources and services in the presence of uncertainty and faults
Effective provenance of data to increase the trust that researchers/users will place in a given dataset for research
Making cloud computing affordable for academic institutes who don't have the financial clout of many players in the private sector
Getting a body of researchers in place across institutes who have experience in using and setting up cloud services, as this is a relatively new area.
Looking at the best ways of moving forward with 'cloud clusters'. Some parallel research models etc will not be able to take advantage of cloud computing as they require fast, low latency networking between nodes. Amazon are moving in this direction, but this is a new development so it will be interesting to see how things develop, and how good the performance is.
Opportunities for collaborative working-Cloud based compute resource can provide a platform for national and trans national collaborative effort in data and analysis sharing, helping to increase the productivity of collaborative efforts .
Shared methodology- Once methods and analysis techniques have been developed they can more easily be shared amongst a research community, this has the potential to reduce effort in new tool and analysis creation
Economic model - cost of Cloud vs. own provision and support
Marginal costs - how far could HPC support be extended
Availability of top end HPC processors
Commercial cloud offerings still too expensive. Explanation: Academic cloud infrastructure (e.g. NGS cloud) is in its infancy, so researchers must rely on commercial cloud services. At present these are only affordable with large discounts from market rates, which are given at the discretion of service providers and not always available.

Academic clouds could save institutes money and enable cloud use by data intensive applications. Explanation: The report recommends that production cloud services should not be set up by the academic community in competition with commercial clouds. However, academic cloud services, provided by the NGS for example, would enable the academic community to pool resources in order to take advantage of economies of scale in terms of provision for power, cooling, system administration and bulk buying. Academic services could also enable applications involving very large volumes of data to take advantage of cloud computing without being prohibitively expensive.
Specialist skills requirement makes research project infrastructure involving clouds hard to maintain in the long term. Explanation: Cloud computing support is not available from centrally managed IT services in many (or most?) institutes. Therefore research projects wishing to take advantage of clouds must develop and maintain the necessary specialised skills themselves. This leaves cloud based research projects more than usually vulnerable to staff turnover, particularly in an environment where the majority of employment contracts are temporary. This has implications for the sustainability of cloud based research computing infrastructure.

Appendix IV. Membership of breakout groups

Tools, software and standards	Donal Fellows Tom Jackson John Darlington Jie Xu Glenn Proctor Alistair Gemmell James Farnhill Martin Dove Paul Flick Steve McGough Dave Snelling Andrew Richards
Supporting the research data life cycle	Sean May Richard Baldock Mark Birkin Malcolm Atkinson Mark Thorley Alex Heneveld Vanessa Cuthill Carole Goble
Financial models and adoption criteria	Anne Trefethen Stefan Zasada Steve Lloyd John Ainsworth Dan Bretherton Mark Hayes Andrew Herbert Jano van Hemert Simon Cox Neil Chue Hong Matt Wood
Use of cloud computing for Cross-Council themes	Margarete Donovang-Kuhlisch David De Roure Paul Watson Mario Caccamo Sheila Anderson Neil Geddes

Appendix V. Briefing paper on "Cloud Computing for Research" provided to delegates in advance of the workshop

Abstract

This paper introduces some of the potential benefits and challenges related to using cloud computing for research in the UK. The focus of the paper is on commercial rather than private clouds. Many of these opportunities and challenges are discussed in more detail by a recent JISC study[2]. This paper, together with the JISC studies, provides a useful starting point for discussion at the workshop.

Cloud Computing

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

Commercial cloud services offer a 'utility' model of computing where individuals do not have to invest in hardware and can instead buy or rent compute cycles and storage capacity from service providers. Costing models vary, but are typically by CPU hours. Components of commercial clouds include 'Platform as a Service' and 'Software as a Service'. In 'Platform as a Service (PaaS)', an additional layer of capability is provided to allow users to develop bespoke applications to run on the cloud. Examples of PaaS include Amazon's Web Services, Windows Azure, and the Google App Engine. In 'Software as a Service (SaaS)', the software remains the property of the provider and access is provided by subscription or on a pay-per-use basis. Alternatives to commercial clouds are private clouds, formed from pooled resources within the closed infrastructure of a single or group of organisations. The potential benefits and challenges vary between commercial and private clouds.

Benefits and Opportunities

- Reduced infrastructure costs

Since the user will access resources that are maintained and managed by the service provider, cloud computing has the capacity to cut down infrastructure costs considerably, both in terms of hardware and IT staff costs. Computing is provided as a utility, with users billed typically by CPU hour or bandwidth, thus removing the need for capital investment. For new users the barriers to entry are low, as initial trials can be low cost. This is an obvious benefit to academic users with more access to resources for consumables than capital.

- Scalability

[2] http://www.jisc.ac.uk/whatwedo/programmes/researchinfrastructure/usingcloudcomp.aspx

One of the attractive features of cloud computing is scalability. For example, where charging operates via CPU hours, it is the same cost to rent a 10-node cluster to compute for 40 hours, as it is to utilise a 400-node cluster for one hour. It therefore has the potential to offer high-performance computing to researchers who would not otherwise have access.

- Flexibility

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

- Data sharing

Cloud vendors can also provide data facilities, providing alternative strategies for storage, recovery and management of data. Cloud computing also provides potential opportunities in data sharing. Potentially, researchers could place data in clouds, and make them accessible for third party use. As software can also be provided through clouds, tools used to interrogate the data can also be made available without having to upload them separately.

- Competitive pricing

If moving data and software between service providers is relatively straightforward, users may have the opportunity to take advantage of competitive pricing. Subscriptions or contracts, however, may complicate this.

- Green computing

Cloud providers may be able to provide similar services using less energy and/or energy from renewable resources.

Gaps and Challenges

- Capability computing

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

- Charging systems

According to a recent study, the charging systems operated by service providers may be cost effective for small and medium-sized users, but not for heavy users who own their own compute infrastructures already.

- Access and usability

One of the primary algorithms used in cloud computing is MapReduce, developed by Google, and the open-source version Hadoop. It is unclear how

simple utilisation of software and such algorithms will be for users with little coding experience. Programs such as Hadoop and MapReduce are dedicated to manipulating data within clouds, but conventional software may require modification to use in conjunction with clouds.

- Security and risk

The most highly discussed concerns are with data security. Concerns arise because data are secured in the servers of the service providers, and the user has much less direct control over security. Cloud computing providers can potentially offer a range of possible security levels to users, but discussions are ongoing over the holding of sensitive and personal data by third parties. Other risks, such as company failures whilst holding research data, should also be compared with alternative provisions.

- Bandwidth issues

As the hardware is remote from the user, there is the potential for users to access their work using much more lightweight, portable web-enabled interfaces. However, for data-rich users, bandwidth can be a considerable problem. There are a number of potential solutions to bandwidth problems, including using data distribution tools, to sending physical drives with data on them to the service provider. Linking physically to a hub via a dedicated link bought or leased from a telecommunications company is a further option, but apart from being expensive, reduces the agility of being able to change service providers if a rival service becomes less expensive. There may also be a need for software applications that deal with data submitted to clouds, particularly where the data are particularly large or complex.

- Virtual machines

Most cloud computing uses virtual machines which encapsulate the users' software and data. Prepared collections of software to run on these machines are called 'images'. There may be benefits to organising arrangements for finding, maintaining and creating images to meet research requirements.

Michael Ball, BBSRC
Matthew Davis, EPSRC