



EPSRC-JISC Cloud Workshop 2011

Monday 26th September 2011

All Hands Meeting, Ron Cooke Hub, University of York

Workshop Annexes



Annex 1: Cloud Pilot Project Summaries

Title: *Flood Modelling for Cities using Cloud Computing*

Investigators: Dr V Kutija (Newcastle University), Professor C Kilsby (Newcastle University)

Summary: Cities are concentrations of flood risk because they are densely developed and tend to be situated in low-lying areas. Floods in UK cities of 2005 and 2007 have been a major driver for planning and research. However, risk from flooding from intense rainfall is particularly difficult to assess because it is sensitive to the spatial-temporal characteristics of rainfall, local runoff and surface flow processes, the performance of urban drainage systems and the exact location of buildings. Detailed models of water flow over complex surfaces are needed so that the effects of buildings, roads, vegetation and other features can be captured. Such models take a very long time to run on conventional computers and therefore need special facilities which have hitherto not been readily available. Fortunately, Cloud computing is now available. Clouds are systems where very many large computing jobs can be submitted to flexible, large facilities at relatively cheap cost without the user having to buy and manage the facility themselves: they simply pay a fee for time and storage of data. The outputs of flood models for cities are of great interest to a range of end-users in local government, the Environment Agency and water utilities. Until now they have been unable to generate and use such models because they could simply not be run for the large areas needed or for the number of cases of rain storms required to cover all eventualities. This is now a real possibility with the Cloud, so we plan to demonstrate its use not only for flood mapping for a large city, but for a wide range of rain storms covering possible climates for now and the future to allow planning and adaptation to take place.

Title: *GATE Cloud Exploratory: Adapting the General Architecture for Text Engineering to Cloud Computing*

Investigators: Professor H Cunningham (University of Sheffield), Dr K Bontcheva (University of Sheffield), Dr V Tablan (University of Sheffield)

Summary: When you plug your fridge into the mains electricity supply you don't worry about all the technology sitting behind the wall socket -- it just works. Cloud computing is starting to supply IT in a similar fashion. No more worrying about backups, no more hours spent configuring a new or repaired machine -- just plug into the network, fire up your web browser and away you go. Researchers have tougher and more specialised IT needs than most, so to realise the same ease of use that the cloud now provides for email or word processing requires work in several areas. One of these areas is to adapt existing established research tools to the cloud, and that is what this project will do. Our tool is called GATE, a General Architecture for Text Engineering. Over the last decade the UK's GATE system has become a world-leader for research and development of text mining algorithms. Text has become a more and more important communication method in recent decades. Our children are now spending over 6 hours in front of screens; our evenings often include sessions on Facebook or writing email to friends and relatives. When we interact with the corporations and governmental organisations whose infrastructure and services underpin our daily lives, we fill in forms or write emails. When we want to publicise our work or share details of our leisure activities we create websites, post Twitter messages or blog entries. Scientists also now use these channels in their work, in addition to publishing in peer-reviewed journals -- a process which has also seen a huge expansion in recent years. This avalanche of the written word has changed many things, not least the way that scientists gather information. For example, a team at the World Health Organisation's cancer research agency recently found the first evidence of a link between particular genetic mutation and the risk of lung cancer in smokers. Their experiments require large amounts of costly laboratory time to test hypotheses, based on samples of mutations in gene sequences from their test subjects. Text mining from previous publications makes it possible for them to reduce this lab time by factoring in probabilities based on association strengths between mutations, environmental factors and active chemicals. A second area that has been revolutionised by new media is customer relations and market research, which are no longer about monitoring the goings on of the corporate call centre. Keeping up to date with the public image of your products or services now means coping with the Twitter firehose (45 million posts per day), the comment sections of consumer review sites, or the point-and-click 'contact us' forms from the company website. To do this by hand is now impossible in the general case: the data volume long ago outstripped the possibility of cost-effective manual monitoring. Text mining provides alternative, automatic methods for dealing with text. GATE provides four systems to support scientists experimenting with new text mining algorithms and developers using text mining in their applications:- GATE Developer: an integrated development environment for language processing components- GATE Embedded: an object library optimised for inclusion in diverse applications- GATE Teamware: a collaborative annotation environment for high volume web-based semantic annotation projects built around a workflow engine- GATE Mmir: (Multi-paradigm Information Management Index and Repository) a massively scaleable multi-paradigm index. We have identified a need for a particular type of cloud service in our research field and this project will implement it such that there is close to zero barrier to entry for researchers. Based on our preliminary investigative work, we expect to complete a production quality service within this project. In simpler terms - this project will work towards making use of GATE on the cloud more like electric sockets and fridges!

Title: *Rapport: Robust Application Porting for HPC in the Cloud.*

Investigators: Professor J Darlington (Imperial College London), Dr D Colling (Imperial College London), Dr SA Butcher (Imperial College London), Mr NP Chue Hong (University of Edinburgh)

Summary: We propose to examine the feasibility of Cloud Computing for Physics and Bionformatics e-Science by porting a set of scientific applications in these domains to a variety of Cloud platforms. This work will be carried out in conjunction with application scientists in Bioinformatics and Physics. The applications that have been selected for piloting on Cloud platforms have been chosen because they present challenges in data volume, processing requirements or close-coupled parallel execution that are typical of most scientific HPC software. Project outputs will be a study of HPC Cloud feasibility which will compare performance of the selected software on a variety of Cloud platforms, seek to determine best practice for Cloud deployment of HPC software both generally and in the chosen domains, and make general recommendations about the use of Clouds for scientific HPC software within an institutional context.

Title: *Elastic Virtual Infrastructure for Research Applications*

Investigators: Professor I Sommerville (University of St Andrews), Dr A Barker (University of St Andrews)

Summary: This proposal is focused on enabling researchers to simply and rapidly deploy, execute and monitor scientific software on elastic cloud computing infrastructures. Current interfaces to cloud resources are relatively low level and do not allow researchers to easily benefit from the elasticity that cloud infrastructures offer. Researchers have to deal with time-consuming and often error-prone tasks such as managing access credentials, selecting instance types, managing elastic IP addresses, as well as monitoring resource usage and starting, stopping and terminating instances in response; this keeps researchers from focusing directly on their scientific research. In order to address this problem and to further the uptake of cloud computing services in research we will develop an elastic wrapper for scientific applications. The elastic wrapper will provide an abstracted gateway to cloud resources and will provide a one-stop-shop interface for researchers wanting to take advantage of cloud resources for their scientific research. It will abstract the complexities of setting up, configuring and managing cloud resources for scientific research applications and provide facilities for execution and collaboration between multiple research sites working on the same problem. The system will take care of issues such as managing resource usage using the elasticity of cloud resources as well as fault tolerance to insure against resource failure. This project will provide a pilot implementation of the elastic wrapper that will be a generic solution but specifically support two exemplar scientific applications and their usage models: Groups, Algorithms, and Programming (GAP), a free, open source system for discrete computational algebra with an emphasis on computational group theory and IDL is a commercial package for statistical and numerical analysis and visualization of scientific datasets.

Title: *myTrustedCloud: Towards a virtual private cloud*

Investigators: Dr D Wallom (University of Oxford), Dr A Martin (University of Oxford), Dr G Taylor (Brunel University)

Summary: The current profusion of cloud infrastructures, built both within the public but also private space have enabled a significant body of research to move their computational requirements into this new paradigm. There are though a collection of users that are not able to make use of this new paradigm though it is clear that this would improve the provision of computational and data resources available to them. With the pilot energy applications that we are going to use, looking at the provision of Advanced Metering Infrastructure, Condition Monitoring and Distributed State Estimation we will prove that the utilisation of hardware trust within the system for attestation of state and identification of both the data and algorithms and their hosting virtual instances would mean that this high value critically important system could utilise cloud computing. Specific outputs will also include a detailed threat analysis of using IaaS cloud systems and the specific countermeasures that trusted platform allow within the system, an exemplar software framework in which energy researchers are able to start making use of commercially sensitive information while at the same time make full use of cloud computing. This framework will be documented such that other research areas will also be able to make use of the findings within the project in such sensitive areas as medical and social research, both whom have traditionally tight restrictions on data.

Title: *Fair Benchmarking for Cloud Computing Systems*

Investigators: Dr L Gilliam (University of Surrey), Professor M Baker (University of Reading)

Summary: Compute resource benchmarks are an established part of the high performance computing (HPC) research computing landscape, and are also in general evidence in non-HPC settings. As Cloud Computing technology becomes more widely adopted, there will be increasing need for well-understood benchmarks that offer fair evaluation of such generic systems in comparison other kinds of computing systems that have been optimized for specific purposes. Cloud Computing benchmarks need to be able to account for all parts of the lifecycle of cloud system use., and most existing benchmarks do not allow for this. Cloud-specific benchmarks will increase in importance because clouds have a wider range of possible applications than are offered by HPC, and also because the variety of options and configurations of cloud systems, and efforts needed to get to the point at which traditional benchmarks can be run, have various effects on the fairness of the comparison. In this pilot, we set out to create an academically focused cloud benchmark site that accounts fairly for such variations. The principal outcome will be a web portal that embodies such considerations and which can be used to access data about benchmark runs, and potentially to adapt benchmarks to run on other Cloud systems. The proposed portal will offer a service to a knowledgeable user that returns the closest matches, based on the closest portfolio of benchmark elements, to a set of requirements specified about their own application as a Service Level Agreement (SLA). The portal will also offer access to bundled benchmark tests (virtual machines containing such applications) that have been constructed during the project, and which will alleviate the need for other researchers to repeat such work and the associated costs of Cloud as well as of effort, in doing so.

Title: *CloudBIM: Exploring the Feasibility and Potential for Cloud Research in the Architecture, Engineering and Construction Sector*

Investigators: Professor Y Rezgui (Cardiff University), Professor O Rana (Cardiff University)

Summary: The AEC (Architecture, Engineering and Construction) industry is a highly fragmented data intensive project-based industry depending on a large number of very different professions and firms, with strong data sharing requirement across lifecycle stages from concept design to demolition. The process of designing, re-purposing, constructing and operating a building involves not only the traditional disciplines (Structure, Mechanical & Electrical, etc.) but also many new professions in areas such as energy, environment, waste, and assisted living with large data sharing requirements. In this context, data management support for the project lifecycle tends to be fragmented with a lack of an overall (project wide) data management policy. Additionally, data sets relating to a particular project can often be stored in: (i) local computers of designers/architects - often with limited network connectivity, persistence and availability; (ii) independently managed, single company-owned archives - where access is dictated by a company specific policy or by a charging model; (iii) shared archives owned by a consortium, often in the context of a particular building project - based, at best, on access policy associated with the project. The CloudBIM proposal explores the feasibility and potential for utilizing Cloud capability to address data storage and processing needs of stakeholders in the AEC (Architecture, Engineering and Construction) sector, with a view of delivering a cloud platform for research. CloudBIM will involve close consultation and interaction with major participants in the area to assess stakeholders' perceptions about outsourced, virtualized Cloud storage for supporting multi-site, multi-team collaborative projects. A prototype cloud platform (based on CometCloud - www.cometcloud.org) and associated governance model will be developed and made available to the AEC research community. The project will deliver several reports based on a number of people-based activities, involving BRE (Building Research Establishment) and MBEKTN (Modern Built Environment Knowledge Transfer Network), along with a prototype using real project case studies BIM (Building Information Model) data to be provided by Bentley. A key outcome will be to spur a wide range of research-oriented activities through a strategic roadmap aimed at the exploitation of the resulting CloudBIM platform.

Title: *My Private Cloud*

Investigators: Professor DW Chadwick (University of Kent)

Summary: This research is designed to increase - a) the trust that users may have in cloud providers, as well as - b) the control that users will have over their data when it is stored in the cloud. It aims to do this in a number of ways. Firstly existing cloud users can provide their feedback about their existing cloud service providers to a cloud reputation service. This will compute the reputations of the various cloud providers, so that new potential cloud users can query it in order to determine which cloud providers are the most reputable. Then, when a user has chosen a cloud provider that (s)he believes to be trustworthy, the user can set their own fine grained privacy policy on the data that they submit to the cloud. This policy will be stuck to their data so that it is always enforced by the cloud infrastructure. In this way the user has full control over all accesses to and processing of their (possibly very sensitive) data. If their data is moved between cloud providers, then the sticky policy will move with the data, thereby ensuring continuing control by their policy. The privacy protecting infrastructure has built in audit support to allow the cloud provider to send the user summary audit information which will detail who has accessed the user's data, at what time and for what purposes. This provides users with visibility into the cloud, and reassures them that their data is safe. Users may alter their privacy policy at any time, should they decide it is too strict or too lax. Finally, users will be able to delegate access to their data to other users or processes, in order to provide the flexibility that is sometimes needed in workflows and other data access scenarios. The fine grained privacy policies and protocols that are supported by the infrastructure allow requestors to collect their various attributes and roles from multiple issuing authorities (a process termed attribute aggregation), even when they are known by different identities at the different authorities. This mirrors the reality of today's plastic card credentials and allows a new generation of virtual cards to be created. The cloud provider is cryptographically assured that all these different attributes and roles do indeed belong to the same requestor, without the requestor being required to reveal his real name. The fine grained policies also support emergency over-rides, so called Break-The-Glass policies. These allow responsible requestors, who are initially denied access to the data in the cloud, to break the glass and be granted emergency access, in the full knowledge that they will be held accountable and have to answer to their line management at a later time. This is achieved by having an obligation service that can perform pre-defined actions when an authorization decision is made. In the case of break the glass, these obligations might be to email the requestor's line manager, and record the incident in a secure audit trail. One example of Break the glass use is in medical applications, e.g. it allows accident and emergency staff to access a patient's medical records that they otherwise would not be allowed to see.

Title: *Optimal Scheduling of Scientific Application Workflows for Cloud-augmented Grid Infrastructures*

Investigators: Professor S Winter (University of Westminster), Mr T Kiss (University of Westminster), Professor P Kacsuk (University of Westminster), Dr GZ Terstyanszky (University of Westminster), Dr C Reynolds (University of Westminster)

Summary: Research scientists need cloud computing to flexibly support their computational requirements but also require user-friendly tools in order to engage. Demanding computational requirements for science research have been addressed through grid computing, which until recently has been based on a fixed physical wide-area infrastructure. Physical grid resources can become overloaded however - scientists may find this restricts their needs where the performance required at a particular moment in time cannot be addressed in real-time. User-friendly tools and environments are required to support grid computing in science communities. P-GRADE is a web-based portal, co-developed and by the investigators, for designing, submitting and monitoring workflows on the grid. It is highly graphical, and thus very appealing to a wide range of non-computing specialists. Familiarity is important to users. In order to develop the potential for cloud computing for the same community of science researchers, it is highly desirable, from their perspective, to retain continuity of computational support environment, as much as possible. The P-GRADE environment has been influenced by user requirements and aspirations over many years, and has already achieved a high degree of acceptance within the research science community, in view of its usability. In any case, to redesign a user interface from scratch is a very costly development activity. Interoperability between service grids and cluster grids was achieved by the investigators in the EU FP7-funded EDGeS project, so that users of each community may now access the resources of the other. Grid computing can also be extended onto cloud resources. The investigators have recently developed solutions for extending both service and cluster grids with virtual cloud resources. The usability of the tools and environment developed for the grid has been experimentally evaluated by the investigators in collaboration with biological research groups at the University of Westminster and Imperial College. These collaborative experiments have been very successful, in that the research biologists have found them very acceptable working environments, and have actively adopted them for their live research programmes. The collaborations have also confirmed that the computational performance of the grid solutions was significantly enhanced. The combined effect of a high degree of usability and enhanced computational performance has been to facilitate significant shifts in biological experimental methodology, leading to increased research productivity. Nevertheless, there are limits to the performance achievable on fixed grid implemented on physical infrastructure, due to the physical limitations of the infrastructure itself. Cloudbursting, i.e. the ability to provide additional compute capacity that may be required on demand at specific times to cope with unpredictable peaks of research computing, is an attractive aspect of cloud computing which has the potential to break through the current performance constraints achievable on fixed grids. It is proposed to attach a cloud infrastructure to existing fixed grids to achieve a mixed grid infrastructure comprising both existing physical and virtualised cloud resources. It is then proposed to re-engineer the tools and environment to exploit cloudbursting in the mixed grid to achieve even greater performance. Finally, the performance of the mixed physical/virtual grid will be evaluated, to include statistics on resource utilisation, to inform the development of a costed performance model.

Title: *Clouds in Space*

Investigators: Professor SJ Cox (University of Southampton), Dr HG Lewis (University of Southampton), Dr K Takeda (University of Southampton)

Summary: Within the last two decades, space-based technology has become a ubiquitous component of everyday life. For example, British Sky Broadcasting (BSkyB) produces subscription television services for over 12 million direct and indirect customers in the UK and Ireland, and 40 million people daily use the Dutch TomTom company's solutions which provide in-car navigation systems and tracking systems for fleet management. Emergency services rely extensively on satellite technology. In this research we will investigate the applicability of Cloud Computing and data handling for the important international problem of Space Situational Awareness (SSA) and Space Debris removal and mitigation. This is an important theme area within the European Space Agency, which protects Europe's citizens and satellite services by detecting space hazards. One of the key goals recently set by Obama was to Strengthen stability in space through: ...; improved information collection and sharing for space object collision avoidance; protection of critical space systems and supporting infrastructures, with special attention to the critical interdependence of space and information systems; and strengthening measures to mitigate orbital debris. A highlighted area of interest was to Develop, maintain, and use space situational awareness (SSA) information from commercial, civil, and national security sources to detect, identify, and attribute actions in space that are contrary to responsible use and the long-term sustainability of the space environment. On Feb 10th the US Iridium-33 and the Russian Cosmos 2251 collided, resulting in debris which could place at risk many other active satellites in this low-earth orbit position. Currently we track around 20,000 objects larger than ~10cm in satellite orbits. In the future, with higher fidelity systems, we will be tracking 500,000+ objects of size 1cm+. This poses significant challenges to be able to scale up the compute resources and complex algorithms required to process the data which arrives twice-daily. We propose to investigate how Cloud computing can be used to tackle these challenges. Cloud computing is internet based computing which allows resources, software, data and services to be provided on demand. Many individuals and businesses use Cloud based services for email, web searching, photo sharing and social networking. Scientists and Engineers are using a similar paradigm to make use of massive amounts of compute and data handling resources provided by companies such as Amazon, Microsoft and Google. Specifically we will investigate the efficacy of the Cloud to develop and test algorithms to target debris for removal and understand the efficiency of the new algorithm - this is important to enable us to perform the modelling required as we track more objects. This is based on the Travelling Purchaser Problem, a variant of the widely used/solved Travelling Salesman Problem. We will investigate whether the Cloud can provide a scalable, reliable and robust infrastructure for the ongoing requirement to aggregate and process ever-increasing volumes of data to propagate orbits, detect events, and plan missions. Such mission planning is an important aspect of launching new satellites and removing existing debris from orbit. We have already built a prototype using Microsoft's Azure Cloud platform and this research will enable us to increase the efficiency of the calculations and improve their scalability. Understanding how the Cloud can be used in this area of science and engineering will also help shape how Cloud providers, such as Microsoft, will provision services in the future which can be used to perform research in disciplines as diverse as healthcare, environmental management, bioinformatics and energy production, which are important challenges for society as a whole.

Title: *ECHO - Enabling Cloud Hosted Organisations*

Investigators: Dr A Stewart (Queen's University of Belfast), Professor R Perrott (Queen's University of Belfast), Dr T Harmer (Queen's University of Belfast)

Summary: This project proposes an investigation into the use of software for the automated deployment and management of heterogeneous resources as utility resource clouds based on previously funded EPSRC activity. It offers automated multi-provider cloud management and will bring an establish field tested technology with commercial features such as cost and budget setting, service level agreements, robustness and resilience to researchers and research applications. The proposal focuses on answering two questions:1. Can the cloud utility model replace local compute clusters or groups of clusters and provide an efficient, extensible and customisable computation platform for the majority of the research community?2. Can the utility cloud model bring a more flexible and dynamic compute infrastructure to researchers and free them of the need to spend time and money supporting their infrastructure? The project focuses on developing the EPSRC investment in Zeel/i technology to manage application access and deployment to cloud providers-the work will focus on supporting research applications rather than the commercial applications that have been the focus for the last 5 years. The pilot project will focus on two representative research application areas that provide distinctly different challenges for cloud as a research computation platform and where PI/co-PI expertise is available. Exemplar 1 is a suite of numerical mathematical applications which relies heavily on BLAS libraries. This application suite is often run in small and medium scale to test algorithm implementation but is designed for very large scale production runs when matrix sizes of 100k by 100k are normal, and where efficient BLAS implementation, multicore resources and fast storage are essential to give satisfactory execution times. Exemplar 2 is a suite of applications to support gene analysis and relies on a several large application databases and a collection of online remote databases and analysis services. The plan of work focuses on the execution, management and analysis of cloud based implementations of the exemplar applications on commercial utility providers-these are judged to offer the most diverse cloud environments and offer distinct cost savings when compared to owned and managed infrastructures. This approach also enables researchers to access the technology and cost savings that the rapid adoption of cloud within the commercial domain is bringing.

Annex 2: Attendees List

Forename	Surname	Institution / Organisation
Simon	Cox	Southampton
Steven	Johnston	Southampton
Alan	Stewart	Queens University Belfast
Hamish	Cunningham	Sheffield
Valentine	Tablan	Sheffield
David	Chadwick	Kent
John	Darlington	Imperial
Jeremy	Cohen	Imperial
Stephen	Winter	Westminster
Yacine	Rezgui	Cardiff
Omer	Rana	Cardiff
Ian	Sommerville	St Andrews
Alex	Voss	St Andrews
Stephen	McGough	Newcastle
David	Wallom	Oxford
Lee	Gillam	Surrey
Paul	Watson	Newcastle
David	Salmon	JANET
Tony	Wheatman	Nottingham
Andy	Powell	Eduserve
John	Brooke	Manchester
Colin	Bannister	Nottingham
Neil	Chue Hong	Edinburgh
Rob	Hawtin	Curtis and Cartwright
Jim	Austin	York
John	Overington	EBI
Louise	Tillman	EPSRC
Torsten	Reimer	JISC
Matthew	Dovey	JISC

Annex 3: Workshop Agenda

Time	Activity
11.30	Registration: RCH001
12.00	Lunch and Poster Session: RCH001
1.00	Welcome and Introduction: RCH248/250 Louise Tillman and Torsten Reimer
1.30	<p>Lightening Talks: RCH248/250 3 minute talks from the Cloud Pilot PIs describing the key findings of their pilot study:</p> <ol style="list-style-type: none"> 1. Flood Modelling for Cities Using Cloud Computing: Stephen McGough 2. GATE Cloud Exploratory: Adapting the General Architecture for Text: Hamish Cunnigham 3. Rapport: Robust Application Porting for HPC in the Cloud: John Darlington 4. Elastic Virtual Infrastructure for Research Applications: Alex Voss 5. myTrustedCloud: Towards a virtual private cloud: David Wallom 6. CloudBIM: Exploring the Feasibility and Potential for Cloud Research in the Architecture, Engineering and Construction Sector: Yacine Rezugui 7. My Private Cloud: David Chadwick 8. Fair Benchmarking for Cloud Computing Systems: Lee Gillam 9. Optimal Scheduling of Scientific Application Workflows for Cloud-augmented Grid Infrastructures: Stephen Winter 10. Clouds in Space: Simon Cox 11. ECHO – Enabling Cloud Hosted Organisations: Alan Stewart
2.15	Break Out Session 1: RCH248/250 Lessons Learnt
3.00	Tea and Coffee
3.15	Cloud Cost Study Presentation: RCH248/250 Rob Hawtin (Curtis and Cartwright)
3.30	Break Out Session 2: RCH248/250 What Next?
4.30	Wrap Up and Close

Annex 4: Outputs from Lessons Learnt Session

“What are most important lessons that you have learnt as a result of the Cloud Computing Pilots”

“What advise would you give to a researcher looking to use the cloud based on your experience”

Group 1

- Separation of resource and AAA
- Understand your problem before you start – less information available from Cloud
- Costing is complicated
- Do not think Cloud is simple

Group 2

- Security
 - Need improved access control mechanics (not ID only)
 - Not standard based
 - O and D technology
 - not trust (internal security only)(researchers ↔ storage providers)
- Data
 - Big data
 - Bandwidth charges
 - Bandwidth
 - Need improved cost metering
- Changing business models
 - Keeping up to date
- Performance
 - Virtualisation
 - Masks underlying hardware (to some extent)
 - Issues for some HPC apps
- Brokering – more efficient use of resources
- Advice
 - Understand your application
 - Does Cloud work for you?
 - Data volumes
 - Security: risk analysis
 - Keep up to date
 - Software/services change frequently

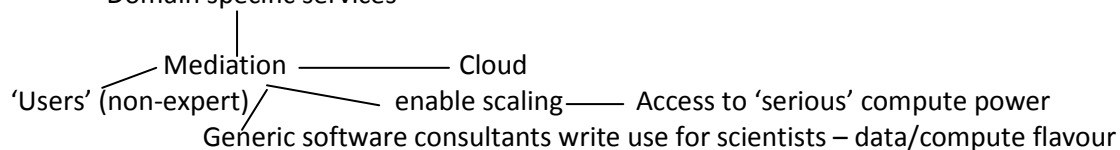
Group 3

Advice: Know resource

- Know programme environment
- Know application
- Know lifecycle

Lessons Learnt: Know lifecycle of application: 1of L → long life?

- Domain specific services



Lessons Learnt:

- Application
- 'Generic middleware functions'
- H/W

Advice:

Generic software consultants write use for scientists – data/compute flavour

Advice:

- Look at bill
 - Meaningful
 - Accurate
 - Current?
 - Itemisation
 - Accounting
- Bypassing ICT services
 - Reducing threshold

Lessons Learnt:

- Appear to be genuine use cases
 - Citizen access

Group 4

Issues:

- Make sure that Cloud has the features your application needs
- Needs to organise governance of the data
- Need to clarify legal issues, e.g. privacy/location with Cloud provider

Advice:

- Find out cost implications – can you claim back e.g. via expenses. Try incrementally . don't put all eggs in one basket
- Understand your requirements. How much can be met by local resource? How much needs outsource and Cloud
- Put pressure for uniform interfaces/users need to act together

Group 5

1. Understand what you need and why:
 - Application specific info
 - Brokerage
2. Do users understand application costs and resource requirements
 - Benchmark and size
 - Who pays?
 - How?
 - Bulk buying
3. Talk to knowledgeable people
 - HPC/Cloud/General computing (who don't frighten the users)
4. Science limited by lack of understanding of the apparatus. Ambitions grow once this is addressed
5. Think about and explore the range of possibilities – justify need of Cloud to yourself first
 - Cost remains ill-understood difficult to justify
6. Engagement with users is vital
7. CLIs are bad (for some- barrier to use)

Top 3 pieces of advice from each group:

ADVICE:

- Need generic software consultants to work with scientists and engineers:
 - Data expertise
 - Compute expertise
- Know yourself:
 - Understand your application its needs, profile and =possible life cycle
 - Understand what is being offered, its capabilities, costs, saleability, what offers what
- Understand your problem before you start
- Look at your bill:
 - Presumes that a bill exists!
 - Itemisation
 - Accounting – who/what
 - Meaningful
 - Accurate
 - Current
- Find all cost implications before you commit larger resources. Use an incremental approach
- Do not think Cloud is simple!
- Understand your needs:
 - How much can be met from local resource
 - How much really needs external resource
- Talk to knowledgeable people:
 - HPC/Cloud? general compute
 - And look for those who understand users
- Thin about and explore the range of possibilities – justify need of Cloud to yourself first
- Work together with other users to apply pressure for uniform interfaces and APIs
- Understand your application and requirements:
 - CPU time
 - Data volumes
- Understand what you need an why:
 - Need application specific information so that brokerage and cost models can be clarified
- Keep up to date:
 - Regular changes to features, pricing models and available services
- Perform a risk analysis (business benefits vs. potential losses)

LESSONS LEARNED:

- Decoupled:
 - Application
 - Generic middle ware functions
 - H/W
- Make sure the Cloud has the features your application needs
- Need to decide on and organise governance of data in the cloud
- Appear to be genuine use cases for Cloud:
 - Accessible to scientists
 - Citizens
 - Users
 - Cost effective
- Engagement with users is vital:

- Choose appropriate UI style some users will not use CLI, others require it
 - Make sure they understand costs, application requirements
- Understand your problem before you start
- Need domain specific services:
 - Non expert users – mediation – cloud resources – access to serious compute power
- Costing is complicated:
 - Variability in charges
 - Variability in performance
 - Lack of information about ‘physical’ attributes
- Science limited by lack of understanding of the apparatus:
 - Ambitions may grow with understanding
- ‘Interesting’ separation of resources and AAA – the credit card issue
- Understand cost model for data storage and transfer:
 - Need for metering/dashboard
- Clarify legal issues e.g. privacy/location of data with Cloud provider
- Monitor application performance
- Poor security:
 - 20 year old technology (un/pw and ACLs)
 - Non standards based
 - Need FIM and attribute based access controls
 - No trust (CSPs don’t trust external services researchers don’t trust cloud providers)
- Cost remains ill understood so difficult to justify
- Benchmarking and monitoring required

Annex 5: Outputs from What Next? Session

From your experience what conditions do you think need to be met for a wider uptake of cloud computing in research in terms of?

- *The technical conditions to enable wider uptake*
- *The conditions for institutional uptake*
- *The conditions for individual researchers/research group uptake*

What specific actions are required to support this?

- *Funder actions e.g. what do EPSRC, RCUK and JISC need to do?*
- *Institutional actions*
- *Researcher and community actions*
- *Industrial actions*

Group 1:

1. Simple standards/APIs:
 - Links with popular environments
 2. Trust/legal service provider:
 - Institution
 - User
- Cost:
- Predictability
 - Defined business models
- Learning curve:
- Education
 - Funding

Funders:

- Funding models for applications and tools and techniques
- No more shotgun weddings

Institutions:

- FEC their IT provisions
- Undertake risk analysis

Researchers:

- Try it out before making commitment...
- Understand application requirements and trade offs between optimisation and costs

Industry:

- Make costs transparent
- Institution for friendly billing
- Provide Cloud middleware to help users
- Provide SLAs

Group 2:

Technical:

- Still difficult to do multi cloud:
 - Commission examples (funders), disseminate results (community), good for classic early adaptors

- Heterogeneity is a problem to understand performance, even if you have e.g. an m1 instance
- There's only a single 'queue' model – better quotas/cut-offs? How to get this in environment?

Institutional:

- Confidence in billing/accounting/authorising/authorising caps
- Local points of authority:
 - Champions – train campus champions/ community SSI/ champions NGs
 - Institutional 'how tos' – 'come and try our cloud'/roadshows
- Better 'sandboxing' – for learning/teaching – credit card responsibilities
 - How to get this into environment?

Research:

- Getting the right knowledge to the people already doing application support (the 'computing' guy in the group) – show that funders accept CC
- Knowing whether the Cloud will be 'worth it': Same as top, good and bad examples
- Better 'sandboxing' – for learning/teaching – credit card responsibilities
 - How to get this into environment?

Group 3:

Top 4:

1. Collaboration industry/researchers in standard bodies - RCs fund participation
2. Need to change research culture
 - Incentives/penalties/training
3. Develop community action at European level
4. Clear comparable business models, joint industry/academia discussion

What Next:

Funders:

- Changes in rules for equipment funding
- Training/guidelines for referees

Institutions:

- Rules for sharing resource/default and steer
- Credit card access integrated into research support
- Training in E-infrastructure

Researchers/Community:

- Develop community/brokerage and advice
- Engage in European effort

Industrial:

- Implement standards
- Engage in standards body
- Clear business model for academia – also a responsibility of academia

Group 4:

Technical Conditions:

- Improved range of resource/services
- Billing systems
 - Usage caps
 - Central accounting

- Long term persistence of data
- Improved bandwidth for external data transfers

Institutional uptake:

- Interoperability
 - Between local resources and Clouds
 - Between different Clouds
- Local support
- Business case
- Lower total cost of ownership
- Billing

Researchers/Group uptake:

- Access to faster/more powerful machines
- Availability of skills/training
- Billing

Funders:

- Support requests for pay-per-use computing as alternative to capital expenditure
- Clear guidance on when different purchase models are acceptable

Institutions:

- Clear internal charging model for access to IaaS services
- Transparent usage monitoring
- Volunteer/community computing:
 - Make effective use of local cycles (institutional policies)
 - Research led – contribute locally managed resource
- Improved data management/usage policies

Researcher/Community:

- Provide example use cases
- Provide reference models for Cloud use
- Research led – contribute locally managed resource

Industry:

- Improved engagement with user communities:
 - Improved services
 - More flexible billing
 - Reduce lock-ins

Group 5:

Technical Conditions:

Common interface:

- Not necessarily a standard?
- Data of networking

Institutional Uptake:

- (Seamless) cloud bursting service
- Common interface
- Brokerage

- Charge back
- Hedging and insurance
- Monitoring
- (management) understanding
- Bridge the gap between management and researchers
- Green league table

Researcher and Research Group Uptake:

- Awareness
- Charges and costs
- Hand holding
- Low barrier to access

General actions:

- Raise awareness

Industrial actions:

- Academic discounts
- Exchange market for providers
- Exchange rate mechanism?
- Not just AWS/Azure/RACHspace etc and existing DCs *reengaged*

Funder actions:

- Offer more money
- Fund more Cloud projects
- Cloud apps: cloud equivalent applications = demonstrators
- Market place ecosystem?

Industrial actions:

- Create private clouds/services
- Sell of spare capacity
- Federations (of federations)
- Fund (and promote) best approaches

Researcher/Community actions:

- Feedback

2.2 Top 4 actions for: Funders, Institutions, Researcher and community and Industrial

FUNDER

- Improve guidance on support for/use of Pay-Per-Use computing as an alternative to capital expenditure
- Change researcher culture – incentives/penalties/training
- Funding for application work using Cloud experts in domains
- Funding for tools and techniques research on Cloud (informed by users)
- Not ‘shot gun’ wedding funding
- Give **us** more money

INDUSTRIAL

- Build a better sand box – but should this be done in a co-ordinated community fashion?
Is this JISC/UMF?
 - One
 - Two
 - Many
- Collaboration of industry/researchers in SDO:
 - RC fund participation STDs matter
- Clear comparable business models joint industry/academia discussion
- Improved engagement with user communities:
 - Improved services
 - More flexible billing
 - Reduced lock-in
- Make prices transparent
- Institution-friendly billing
- Provide better Cloud middleware
- Provide proper SLA
- Academic discounts

RESEARCHER

- Provide examples use cases and reference models for Cloud use
- Develop community action at EU level c.f. EGI/PRACE
- Try out the Cloud before making commitment
- Understand application requirements
- Mandatory to publish raw data and algorithms
- How do I know if it's 'worth it'?
 - Commission exemplars good and bad
 - Disseminate at community and institution
 - Incentives for helping/sharing SSI/NGS 'champions'
- Focus on science

INSTITUTION

- Clear internal model for use and access of IaaS services
- Promote use of such services
- FEC their IT provisions
- Undertake proper risk analysis
- Show that costing preconceptions are not well founded
- Cloud vs. but a server
- Commission cost and performance studies
- Get institutions to publish their costs for services
- Institutional buy-in:
 - From the top ('I think we should use this')
 - On the ground ('institutional how tos' roadshows)
 - But who pushes?
- Embrace the technology – understand it first please!