Practical Cloud Evaluation from a Nordic eScience User Perspective

Åke Edlund
PDC, KTH Royal Institute of Technology
SE-100 44, Stockholm, Sweden
edlund@pdc.kth.se

Maarten Koopmans
vrijheid.net
1094 KC Amsterdam, Netherlands
maarten@vrijheid.net

ABSTRACT

In this paper, we describe the findings of the NEON project – a cross-Nordic – Sweden, Norway, Denmark, Finland and Iceland - project evaluating the usefulness of private versus public cloud services for HPC users. Our findings are briefly that private cloud technology is not mature enough yet to provide a transparent user experience. It is expected that this will be the case mid 2012. The cost efficiency of both public and private cloud should be continuously monitored as there is a strong downward trend. This conclusion is supported by NEON experimenting as well as larger initiatives e.g. StratusLab reports. Public cloud technology is mature enough but lacks certain features that will be necessary to include cloud resources in a transparent manner in a national infrastructure like the Norwegian NOTUR (www.notur.no) case, e.g. with respect to quota management. These features are expected to emerge in 2011 via third party management software and in the best of breed public cloud services. Public clouds are competitive in the low end for non-HPC jobs (low memory, low number of cores) on price. A significant fraction (ca. 20%) of the jobs running on the current Nordic supercomputer infrastructure are potentially suitable for cloud-like technology. This holds in particular for single-threaded or single-node jobs with small/medium memory requirements and non-intensive I/O. There is a backlog of real supercomputer jobs that suffers from the non-HPC jobs on the supercomputer infrastructure. Off-loading these non-HPC jobs to a public cloud would effectively add supercomputing capacity. Another finding is that available storage capacity is not accessible in a user-friendly way; most storage clouds are only accessible via programmable interfaces. A number of experiments and piloting are presented to support these claims.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; C.2 [**Computer-Communication Networks**]: General, Distributed Systems; K.6 [**Management of Computing and Information Systems**]: Miscellaneous; D.2.8 [**Software Engineering**]: Metrics—*complexity measures, performance measures*

General Terms

Management, Measurement, Performance, Design, Economics, Reliability, Experimentation, Security, Human Factors, Standardization, Legal Aspects, Verification

Keywords

Cloud computing, experimenting, evaluation of performance cost, usability

1. INTRODUCTION

One of the main goals with this project was to report on the state-of-the-art including the HPC for cloud scenarios, with pilots and recommendations focusing on non-HPC cloud scenarios. This is also the most urgent issue for today's Nordic HPC centers – that non-HPC jobs are taking considerable shares of the runtime on HPC resources. The article is organizes as follows: describing the evaluation work, followed by a discussion on the cost and risks when using cloud, ending with a longer conclusion section including short-term and long-term recommendations.

2. HOW TO MOVE NON-HPC JOBS TO A CLOUD COMPUTING ENVIRONMENT

This section will list the scenarios for moving non-HPC jobs off the supercomputing infrastructure to a cloud-like environment. The scenarios are:

- Existing private cloud solutions in an "Enterprise" (licensed) variant
- Open source private cloud solutions
- Local solutions, completely do-it-yourself
- Public clouds

2.1 Existing private cloud solutions using licensed software like Enterprise Eucalyptus/ OpenNebula, together with management software such as RightScale

Though this solution is feature-rich it is also complex and expensive. Extra features are things like quota management, cloud portability and policies on the usage of cloud services per end user. Quota management is often done by setting a virtual price on the local infrastructure and then assigning a budget to a user or a project. Quotas are ”soft” quotas, meaning that the user does not get shut down but rather that the administrator gets signaled on reaching a certain usage threshold. Note that via this pricing mechanism one budget can mix and match public and private clouds. Cloud portability is implemented by adding a meta package system on top of ”identical” virtual machines in every cloud. The user then creates a virtual machine by adding packages to the base image in a web based application. Additionally, multiple virtual machines may be bundled in a deployment. Such a deployment can then be launched on a cloud of the user’s choosing. Also, administrators may provide rich template virtual machines – we have seen examples from computing instances to Wikimedia services. This concept is often referred to as ”IT vending machine”. However, setting up a private cloud has a steep organizational learning curve. Doing this in such a way that it integrates with cloud management software and fully utilizes the added benefits of the management software will add man-years to the initial set up compared with plain old hardware. And as private clouds are rapidly evolving with at east two major releases per year, intensive update cycles will keep this labour-intensive, requiring an organization to acquire new specialistic knowledge of fleeting nature. We expect as Enterprise private clouds mature more and more the need for separate management software will decrease and the required effort to keep a private cloud running will go down. This will be at least another 18 months though.

2.2 Open source alternatives of Eucalyptus and OpenNebula

We have looked in depth at two leading open source private clouds offerings, OpenEucalyptus and OpenNebula. Both solutions are open source, though Eucalyptus also has an Enterprise variant. During the project a third solution became open source (CloudStack) and started rising, specifically in the US realm. We have deployments of OpenEucalyptus and OpenNebula in Finland and Sweden, with the Swedish one (RedCloud) still continuing. There is a strong tendency towards API compatibility with the Amazon public cloud, recognizing that Amazon currently has set the gold standard. Eucalyptus differs in that it also offers storage, OpenNebula only offers computing at the time of writing. In using private clouds we have gained a number of insights:

- Private clouds require above average control of the network topology. As private clouds effectively provision virtual machines to remote users, the network infrastructure needs to support the cloud model that has been chosen, or vice versa. This requires system and network administrators to closely work together. In practice, this adds significant calendar time in initial set up.
- Private clouds need ”current” hardware. As the older hardware does not support virtualization we have had e.g. to resolve this using Xen’s paravirtualization. Xen provides an efficient virtualization platform with low virtualization overhead. Its shortcoming is that it requires modifications to virtual machine’s kernel and drivers and is hence difficult or impossible to apply to non Linux operating systems like Microsoft Windows. In other words private clouds function best with hardware support for virtualization.
- Private cloud stacks are not as complete as public cloud offerings. This can be seen quite easily when comparing features of the cloud offerings. Where Amazon offers queueing services, relational database services, load balancers, VPNs (Virtual Private Networks) to a ”private cloud within the public cloud”, most private clouds offer only computing. Eucalyptus offers storage as well, but it can be considered sub-par compared to public providers. This implies that significant effort is required to offer a full scale cloud ”experience” with private clouds.

Note that the above three points also hold for the Enterprise variants, though the first two points will most likely be solved by the included consultancy in an Enterprise offering. Additional services and middleware (the third point) are often the value-added services of the Enterprise stack, but at significant cost compared to a public cloud. Finally, review by our peers has revealed that long-running larger private clouds with Eucalyptus were deemed fragile, were OpenNebula was considered more stable, but lacking storage.

2.3 Local solutions

Computational high energy physics (HEP) requires very large, non-standard software installations. This last requirement makes it difficult to use all potentially available resources. A DIY (Do-It-Yourself) solution has been implemented on a 24-core farm at CERN: A small library of virtual machine images was put in place and used for high energy physics applications and medical image analysis. For the case of HEP, the images were provided by the CernVM project, for medical image analysis, a minimal Debian image was used. The main lesson learnt here is that the current state of private clouds is such that for small to medium sized problems, a DIY solution might provide significant benefits in terms of time, money and effort spent while providing an end user solution that is more tailored to the needs of the user base.

2.4 Public cloud solutions

Public clouds seem to be the most stable and feature rich offerings – the biggest downside is their price, as some components of the pricing structure appear free for NGIs (National Grid Initiatives) and NRENs (National Research and Education Network), e.g. bandwidth. The biggest advantages of the public clouds are that they have become so accessible that an average system administrator does not need to build a whole lot of new expertise. It is wise to have system administrators build custom virtual machines on the public cloud to minimize support efforts. Another additional benefit of public clouds is the community around a cloud – for end users and system administrators alike. Within one domain/brand/flavor of public cloud access management can often be implemented, though not full integrated with existing federations. Quotas are not directly implemented, so monitoring budget is a manual operation unless cloud management software is used. But as soon as one chooses one specific public cloud (as is often the outcome from a tender

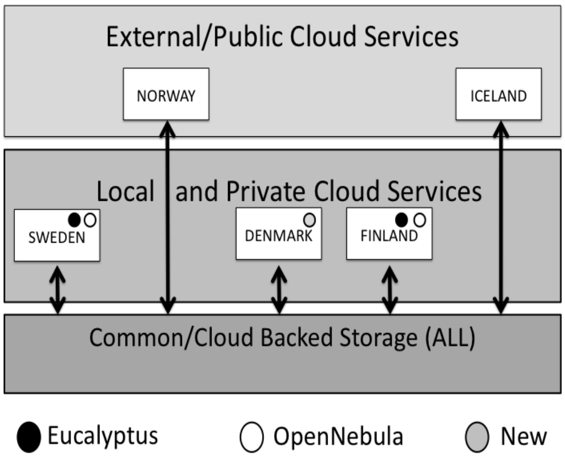

Figure 1: Big picture of Nordic cloud pilots

procedure) most other benefits of cloud management software might be too expensive for what they offer. Finally, public clouds offer instant availability – and thus low turnaround times. This is an excellent feature for organizations starting with cloud computing; it allow for setting up the optimal organizational structure and work on basic cloud knowledge without having to acquire lots of highly specialistic knowledge up front.

3. TESTS, PILOT IMPLEMENTATIONS, GAP ANALYSIS

This section will give a brief finding on the tests that have been performed and which tools have been used. We'll also provided a gap analysis concerning the currently available private and public clouds and their management software on the one hand, and the desired and needed functionality on the other hand.

1. Shortlists public and private cloud infrastructure providers, as well as a shortlist cloud management software. *Rationale: the goal was to have three shortlists that "work well together", so that a) hybrid clouds and b) cloud management becomes realistically testable.*

Public clouds: Amazon, Rackspace
Private clouds: OpenEucalyptus, OpenNebula
Management software: RightScale

These shortlists reflect the most mature products when the project started in early 2010.

2. Shortlist pilot applications: At least one computationally oriented and at least one storage oriented application. *Rationale: it was expected that besides computationally intensive also storage based applications will benefit from the cloud.*

In Norway two pilot applications where launched on the public cloud:

– **Cloud backed storage:** can we offer storage that is cloud backed as if it is a normal disk partition, completely metered, encrypted and elastically scalable? This pilot currently runs successful in the Amazon cloud.

– **eSysbio (www.esysbio.org) pilot:** can we run part of the eSysbio project in the public cloud? eSysbio aims to develop an e-science environment for supporting systems biology research – and use it to drive Norwegian research within this field. It will conduct research on Web services and service-oriented architecture (SOA), and use the results to build a collaborative virtual workspace that will facilitate the interdisciplinary exploitation of data, tools and computational resources relevant for systems biology research.

– **Cloud deployments:** At the University of Iceland, a pilot case study has been performed to study issues related to software architecture and clouds. Support for changes of the run-time and deployment-time architecture has been investigated. Instead of needing to, e.g., manually fire up new Virtual Machines (VM) and manually start services on them, a cloud provider independent scripting language has been applied to automate deployment of applications in the Cloud.

– **Private cloud infrastructure pilots:** what is the current state of private clouds and can we run pilots on them? To this end we have set up OpenEucalyptus and OpenNebula in Sweden and Finland. Based on the maturity level of these products we deemed it impossible to do a user-based pilot in those environments within the scope and manpower of the NEON project.

3. Choose and set up a public cloud infrastructure for a pilot application. *Note: the goal was to do this on local sites and then make this work together with other sites and the management software.*

Public clouds have, after initial tests of private clouds, been set up by Norway and Iceland.

4. Choose and set up a private cloud infrastructure for a pilot application. *Note: extra attention will be given to multiple-site management.*

This has been done in Finland and Sweden; Sweden has also set up a private cloud with multiple availability zones.

5. Choose and set up management software for cloud infrastructure. *Rationale: the management software will be crucial in provisioning, metering and in migration between clouds (hybrid clouds).*

We have set up accounts at RightScale to test the management infrastructure. Though RightScale's feature set is impressive and covers most needs for (very) large deployments, discussions about the cost of Rightscale deemed it too expensive compared to spending on the rest of the cloud infrastructure, especially in an initial phase. As part of a large scale tender process RightScale and similar services could be expected to play a role in offerings with private/public clouds. Among the features of RightScale are access management, quota management, a cloud component repository, attaching a virtual price on private cloud usage and a system for setting up virtual images in a cloud-independent manner.

6. Feasibility of multi-domain support (availability zones) for private clouds. *Rationale: one could consider a NGI as an "availability zone" – here we have learnt whether this is a useful comparison and how to implement cross-domain private clouds using the management software*

This has been implemented by Sweden with two sites locally, and a possible third German site joining later. Although it is certainly doable, tight coupling with network provisioning that will differ across borders make this impracticable. This is due to the fact that as soon as the network provisioning has been chosen on the main site, all other sites must follow the same model. If network provisioning is the same, adding multiple zones is straightforward though. Thus feasibility of multiple domains is largely organizational.

7. Gap analysis on cloud offerings versus needs (qualitatively) in user base, based on pilot experiences. *Rationale: when doing pilots, user feedback will give a qualitatively oriented feedback on using HPC in a cloud for private and public clouds, as well as the management software. This guarantees that cost is not the only metric.*

Based on the eSysbio pilot, the cloud backed storage pilot, and the pilots that have set up private cloud infrastructures we have concluded that the current public cloud offerings are superior to private cloud offerings. As needs we simply consider mimicking the current state of affairs on HPC computers, i.e. providing a similar user experience in a transparent way. Resulting gaps:

Gaps private clouds:

- No mature storage offering – storage offerings are either no part of private clouds, or perform sub-par.
- No other middleware offerings out of the box – services like databases, queues etc. must be deployed and maintained by the providing site.
- Heavy reliance on system and network administrator expertise for end users – essentially the downside for users when having full control.
- Separate management software – either you pay for an Enterprise version, or you pay for separate management software/services.
- Identity management integration limited to users uploading their own certificates.
- Rapid update cycle of core infrastructure software.
- Quota management hard or non-existent.

Gaps public clouds:

- More reliance on system administrator expertise for end users – same downside as above, though network knowledge might be less of an issue.
- Separate management software – this depends on the particular cloud solution and the level of management required. Obviously, public cloud offering offer some level of management if only for billing purposes. Public clouds develop themselves rapidly in this area, but not in a compatible way.
- Identity management integration limited to users uploading their own certificates.
- Quota management hard or non-existent – this is obviously not in the interest of public cloud providers. Why limit usage if that's what they sell?

On the plus side, from the pilots we have learned that the immediate availability speeds up project set up significantly. Also, the specialistic expertise at public cloud providers seems to be such that e.g. hardening of the infrastructure might be better than you could do it yourself. This is due to economy of scale.

8. Gap analysis on integration with existing AAI (Authentication and Authorization Infrastructure) for public and private cloud offerings. *Rationale: AAIs are well deployed, both locally and internationally (eduGAIN). Utilizing federated identity management should greatly foster adoption by easing access.*

This has turned out not to be an option. The best both public and private clouds currently do is allowing users to upload their own client certificates for starting/stopping services. This is after they have signed up in a custom identity management system. These systems tend to be closely coupled with the provisioning infrastructure, rendering AAI integration nearly impossible without a significant update from private and public cloud providers alike.

9. Gap analysis on integration with existing metering infrastructures. *Rationale: do the current metering infrastructures have what it takes and how do they compare to the metering built into clouds and cloud management software?*

Current cloud solutions provide APIs for monitoring usage, but do not offer quota management or cost control. Third-party management systems do provide soft quotas. Currently there exists a wide gap with respect to quota management and cost control for every cloud provider. Cloud management software bridges this gap, but adds an extra management infrastructure. We expect that the gap will close from three sides in the 2011-2012 timeframe:

- Commercial cloud management software will be able to manage "bare metal" more and more
- Cloud solutions will slowly add cost control and quota management features
- Open source cloud management solutions will thrive and add quota management and cost control

10. Architecture and cost analysis for integration with existing AAI infrastructure. *Rationale: once we know the gaps with the current AAIs the questions will be:*

a) can we integrate
b) if so, how?
c) at what cost?

This is not an option, further cost analysis on this point has proven useless until public and private cloud services enhance their identify management. It would result in an large rewrite of e.g. private cloud software or DIY solutions as per the Danish pilot.

11. Architecture and cost analysis for integration with existing metering infrastructure. *Rationale: see previous point.*

As clouds currently have no cost control only monitoring at best, there are multiple options. The first option is using a (hybrid) cloud management solution such as RightScale. These solutions provide charging of private clouds infrastructures as well, so costs for a project can span multiple cloud types. Also, these solutions provide soft quotas (signaling) which gives a weak kind of cost control. A second option would be to use the APIs provided by the cloud providers to continuously polling for usage information and then implementing a quota system yourself (Fig. 2). The cost of developing such a system is believed to be \$50,000-\$100,000 and the functionality will be basic. Maintenance (given the rapid cloud infrastructure developments) will be at least another \$50,000 on a yearly basis.

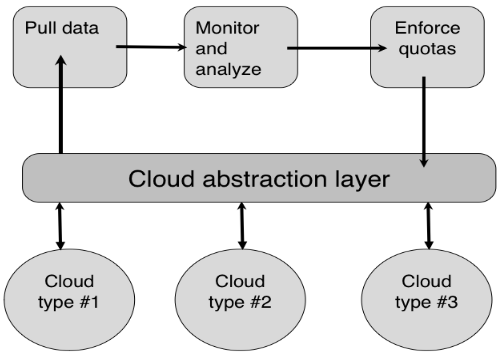

Figure 2: Cloud Abstraction Layer

Another option might be donating to an open source cloud management project. This will largely eliminate maintenance costs. The cost should be no higher than \$50,000 given the nature of open source projects. Finally, simply waiting while the market or open source community solves these problems is a valid option as well.

4. THE COST OF USING CLOUD COMPUTING

Economies of Scale and Flexibility in Use

The larger a data center is, the more value technologies such as virtualization and multi tenancy and efficient energy choices can bring. Very large data centers (≈ 50,000 servers), like the one below from Google in Oregon, can be up to 7 times [3] more efficient on administration, and 5 times (US numbers, [3]) more cost efficient on energy compared to mid-size data centers (≈ 1,000 servers). For smaller data centers these differences are even larger. Through this economy of scale big corporations – like Google, Amazon, Microsoft and Facebook – deliver compute and storage services on demand with a competitive price model. For customers using these cloud services a new flexible way of doing IT evolves – a service based economy – where the user only pays for what they use, and when they need the service. In a report from Microsoft [9] there is a recent (November 2010) calculation on the cost benefits when using cloud either using external resources (public clouds) or internal resources (private clouds). As can be seen from this figure there is a clear cost benefit in using cloud resources for smaller up to larger size cases, with a larger value in using public offerings. We'll come back to this picture later on. To decide which service to use the user need to consider a number of things: for how long will the cloud service be used, what size of service, and usage pattern. Usage pattern is hardest to predict, i.e. will the usage be evenly distributed over the period or will there be temporary peaks in usage. Amazon [2] describes the benefit in using clouds in relation to usage patterns illustrating the cost when the providers, the traditional data center, either over provisions or under provisions its resources. In case of over provisioning the user pays too much for the delivered service. In the case of under provisioning the user of the service is either rejected due to lack of resources or affected by the overload of the resources. To help the user of the cloud resources to decide which service to use we have constructed a guiding decision tree model below.

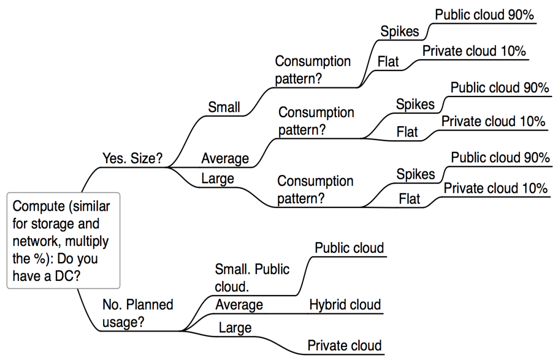

Figure 3: Cloud Decision Tree

In this decision tree the user is asked to consider size of use and usage pattern, resulting in below check list.

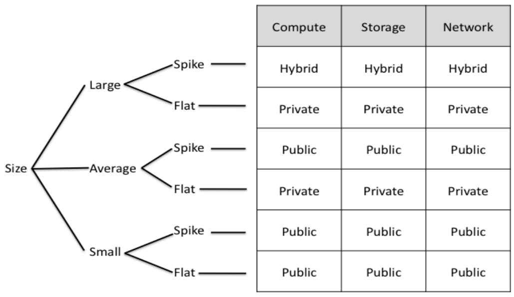

Figure 4: Cloud Decision Tree Check List

If the user for example wants to conduct work of the size of a large data center with a flat predictable behavior on compute, storage and network usage – the optimal choice is to deploy a private cloud on the user's own resources. If on the other hand, for the same case, there is an expected spike in usage of compute resources a hybrid (a private cloud using public cloud resources when needed) is to be considered.

Security and Legal considerations

Above reasoning does not consider security need of the user's data. For many applications, e.g. studies of consensus and medical data, public clouds are not allowed due to legal regulations. Time of usage is another factor to take into consideration. If the usage level is high and over longer time (what is called 'Flat' above) there are break points when a private solution is more cost efficient. Some early studies [6] suggest to use clouds if the total need of compute time is below 12 months, and extra storage need below 6 months. These break points are moving with the decreasing pricing (e.g. 15% price reduction between June 2010 and November 2010) and in the [6] study examples of collaboration between federated private clouds show how these break points can be adjusted.

Cloud computing and environmental costs

Electricity cost is rapidly rising to become the largest element of total cost of ownership, currently representing 15%-20%. Power Usage Effectiveness (PUE) describes how much extra power is needed to deliver the requested IT service. If PUE equals 1, this means that all inserted power is used to deliver the requested service. PUE equal 2 means that for each kW added for compute, storage and network one more kW is lost on (mainly) cooling. PUE does not show the full picture but gives us an idea of how well the data centers are doing with respect to energy consumption. High PUE renders high financial and environmental cost. From "Sustainable IT – a Year in Review", Joyce Dickerson, Nov. 5, 2009 [7] Normal data center PUE is 2 and above. Exceptional data center PUE is 1,5. Google and similar have a PUE of 1.2. Lower PUE gives a competitive advantage on pricing. In the overall environmental picture the source of energy need to be taking into consideration, i.e. where the energy is produced and how. In the Nordic region there are a number of alternatives, including thermal alternatives (Iceland), water power plants (Norway, Sweden, Finland).

Cost – public cloud offering

The pilots of the NEON project found that the cost of administration of today's private clouds is too unpredictable, resulting in a overall recommendation towards public clouds. The pricing models of public clouds is highly competitive lead by Amazon, from which we make below comparisons: to deliver the services on internal resources, or to buy it from the (public) cloud. This brings us back to the TCO (Total Cost of Ownership) results from the most recent report on Cloud and Cost [9]. As mentioned at the first view of below picture (Figure 5.), public cloud offering is overall more cost efficient than the private cloud alternative – for the use cases where both could be used. The conclusion from this study is: "— for organizations with a very small installed base of servers (<100), private clouds are prohibitively expensive compared to public cloud. The only way for these small organizations or departments to share in the benefits of at scale cloud computing is by moving to a public cloud. For large agencies with an installed base of approximately 1,000 servers, private clouds are feasible but come with a significant cost premium of about 10 times the cost of a public cloud for the same unit of service, due to the combined effect of scale, demand diversification and multi-tenancy. In addition to the increase in TCO, private clouds also require upfront investment to deploy – an investment that must accommodate peak demand requirements. This involves separate budgeting and commitment, increasing risk. Public clouds, on the other hand, can generally be provisioned entirely on a pay-as-you-go basis. " To make a fair comparison private-cloud vs public cloud – private clouds can be more customizable, but comes with a installation cost and with lower flexibility. To summarize: we only compare to public cloud in our cost analysis, leaving private cloud alternatives for future studies (see conclusion and recommendation section). Two things are needed now: to understand the cost of our current data centers, and to get good estimates on what these (or parts of these) services would cost on a public cloud (in this case Amazon).

Current cost

There's a number of studies on this field, and we used approximate numbers from [3], distributed these numbers among the NEON partners for comments.

Table 1: Current cost – from IBM article [11]

Amortized Cost	Component	Sub-Components
45%	Servers	CPU, memory, storage systems
25%	Infrastructure	Power distribution and cooling
15%	Power draw	Electrical utility costs
15%	Network	Links, transit, equipment

The overall finding among the NEON partners was: that above numbers are fairly close to their experiences, that most partners don't have their data center (DC) costs on that detail, and that especially administration cost was very hard to estimate. The administration part of the cost is one area where cloud computing has its benefits – less administration is needed when the actual machines resides on an external site. Still, the difficulty of estimating the current administration cost is due to the administrator's many varying duties, including participation in projects not directly related to the handling of the data centers. Nevertheless: we want to point out this lack of data, and find it most likely that if we could include the actual administration cost of today, cloud services would not make this expense higher (more likely lower).

Specific Cloud Migration Cost Estimates – Norway

One of the NEON partners, NOTUR, was especially studying public cloud services – with a specific user case (eSysBio) using AWS. During these studies a more detailed cost estimate was done.

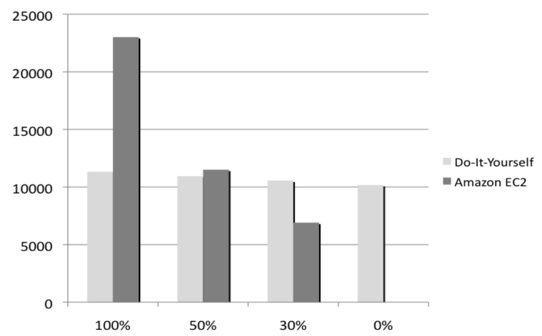

Figure 5: NOK/node/year versus utility level

Actual number of NOTUR centers were compared to AWS using the Amazon 'price calculator' . In above example we see a break-even point at 50% utility level. All examples are Linux instances, based on November 2010 data.

Storage and Network The factors deciding the storage cost is detailed on:

- Storage/Day/Week/Month
- Reduced Redundancy Storage/Day/Week/Month
- Data Transfer In/Day/Week/Month
- Data Transfer Out/Day/Week/Month
- PUT/COPY/POST/LIST Requests
- GET and Other Requests

The basic rule for Amazon S3 service of today (November 2010 data) is: 12 euro cents per GB and month – 35 euro cents to get 1 GB up and back with options of direct links to, e.g. Amazon Ireland. For the upload and download of larger amount of data there are services for direct transport of media storage through e.g. DHL.

Add-on Services

Comparing to Do-It-Yourself, Cloud services comes with a number of add-on services. For example from Amazon the list of services is steadily growing, as is from the private cloud vendors. To complete the picture when deciding on alternatives, the user should consider the different use cases he/she expects to better evaluate the final costs and usability. E.g. load balancing and autoscaling, management tools are most likely to be needed, as is ID management and metering. Cloud is usually look at as a way to lower existing costs, but is also about adding flexibility and new services.

4.1 Additional models and studies for cost estimates

We have in parallel been studying the current work on cloud cost modeling and found a number of interesting work - all at an early and rather theoretical level for our current use. With this note we want to make the reader aware of near-future help in choosing the optimal level of cloud usage. Examples are given here: [1, 11, 5, 4, 10, 8, 12, 13]

5. CONCLUSIONS AND RECOMMENDATIONS

Private clouds are not mature enough for our Nordic eScience users, public clouds offer already a value, both in cost and in usability, for the non-HPC user. Private cloud services are still at an early stage, and not easy to use for a regular administrator. While Private clouds (e.g. Eucalyptus, OpenNebula) have some way to go, Public clouds, especially from Amazon, are now in a more mature stage: well documented, easy to use, predictable and feature rich. In addition, Amazon has a number of initiatives for academic use, e.g. the Amazon Education program with recurring grants for research applications. In the NEON project we applied for one of these (USD 5,000), and got the application accepted within two weeks – and started to use it the same day.

Many pilot installations, one bioinformatics pilot, and a common storage layer to connect all together.

During the project a number of alternatives were tested (see picture below) and in addition a common open source cloud-backed storage service was set up to link these together. A pilot in bioinformatics, eSysbio, was launched on a public cloud with promising results (mostly based on the above mentioned Amazon grant).

Cost: Public clouds are on par with local alternatives. Private clouds are not predictable

Looking at the cost of using clouds, we focused on the Public side comparing a real-life HPC cluster with a non-HPC cluster offering from Amazon. This comparison gave at hand that the costs are comparable, with a higher flexibility on the Amazon alternative. Private clouds are still too hard to install, manage and maintain – making the cost calculation futile. There are management tools, not for free, like

RightScale, that mitigate this. The cost of RightScale and similar tools was considered too high for our community, and it would still not remove the current issues with the private clouds' immaturity level.

Note: we did manage to install private clouds on a number of sites but the experience shows that the work and support needed was too high to be of practical usage for large scale deployments without significant extra manpower.

Risks: Lock-in effects continue to be an issue, limiting the usage for some researchers

Lock-in effects are the top risk when using both the public and private cloud. This has to be evaluated on a case-by-case basis, emphasizing the limitations in publishing private/sensitive data for public clouds. Specifically, data transfer costs for larger deployments may cause an economic lock-in when moving away from a public cloud. Private clouds lock-in is migration cost; given the maturity level it is likely that cross-grades need to take place because of enhanced and new (required) features, adding extra cost for manpower.

5.1 Near-term Recommendation – use Public clouds for non-HPC and some HPC users

Due to the immaturity of the private cloud offerings, we recommend users to take into account the need of advanced system administrators to install, use and manage private clouds of today. If not comfortable with these services, better wait for them to mature and meanwhile focus on public cloud offerings, and/or on improving the local virtualization efforts. This way, an organization can become familiar with the underlying technology and build operational excellence with regards to cloud technology. Public clouds are ready to be used, as e.g. the example of eSysbio shows. There are many ways and services for simplified public cloud deployments, e.g. like Heroku, Rightscale or Amazon's own Elasticfox or web based management console. Support of the conclusions above can be found in the Norwegian initiative for 2011 with the following strategy: "—NOTUR infrastructure should actively try to start moving non-HPC jobs to cloud technology in 2011. Initially this should be done by offering a cloud computing service on a small scale in a public cloud. This will lead to organizational and operational experience and excellence and bootstrap and organize the user communities. After 2011 a private cloud may be set up or a public cloud can still be used – depending on maturity and pricing of both types of clouds after 2011."

5.2 Long-term Recommendation

Wait and learn, continue testing public cloud offering, cooperate internationally - take lead on public cloud

"Wait and see" is sometimes a good option, but in this case it is better to "Wait and learn", i.e. continue with the above public cloud experiments by deploying a small-scale cloud service for non-HPC jobs while this new field matures. This is similar to what is described above for NOTUR 2011. Running non-HPC jobs, and even smaller HPC jobs, on a public cloud, learning more about very long-term stability and cost fluctuations – and mobility – is what we recommend. The eSysbio pilot started in July and has been running since, giving good first input for a bigger "Virtual Data Center" in the public cloud. Another development during the "wait and learn" period could be to establish a cloud backed storage usable for all cloud users, independent of cloud focus. This storage service would be the "glue" of a cross-nation wide cloud service. Alternatively, current storage resources could be opened up as a private cloud storage solution, as deploying storage only is inherently less complex than deploying a complete cloud stack. Following the above near-term recommendation is an adaptive way forward, and together with ECEE – the European open collaboration on cloud projects (www.scientific-cloud.org) – a way to minimize risk of double work, and repeating of mistakes already done by others. In addition the ECEE roadmaps gives us a say and insight into future common projects and collaborations – as well as possible interoperability challenges.

Following the above near-term recommendation is an adaptive way forward, especially if based on ECEE and other international collaborations. *[Mitigates the risk of not being part of the future evolution of clouds for eScience]*

Stay ahead of the expected user adoption to public cloud offerings – by being the primary point of contact for any researcher wanting to use e.g. Amazon. This is achieved by creating a shared knowledge among the NGIs on how to best use Amazon and others, and in addition by being a preferred customer for e.g. Amazon – getting better support and consolidated (cheaper) pricing than the direct use of Amazon. In addition we could become the 'grants office' for e.g. Amazon in our region/countries. [*Mitigates the risk of losing users and users losing time and money repeating (by us) known mistakes]*
Keep a small scale private cloud up to date and monitor the feature set, complexity and cost during 2011; determine a go/no-go for private clouds versus public clouds at the end of 2011. *[Mitigates the risk of slow start in adopting clouds for eScience]*

Summary - Risks and their Mitigation

Following above recommendations mitigates the risk of losing the initiative in cloud usage for eScience in our region. I.e. if we do nothing on clouds, the users will, presumably on clouds. The rationale is quite simple: for small to medium sized projects computing cost is small in absolute numbers. Research that has not been awarded "time" might find other ways to get enough money to start in a cloud. This will result in a less structured and cost efficient usage of cloud for eScience. Also, the initiative might be lost. Another risk is the described sensitivity of data and usage of clouds off premises, depending on local legislature. When using clouds special care must be taken to see if the organization participates in the safe harbor program if it is US based. Finally, there is a risk of economic lock-in for private and public clouds and economic DoS for public clouds. As described above, the economic lock-in mostly concerns data transportation costs for public clouds and a high-frequent update cycle and its associated costs for private clouds. The economic DoS is unlikely but potentially devastating: if a set of users accounts of a public cloud should be compromised the providing organization (i.e. the NGI) would end up paying a lot of extra money. Note that most public cloud providers do have thresholds built-in that require manual intervention to scale beyond, but the risk is on the providing site. The following table provides a summary of the risks and the measures:

Table 2: Risks and their measures

Risk	Measure
Doing nothing	Users will, and may disrupt current practices and administration
Data sensitivity	Local data storage, only computing in cloud. Safe harbor policy when dealing with US based cloud providers.
Economic lock-in private clouds	Wait and watch, via e.g. ECEE
Economic lock-in public clouds (data)	Volume deals and tenders. Keep data local when possible.
Economic DoS	Negotiate volume thresholds; active monitoring tools that monitor overall usage and cost structure.

6. ACKNOWLEDGMENTS

All authors would like to thank their home NGIs and the Nordic Data Grid Facility, NDGF (ndgf.org) for the financing and support of this study. Swedish authors want to especially thank the Swedish National Infrastructure for Computing (SNIC), and Danish author thank Danish Center for Scientific Computing (DCSC) for extra funding and support. Maarten Koopmans did his work funded by UNINETT Sigma AS, Norway.

7. ADDITIONAL AUTHORS

Zeeshan Ali Shah
PDC, KTH Royal Institute of Technology, Sweden
email: `zashah@pdc.kth.se`
Ilja Livenson
PDC, KTH Royal Institute of Technology, Sweden
email: `livenson@kth.se`)
Frederik Orellana
Niels Bohr Institute, University of Copenhagen, Denmark
email: `orellana@nbi.dk`
Jukka Kommeri
Helsinki Institute of Physics, Finland
email: `jukka.antero.kommeri@cern.ch`
Miika Tuisku
Helsinki Institute of Physics, Finland
email: `miika.tuisku@hip.fi`
Pekka Lehtovuori
CSC - IT Center for Science, Finland
email: `Pekka.Lehtovuori@csc.fi`
Klaus Marius Hansen
University of Copenhagen, Denmark
email: `klausmh@diku.dk`
Helmut Neukirchen
University of Iceland, Iceland
email: `helmut@hi.is`
Ebba Hvannberg
University of Iceland, Iceland
email: `ebba@hi.is`

8. REFERENCES

[1] I. S. Ali Khajeh-Hosseini, David Greenwood. Cloud migration: A case study of migrating an enterprise it system to iaas. *IEEE 3rd International Conference on Cloud Computing*, July 2010.

[2] Amazon. The economics of the aws cloud vs. owned it infrastructure. December 2009.

[3] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia. Above the clouds: A berkeley view of cloud computing. Technical Report UCB/EECS-2009-28, EECS Department, University of California, Berkeley, Feb 2009.

[4] M. Assuncao, A. de Costanzo, and R. J. . Buyya. Evaluating the cost-benefit of using cloud computing to extend the capacity of clusters. *HPDC '09 Proceedings of the 18th ACM international symposium on High performance distributed computing*, pages 141–150, 2009.

[5] A. Bain, I. Read, J. Thomas, and T. . Merchant, F.

[6] R. Campbell, I. Gupta, M. Heath, S. Ko, M. Kozuch, M. Kunze, T. Kwan, K. Lai, H. Y. Lee, M. Lyons, D. Milojicic, D. O'Hallaron, and Y. C. Soh. Open cirrus cloud computing testbed: Federated data centers for open source systems and services research. *Usenix HOTCloud'09*, 12(2):291–301, June 2009.

[7] J. Dickerson. Sustainable it – a year in review. *EDUCAUSE Annual Conference*, November 2009.

[8] A. Greenberg, A. Hamilton, D. Maltz, and P. Patel. The cost of a cloud: Research problems in data center networks. *ACM SIGCOMM Computer Communication Review*, 39(1):68–73, January 2009.

[9] R. Harms and M. Yamartino. The economics of the cloud. Microsoft whitepaper, Microsoft Corporation, November 2010.

[10] D. Kondo, B. Javadi, P. Malecot, F. Cappello, and A. D. Cost-benefit analysis of cloud computing versus desktop grids. *18th International Heterogeneity in Computing Workshop*, May 2009.

[11] J. Koomey, K. Brill, P. Turner, J. Stanley, and B. Taylor. A simple model for determining true total cost of ownership for data centers. *IBM Corporation*, November 2007.

[12] I. Livenson, G. Singer, S. Srirama, U. Norbisrath, and M. Dumas. Towards a model for cloud computing cost estimation with reserved resources. In *2nd International ICST Conference on Cloud Computing (CLOUDCOMP 2010)*, 2010.

[13] M. Menzel, M. Schönherr, J. Nimis, and S. Tai. (mc2)2: A generic decision-making framework and its application to cloud computing. *International Conference on Cloud Computing and Virtualization (CCV 2010)*, 2010.