SUPPORTING SYSTEM DEPLOYMENT DECISIONS IN PUBLIC CLOUDS

Ali Khajeh-Hosseini

**A Thesis Submitted for the Degree of PhD
at the
University of St Andrews**

2013

Supporting System Deployment Decisions in Public Clouds

Ali Khajeh-Hosseini

School of Computer Science

University of St Andrews

A thesis submitted to the University of St Andrews

for the degree of Doctor of Philosophy

August 2012

I, Ali Khajeh-Hosseini, hereby certify that this thesis, which is approximately 52,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in January 2009 and as a candidate for the degree of Ph.D. in January 2009; the higher study for which this is a record was carried out in the University of St Andrews between 2009 and 2012.

Date ……………… Signature of candidate ………………

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Ph.D. in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date ……………… Signature of supervisor ………………

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Embargo on both all of printed copy and electronic copy for the same fixed period of 1 year on the following ground:

Publication would preclude future publication.

Date ……………… Signature of candidate ………………

Date ……………… Signature of supervisor ………………

Acknowledgments

It feels wonderful writing this page having just finished a decade at Scottish universities, first at the Robert Gordon University, then at the University of Edinburgh, and now at the University of St Andrews. I feel privileged to have access to such a great academic community. I am deeply grateful to:

Ian Sommerville for never saying no and being the best PhD adviser, ever.

Dave Greenwood, James Smith, Ilango Sriram, Derek Wang, John Rooksby, Gordon Baxter and the wider St Andrews & LSCITS crew for the research discussions and feedback over the years, as well as being supportive and open to collaboration.

Frank Gunn-Moore, Juliana Bowles and Alan Boyd for the thoughtful discussions in the car and making the commute from Edinburgh to St Andrews so enjoyable.

Mohammad and Tahereh for being the perfect parents, and Hassan and Fatemeh for being the best brother and sister. I cannot thank you all enough for encouraging me during my studies.

And finally, to Maryam for her love and support, and smiling whenever I said, "no, I need to work on my thesis". It's finished honey.

Abstract

Decisions to deploy IT systems on public Infrastructure-as-a-Service clouds can be complicated as evaluating the benefits, risks and costs of using such clouds is not straightforward. The aim of this project was to investigate the challenges that enterprises face when making system deployment decisions in public clouds, and to develop vendor-neutral tools to inform decision makers during this process. Three tools were developed to support decision makers:

- Cloud Suitability Checklist: a simple list of questions to provide a rapid assessment of the suitability of public IaaS clouds for a specific IT system.
- Benefits and Risks Assessment tool: a spreadsheet that includes the general benefits and risks of using public clouds; this provides a starting point for risk assessment and helps organisations start discussions about cloud adoption.
- Elastic Cost Modelling: a tool that enables decision makers to model their system deployment options in public clouds and forecast their costs.

These three tools collectively enable decision makers to investigate the benefits, risks and costs of using public clouds, and effectively support them in making system deployment decisions.

Data was collected from five case studies and hundreds of users to evaluate the effectiveness of the tools. This data showed that the cost effectiveness of using public clouds is situation dependent rather than universally less expensive than traditional forms of IT provisioning. Running systems on the cloud using a traditional 'always on' approach can be less cost effective than on-premise servers, and the elastic nature of the cloud has to be considered if costs are to be reduced. Decision makers have to model the variations in resource usage and their systems' deployment options to obtain accurate cost estimates. Performing upfront cost modelling is beneficial as there can be significant cost differences between different cloud providers, and different deployment options within a single cloud. During such modelling exercises, the variations in a system's load (over time) must be taken into account to produce more accurate cost estimates, and the notion of elasticity patterns that is presented in this thesis provides one simple way to do this.

Table of Contents

List of Figures

List of Tables

1 Introduction

Cloud computing represents a shift away from computing as a product that is owned, to computing as a service that is delivered to consumers over the internet from large-scale datacentres — or *clouds*. This shift is already having a profound effect on the ways that software is procured, developed and deployed, similar to the effect of moving from mainframes to PCs. Clouds were initially used by technology start-ups such as Twitter and Dropbox as they provided an elastic and relatively cheap infrastructure layer, on top of which companies could deploy their systems. Recently, more established enterprises are have shown an interest in cloud computing due to its potential benefits, which include scalability, reliability, cost-effectiveness and ease of deployment [1].

However, much ambiguity and uncertainty exists regarding the actual realisation of these benefits, as there is currently hype, particularly around the cost savings of cloud computing, which are sometimes based on simplistic assumptions. For example, elasticity – the ability to acquire more computational resources only when needed – is one of the key benefits of using clouds. Amazon Web Services, the largest cloud provider, named its main offering **Elastic** Compute Cloud (EC2) to signify this. However, the existing literature from academia and industry rarely includes discussions around the cost of different ways to provide elasticity.

Furthermore, for cloud computing to deliver real value, it must be aligned to the enterprise rather than simply be a platform for basic tasks, such as application testing or product demonstrations. Cloud computing is not simply a technological improvement of data centres but a fundamental change in how IT is provisioned and used [2]. Thus the adoption of cloud computing is likely to change the work of various system stakeholders in the enterprise. Therefore, the issues around migrating systems to the cloud and satisfying the requirements of key system stakeholders have to be explored. These stakeholders include technical, project, operational and financial managers as well as the engineers who are going to be developing and supporting the systems.

This thesis focuses on identifying cloud adoption challenges for enterprises, and supporting them during their decision making process. Our approach has been to develop a set of practical tools and evaluate their effectiveness using case studies from industry. The tools are the:

- **Cloud Suitability Checklist**: this tool comprises a simple list of questions to provide a rapid assessment of the suitability of public IaaS clouds for a specific IT system.
- **Benefits and Risks Assessment tool**: this tool is a spreadsheet that includes the general benefits and risks of using public clouds, which were identified by reviewing over 50 academic papers and industry reports. This tool provides a starting point for risk assessment and helps organisations start discussions about cloud adoption.
- **Elastic Cost Modelling tool**: this tool enables decision makers to model their system deployment options in public clouds and forecast their costs. The tool includes a unique feature that lets users describe usage patterns of their computing resources and perform 'what-if' style cost analysis to compare different deployment options or cloud providers.

These tools collectively enable decision makers to investigate the benefits, risks and costs of using public clouds, and effectively support them in making system deployment decisions.

1.1 Cloud Computing

Since Amazon announced the public launch of its Elastic Compute Cloud (EC2) in August 2006 [3], the term "cloud computing" has gained traction in industry. One of the reasons for this traction has been the publicity generated by successful start-ups that deployed their web applications on public clouds, and watched as their applications auto-scaled to thousands of servers during peak usage. Animoto.com was one such start-up; they enable users to upload photos and select/upload a music track, with which they generate a video slideshow with various visual affects. When Animoto launched their application on Facebook, they grew from 25,000 users to 250,000 users in three days; at peak 20,000 new users were signing up per hour[1]. This sudden scale-up required Animoto to increase their server count from 50 to 4,000 on AWS' clouds.

[1] http://blog.rightscale.com/2008/04/23/animoto-facebook-scale-up

Zynga, a gaming company, also uses public clouds in an effective manner. Zynga launches new games on one of AWS' public clouds as it does not know, in advance, what the demand for each game is going to be. Each of Zynga's new games that is launched in a public cloud is monitored for three to six months. If the growth-rate of the game becomes flat, or if the usage of the game becomes predictable, the game is moved from AWS' public clouds into Zynga's private cloud – called zCloud[2]. Companies such as Animoto and Zynga demonstrate that the elasticity offered by public clouds is one of its key benefits.

There are many definitions of cloud computing [4–6]. The US National Institute of Standards and Technology (NIST) has published a working definition [7] that has captured the commonly agreed aspects of cloud computing. Based on this definition, the focus of this thesis is on public IaaS clouds as shown in Figure 1.

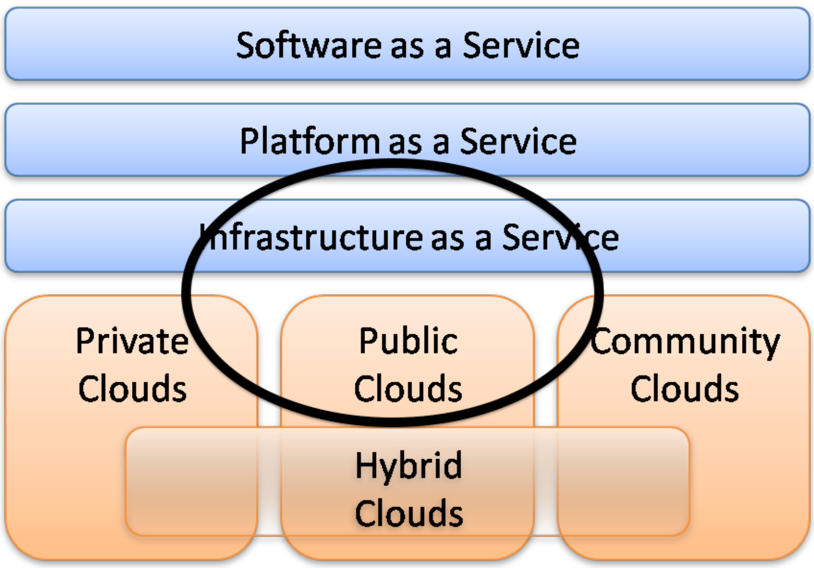


Figure 1: This thesis focuses on public IaaS clouds

The NIST definition describes cloud computing using five characteristics, four deployment models and three service models:

- Characteristics:
  1. *On-demand self-service*: resources can be requested using APIs.
  2. *Broad network access*: resources can be accessed over the internet.
  3. *Resource pooling*: the underlying infrastructure is shared between users.

[2] http://code.zynga.com/2012/02/the-evolution-of-zcloud

4. *Rapid elasticity*: users can request more resources at anytime and release resources that are not needed.
5. *Measured service*: resource usage is metered and billed on a regular basis.

- Deployment models:
  1. *Public clouds*: clouds that are available to anyone who registers, usually using their credit card to pay the monthly bills. Popular public cloud providers include Amazon Web Services[3], Microsoft[4] and Google[5].
  2. *Private clouds*: clouds that are only available to a single organisation, usually the organisation that built the cloud. Popular private cloud software (that is used to build private clouds) include Eucalyptus[6], Open Stack[7] and Cloud Foundry[8].
  3. *Community clouds*: clouds that are built by several organisations for their shared use, usually in one industry.
  4. *Hybrid clouds*: a conceptual cloud that consists of a mixture of the above three deployment models. For example, an organisation could use a private cloud for most of its workloads and occasionally use public clouds to handle peaks in usage.
- Service models:
  1. *Software as a Service (SaaS)*: this is where software is delivered to users using web applications in a browser; this model has been around for around ten years and was popularised by companies such as SalesForce[9] and more recently by Google Enterprise[10]. SaaS applications are usually developed using a multi-tenant architecture, and are deployed on either PaaS or IaaS.
  2. *Platform as a Service (PaaS)*: this is where organisations develop their applications using languages and APIs that are supported by the cloud provider. The provider does not expose developers to servers or the

---

[3] http://aws.amazon.com
[4] https://www.windowsazure.com
[5] https://appengine.google.com
[6] http://www.eucalyptus.com
[7] http://openstack.org
[8] http://www.cloudfoundry.com
[9] http://www.salesforce.com
[10] http://www.google.com/enterprise

underlying infrastructure as they take care of deploying applications, scaling them and ensuring that they are migrated to new servers if the underlying infrastructure fails. Popular PaaS providers include Heroku[11] and AppEngine[12]. Most PaaS providers, including Heroku, use IaaS to deploy their platform.

3. *Infrastructure as a Service (IaaS)*: this is the usually lowest level of abstraction offered by cloud providers, where organisations can provision and control virtual machines, storage, and some of the network configurations such as opened ports. Popular IaaS providers include AWS and Microsoft, which are the type of companies that have the expertise and capital to build and operate large-scale datacentres for their cloud services.

1.2 Project Aims and Motivations

The aim of this project is to support organisations during cloud adoption and assist system deployment decisions in public clouds. Many enterprises are not familiar with the benefits, risks and costs of using public clouds as cloud computing brings a major shift in how IT may be provisioned and consumed within organisations. Our work, therefore, has two broad objectives:

1. To support enterprises in assessing the benefits and risk of using public clouds.
2. To provide enterprises with a tool to model the costs of deploying systems in public clouds.

There are two main motivations behind this project. Firstly, cloud adoption is a topical issue, and there is significant interest from industry in using public clouds. Thus this project is partly driven by demand from industry. Secondly, as academics, we are in a unique position to offer unbiased advice and expertise to enterprises that are interested in using new technologies. Thus this project, which is rooted in academic research and fills a gap in the literature, provides vendor-neutral expertise and tools for enterprises that are interested in deploying systems in public clouds.

[11] http://www.heroku.com
[12] https://appengine.google.com

Enterprises need to have a discussion about the benefits and risks of using cloud computing to raise awareness, and ensure that decision makers understand the issues involved during cloud adoption from different stakeholder perspectives. The Cloud Suitability Checklist and the Benefits and Risks Assessment tool, presented in this thesis, provide a starting point for those discussions.

As part of their cloud adoption decision making process, enterprises may also wish to estimate the costs of using public clouds. However, most enterprises are not familiar with the pricing schemes used by different cloud providers, and pay-as-you-go pricing models provide a number of challenges for cost estimation and forecasting:

1. Public clouds enable elastic systems to be developed that scale on-demand. Elasticity has associated costs, but how should this be modelled?
2. Usually, there are multiple deployment options for a system on public clouds and each has associated costs. How should the various system deployment options be modelled?
3. There are many cloud providers worldwide; they have different pricing schemes and they regularly change their prices. How should these differences be modelled such that users:
    a. Can obtain accurate and up-to-date cost estimates.
    b. Can compare cost estimates in a fair manner.
    c. Perform 'what-if' style analysis to understand the effects on their costs if they cloud providers change their prices or if they consume more/less resources than their original estimates.

These challenges were addressed by developing the Elastic Cost Modelling tool that is presented in this thesis.

1.3 Novel Contributions

There are four novel research contributions made by this work:

1. The notion of *elasticity patterns* that is presented in this thesis contributes to the understanding of cloud computing costs, and enables more realistic cost forecasts to be produced. Elasticity patterns are constructed using a simple language to describe variable usage patterns for computing resources.

2. The Elastic Cost Modelling tool that was developed as part of this thesis. This tool contributes to the existing system modelling literature by describing how IT practitioners can model the deployment of their systems on public clouds. The tool is unique as it enables users to model their systems in a practical and simple manner, and produces cost forecasts by running the model through a simulation that takes into account elasticity patterns and up to date prices from cloud providers.

3. The identification of the potential issues that can arise during cloud adoption in enterprises is novel as it attempts to highlight the overall organisational implications of using cloud computing. These issues have not been discussed to any significant extent in the existing literature, and their inclusion in the Cloud Suitability Checklist and the Benefits and Risks Assessment tool provides a starting point for risk assessment.

4. The in-depth evaluation of the tools developed. The five case studies presented here, the 230 downloads of the Benefits and Risks Assessment tool and the data collected from the 270 users that used the online version of the Elastic Cost Modelling tool, ShopForCloud.com, is considerably more detailed than is usual in academic theses. The evaluation highlighted the simplicity and usefulness of the tools presented in this thesis. The related research in cloud computing, which is reviewed in the next chapter, rarely shows this level of industrial applicability.

1.4 Thesis Structure

The structure of this thesis is shown in Figure 2 and is described below.

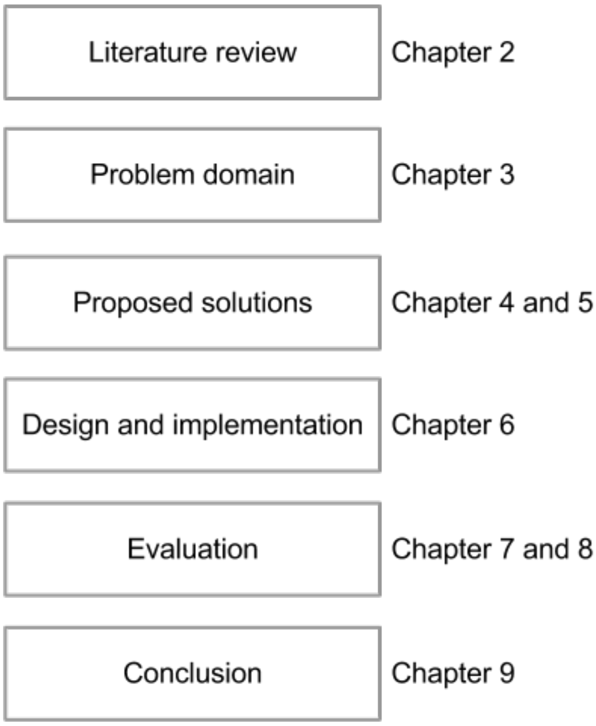


Figure 2: Thesis structure

Chapter 2 reviews the cloud computing literature, focusing on related work in cloud adoption, cloud cost modelling and general system modelling tools and techniques. This chapter also reviews the research methods that were used in this project.

Chapter 3 discusses the challenges that enterprises face during cloud adoption, which include organisational change, security and compliance, support and maintenance of systems, as well as estimating the cost of system deployment in public clouds. This chapter focuses on the problem domain and identifies the challenges that are being addressed by this project.

Chapter 4 identifies the general benefits and risks of using public clouds and describes how the Cloud Suitability Checklist and the Benefits and Risks Assessment tool were developed; decision makers can use these tools as a starting point for risk assessment.

Chapter 5 introduces the key concepts behind the Elastic Cost Modelling tool, which include the modelling notations, and the notion of elasticity patterns that were developed as part of this project. This chapter also describes how a prototype was developed, and the lessons that were learned from this exercise.

Chapter 6 describes the detailed design and implementation of the online cost modelling tool, ShopForCloud.com. This chapter also describes the tool’s key components including the simulation engine and algorithms that are used to estimate and forecast costs.

Chapter 7 reviews four case studies that were used to evaluate the tools developed in this thesis. This chapter also discusses the key findings and feedback that were received from the case studies.

Chapter 8 continues the evaluation of the tools presented in this thesis by describing an experiment that was used to measure the accuracy of the cost modelling tool. This chapter also discusses the data and feedback that were gathered from users of ShopForCloud.com.

Chapter 9 concludes this thesis by reviewing the lessons that were learned from the case studies and providing a critical evaluation of the tools. The chapter ends by describing future work that could further develop this research.

2 Literature Review

This thesis focuses on the challenges that organisations face during the adoption of cloud computing, and proposes a number of tools to support decision makers investigate the benefits, risks and costs of using public clouds. Our aim was for the work to be relevant to industry and useful in practice, thus several industrial case studies were conducted and the tools were all publicly released to gauge their usage in practice.

The first part of this thesis identifies cloud adoption challenges in enterprises. The Cloud Suitability Checklist and the Benefits and Risks Assessment Tool were developed to support decision makers by making them aware of the issues that need to be considered during cloud adoption. Section 2.1 discusses the cloud adoption literature, and includes a review of other tools that also aim to support cloud adoption decisions.

The second part of this thesis focuses on estimating the costs of deploying systems on public clouds. A cost modelling tool, called ShopForCloud.com, was developed to enable decision makers to model their systems to obtain cost estimates and investigate the costs of using different cloud providers and deployment options. Section 2.2 reviews the cloud cost modelling literature and describes other tools that are available in industry. ShopForCloud includes a set of modelling notations that were inspired by the Unified Modelling Language, thus Section 2.3 discusses related system modelling notations and tools.

Finally, this chapter ends with a brief review of the research methods that were used in our work, which included modelling, simulation and case studies (Section 2.4).

2.1 Adoption of Cloud Computing

The cloud adoption literature that is relevant to this thesis can be divided into a number of categories including the organisational change that cloud adoption brings (Section 2.1.1), the security and compliance challenges that organisations face (Section 2.1.2 and Section 2.1.3), and finally the supporting tools that are available to assist organisations during their decision making process (Section 2.1.4). The focus of this thesis is on practical approaches to support decision makers, thus the sections that describe the literature relating to organisational change, security and compliance do not go into much depth.

2.1.1 Organisational Change

The literature has not so far examined the organisational change issues regarding cloud adoption to a great extent. Indeed, this thesis is one of the early works that investigated such issues (discussed in Chapter 3). Yanosky [8] discussed how cloud computing will affect the authority of the IT department within universities; however, his work also applies to IT departments in enterprises.

Yanosky discusses the evolution of IT support in organisations and the role of central IT departments. The IT department gained its authority in the early days of computing when they had the majority of the programming skills and control of mainframes within an organisation. As the use of IT expanded within organisations, system administrators and developers were forced to learn new skills as their role was no longer just about keeping the technology running. Until the invention of the PC, users relied on the services provided by the IT department for systems support.

The adoption of the PC eroded some of the IT department's authority as it provided users with an opportunity to create and use applications without the explicit support of the IT department. Users went on to form online communities to support each other as they were more experienced in solving technical problems. Although the IT department no longer had full control over the technology, it did have "a set of carrots and sticks at hand [...] including the supreme sanctions of refusing support for shadow systems [...] or cutting off network connectivity" [8].

The authority of the IT department is going to be further eroded by cloud computing. Cloud computing is increasingly turning "users into choosers" [8] who can replace the services provided by the IT department with service offered in the cloud. Yanosky believes that users will end-up asking for support from the IT department when they have problems with a cloud. However, this might not be the case as cloud providers, such as AWS, are offering support services as well[13].

[13] http://aws.amazon.com/premiumsupport/

Yanosky suggested that IT departments could respond to this change by either controlling and monitoring the services that are allowed to be used in a cloud, or by providing a certification program, where they only support certified services. Either way, users will continue to have increased political influence within organisations, and the IT department's role will change from "provider to certifier, consultant and arbitrator" [8]. A similar view is also held by Erbes et al. who see a shift in the skillset of IT personnel from "designing, building, and operating IT services to contracting, integrating, and managing relationships with service providers" [9].

The type of organisational change that cloud computing results in can be demonstrated by considering IT procurement within enterprises. Simplistically, procurement is based on obtaining estimates for things, then getting those estimates signed-off by management to allow the procurement to proceed. Capital and operational budgets are kept separate in this process, and it can take several months between the decision to procure hardware and the hardware being delivered, setup and ready to use.

The use of cloud computing can greatly reduce this time period, but the more significant change relates to the empowerment of users and the diffusion of the IT department's authority as pointed out by Yanosky [8]. For example, a company's training coordinator who requires a few servers to run a week-long web-based training course can bypass their IT department and run the training course in a cloud. They could pay their cloud usage-bill using their personal or company credit card and charge back the amount as expenses to their employer. Such a scenario was recently reported in BP, where a group bypassed the company's procurement, IT department and security processes by using AWS to host a new customer facing website[14].

However, currently the typical enterprise IT department is not used to a utility billing model across shared infrastructures; resource sharing across such infrastructures requires a certain level of cultural and organisational process maturity, and the move towards cloud computing will require significant changes to business processes and organisational boundaries [10]. Therefore, users need to consider the benefits, risks and the effects of cloud computing on their organisations and usage-practices in order

[14] http://a6.64.354a.static.theplanet.com/Individual-Case-Studies/bp-fuels-cloud-computing-interest

to make decisions about its adoption and use; the potential for reduced costs could be just one of the significant benefits of cloud computing. Erdogmus [1] lists other benefits as “scalability, reliability, security, ease of deployment, and ease of management for customers, traded off against worries of trust, privacy, availability, performance, ownership, and supplier persistence”. Some of these issues are discussed in subsequent sections of this thesis.

Motahari-Nezhad et al. [11] briefly discuss the benefits and risks of using cloud computing from a business perspective. They highlighted the lack of environments for helping businesses migrate their legacy applications to the cloud. In addition, they pointed out the difficulties of finding and integrating different cloud services for a given set of business requirements. They propose a conceptual architecture for a virtual business environment where individuals and SMEs can start and operate a virtual business using cloud-based services. This conceptual architecture includes four layers: business context, business services, business processes and IT services. Motahari-Nezhad et al. conclude by sketching an implementation of their conceptual architecture and the challenges encountered for each of the four layers. The work performed in the IT services layer of this conceptual architecture is going to be mostly affected by cloud computing.

Elson and Howell [12] provide just one example of the ways in which cloud computing could potentially affect the work of IT departments. They describe how cloud computing can resolve conflicts in system development roles. For example, a startup company that provides a hosting service for bloggers could, without cloud computing, be forced to build its own datacentre even though it only wants to integrate and offer existing open-source software to its customers. Elson and Howell describe this scenario as conflating the roles of a software integrator and a hardware wrangler (someone who sets up and maintains hardware). They describe how IaaS cloud providers such as AWS resolve such conflicts by providing an “explicit and narrow interface” in the form of a virtual machine (VM) image. Software integrators create the VM image and hardware wranglers deploy them.

As part of this thesis, we investigated the feasibility of migrating an IT system at an SME in the Oil & Gas industry to the AWS cloud [13]. We found that despite the promised financial benefits, opportunities to remove tedious work from IT staff and the potential to enter new marketplaces, almost all of the stakeholder groups were neutral or reluctant to support a move to the cloud due to concerns regarding its impact on their work, increased risk of dependence upon third parties and its implications for customer service and support. Therefore, from an enterprise perspective, costs are important but so too are customer relationships, public image, flexibility, business continuity and compliance. The organisational changes faced by enterprises are identified and discussed in Chapter 3.

2.1.2 Security

Security issues are widely acknowledged as being important in cloud computing. Grobauer et al. [14] provide an overview of vulnerabilities in using a cloud, which relate to web applications, cryptography protocols and the underlying communication protocols that are used in the cloud. Jensen et al. [15] also provide an overview of the technical security issues in cloud computing. Most of these issues are not specific to cloud computing as they relate to the underlying security problems of web services and web browsers. However, these security issues become more significant as cloud computing makes heavy use of web services, and users rely on browsers to access services offered in the cloud. For example, denial of service (DoS) attacks were a common concern even before cloud computing became popular, but when an application is targeted by a DoS attack in the cloud, the user or owner could actually end-up paying for the attack through their increased resource usage. This could be significantly higher than the peak usage of that application in an in-house datacentre with limited resources. In fact, Jesper [16] recently reported this scenario with their application running on the AWS cloud but they did not mention if they were charged for the usage generated or if AWS waived the extra costs incurred as a result of the attack.

Such incidents and other security concerns have resulted in the establishment of the Cloud Security Alliance (CSA), which is an industrial group with members from corporations such as Microsoft and HP. The CSA has published a set of best practices and guidelines for organisations adopting cloud computing. These guidelines come in the form of problem statements and issues that need to be considered by cloud service consumers. They cover a wide range of areas including encryption and key management, portability and interoperability, and risk management [17]. The

European Network and Information Security Agency (ENISA) has also published a report about security issues in cloud computing. They identified 35 risks of using cloud computing [18], which are split into the following categories:

- Policy and organisational risks such as vendor lock-in, loss of governance, compliance challenges, and cloud provider acquisition.
- Technical risks such as data leakage, distributed denial of service attacks, loss of encryption keys, and conflicts between customer hardening procedures and cloud platforms.
- Legal risks such as data protection and software licensing risks.
- Risks not specific to the cloud such as network problems, unauthorized access to datacentres, and natural disasters.

In addition to the above risks, ENISA's report pointed out that some security measures become cheaper to implement on a larger scale. Therefore, cloud providers could potentially provide greater security, such as hardened VM instances or hiring experts, that can deal with particular threats [18].

Others have also considered the security benefits of using cloud computing. Armbrust et al. [4] discussed the security levels of systems deployed in local datacentres and compared this with the potential security of using a cloud. In conclusion, they believe that systems deployed in a cloud could be made as secure as systems deployed in local datacentres. They supported their hypothetical argument by mentioning the possible use of technologies such as encrypted storage to improve the security of a system deployed in a cloud. This work, however, falls short of analysing potential weaknesses of running systems in a cloud.

One of the strengths of developing SaaS applications in the cloud is that intruders will not have access to the entire application source code; this could be seen as a security advantage over traditional applications where intruders have access to application executables that might be reverse engineered [19]. Startup companies could use cloud computing to eliminate the costs of developing a secure infrastructure. Kaufman [20] discussed security responsibility and asked if using cloud computing will result in security being a joint responsibility with the cloud providers. The question of responsibility was also raised in Mansfield-Devine [21], who highlighted the lack of control and checks over third parties who develop applications on top of the PaaS layer of the cloud.

2.1.3 Compliance

Most of the security issues in cloud computing are caused by users' lack of control over the physical infrastructure. This leads to legal issues that are affected by a cloud's physical location, which determines its jurisdiction. Joint et al. [22] provided an in-depth review of the legal and compliance issues that UK-based organisations should consider when using cloud computing. These include three of the eight principles of the UK's Data Protection Act (DPA), a law that came into place to protect personal data. For example, the first DPA principle requires organisations to ask for an individual's permission before processing personal data. The use of cloud computing makes it difficult to provide individuals with full details of how and where their data will be processed. This lack of knowledge or transparency makes it difficult for individuals to reach "informed consent" [22]. In addition, Joint et al. pointed to legal issues regarding confidentiality, copyright, and specific rules that govern businesses that are regulated by the Financial Services Authority in the UK.

Physical data location is important as there are no internationally agreed rules for data protection. For example, AWS offer their cloud services from both U.S. and European-based datacentres to be able to deal with differences in each region's rules and regulations. Jaeger et al. [23] reviewed a range of issues that are affected by the geographic location of a cloud, i.e. the location of the physical datacentres. Jaeger et al. emphasized the importance of locality by pointing out that cloud computing increases the control of governments and corporations over resources. This is because cloud computing brings together vast amounts of data and computing resources in centralized datacentres, compared to the current situation of hosting in geographically dispersed locations. While the location of a cloud has a significant effect on the rules and regulations that govern it, it is unclear "whether a cloud will be considered to legally be in one designated location [...] or in every location that has a datacentre that is part of the cloud" [23]. This poses problems for companies that are considering using, for example, US-based cloud providers that have datacentres outside of the US as the US government might still be able to access their data. It is unlikely that these jurisdiction issues will stop the use of cloud services; however, they will have long-term implications that need to be considered by users.

Nelson [24] wrote a report as an advisory document for the US government. Nelson stressed the important role that governments can play in advancing the use of cloud computing. Just as procurement decisions made by the U.S. government in the 1980s were fundamental in the development of the Internet, governments have the power to advance cloud computing usage by being "model users" [24]. Nelson also provided advice to governments and policy makers on how to proceed; this included the need for governments to update their IT procurement rules to support procurement of systems with an externally-provided cloud service model, as well as the need for them to encourage experimentation with the cloud. There are some signs that governments are taking on this type of advice, for example, the U.S. Federal Government issued a Request for Information regarding IaaS offerings [25], and more recently AWS launched their US GovCloud[15].

2.1.4 Tool Support

There are few tools to support decision makers during cloud adoption. Saripalli and Walters [26] describe the QUIRC framework, which is a qualitative risk assessment methodology based on assessing the probability and severity of a threat to calculate its impact. The authors propose the use of the Delphi method to collect the necessary information for the risk assessment. The works of Saripalli and Walters [26] and Grobauer et al. [27] focus on assessing the security risks of using the cloud, whereas this thesis argues that other risks must also be considered during the decision making process. Therefore, this thesis builds on previous works in cloud security risk assessment by identifying other risks, such as organisational and legal risks (further discussed in Chapter 4).

Saripalli and Pingali [28] propose Multiple Attribute Decision Methodology for Adoption of Clouds (MADMAC) to support decision makers. MADMAC is based on the principles of multi-attribute decision making [29] and relies on the Delphi method for the collection of input data to their framework. The use of MADMAC involves three steps:

[15] http://aws.amazon.com/govcloud-us/

1. Definition of decision attributes. Six categories of attributes are predefined and provide a starting point for this
2. Assigning weights to attributes by experts. This can be done either using a Likert-style scale or using a method such as Delphi to enable expert opinions to converge
3. Ranking the alternative cloud adoption options to come to a decision.

Saripalli and Pingali [28] are developing a software system based on MADMAC, but no practical tool is currently available to support decision makers. Although the authors mention that such decision support systems have been used successfully in other industries (such as environmental, agricultural and public health), they have not provided any evidence of their approach being used in practice. Overall, this approach seems to be unnecessarily complicated and reliant on 'experts' whose ratings have to be trusted by decision makers.

Misra and Mondal [30] describe their method of identifying a company's suitability for the adoption of cloud computing, which relies on decision makers assigning credits (weights) to a range of factors, and their tool suggesting one of three possible outcomes based on their calculations: not suitable for cloud adoption, may or may not be suitable for cloud adoption – further investigation required, and suitable for the adoption of cloud. The authors stress that their work acts as an initial guide in arriving at a decision, and should not be used as a definitive tool as in most cases in-depth analysis is needed on various factors. Like the authors of MADMAC, Misra and Mondal have not provided any evidence of their tool being used in practice.

Menzel et al. have developed CloudGenius [31], which is a tool to support the selection of virtual machines and cloud providers for web applications. This tool is based on their $(MC^2)^2$ decision making framework[16] [32] and takes into account cost, performance, latency, uptime and the popularity of the cloud provider. Their framework is similar to MADMAC as it acts as a multi-criterion decision support system and enables users to define scenarios, alternative options for each scenario and a set of criteria that can be assigned weights by experts. The authors mention that their

[16] http://aotearoadecisions.appspot.com

framework is being evaluated at a German IT infrastructure provider but do not provide any details on the nature of the evaluation or its outcome.

Overall, approaches such as [28], [30], [32] advocate automated methods involving the definition of a set of decision attributes, assigning weights to them, and ranking alternative options to select the 'best one'. This can be seen as rather rigid and unnecessarily complicated, but is also relatively high-effort compared to our approach. Furthermore, these approaches mostly focus on the technical aspects of cloud adoption, whereas we argue that such decisions also need to take organisational factors into account (discussed in Chapter 3).

In industry, Accenture's Cloud Computing Accelerator [33] and CSC's Cloud Adoption Assessment [34] are examples of typical offerings from IT consultancies that attempt to support cloud adoption decisions. Such approaches have two problems: they are based on closed proprietary tools that are not widely available, and they are often accompanied by expensive consultancy contracts. In contrast, we argue that given the tools described in this thesis, enterprises can assess the feasibility of using cloud computing in their organisations quickly and cheaply without the need for external consultants. Our tools can also be used by decision makers to check the claims made by IT consultancies and cloud providers.

It should be noted that aside from the references mentioned in this chapter, the references mentioned in Chapter 4 also point to related work regarding the benefits and risks of using cloud computing, which has until now been sparsely reported in various academic papers and industry reports. The comprehensive review presented in Chapter 4 builds upon those existing works, and presents the benefits and risks in a tool that provides a starting point for risk assessment.

2.2 Cloud Computing Costs

This section reviews the literature relevant to cloud computing costs (Section 2.2.1), and describes the tools that are currently available to support decision makers estimate the costs of using public IaaS clouds (Section 2.2.2).

2.2.1 The Cost of Using IaaS Clouds

Cloud providers have detailed costing models and metrics that are used to bill users on a pay-per-use basis. This makes it easy for users to see the exact costs of running their applications in the cloud and it could well be that the design of their system can have a significant effect on its running costs. Armbrust et al. [4] mention short-term billing of as one of the novel features of cloud computing and a number of researchers have investigated the economic issues around cloud computing from a consumer perspective.

Youseff et al. [6] described the three pricing models that are used by cloud service providers, namely tiered pricing, per-unit pricing and subscription-based pricing. Tiered pricing is where different tiers each with different specifications (e.g. CPU and RAM) are provided at a cost per unit time. An example is Amazon EC2, where a small tier Linux virtual machine has the equivalent of a 1.0GHz CPU, 1.7GB of RAM with 160GB of storage and costs $0.08 per hour[17], whereas as a large tier machine has the equivalent of four 1.0GHz CPUs, 7.5GB of RAM with 850GB of storage and costs $0.32 per hour. Per-unit pricing is where the user pays for their exact resource usage, for example it costs $0.125 per GB per month[18] to store data on Amazon's Simple Storage Service (S3). Subscription-based pricing is common in SaaS products such as Salesforce's Enterprise Edition CRM that costs $125 per user per month[19].

More elaborate pricing models exist in the grid and utility computing research community [35–38] but Armbrust et al. [4] point to other utilities such as electricity and argue that simpler pricing models will remain dominant because they are transparent and understandable to users. Researchers are also investigating the risks of using such pricing models by private cloud providers that are considering bursting-out to public clouds [39]. New cloud computing pricing models based on market mechanisms are starting to emerge but it is not yet clear how such models can be effectively used by enterprises. An example of such models is used by AWS Spot Instances[20], which allows users to bid for unused capacity in AWS clouds. AWS runs

[17] http://aws.amazon.com/ec2/#pricing
[18] http://aws.amazon.com/s3/#pricing
[19] http://www.salesforce.com/crm/editions-pricing.jsp
[20] http://aws.amazon.com/ec2/spot-instances/

the user's instances as long as the bid price is higher than the spot price, which is set by AWS based on their datacentre utilization.

Klems et al. [40] attempted to address the problem of deciding whether deploying systems in the clouds makes economic sense. They highlighted some economic and technical issues that need to be considered when evaluating cloud adoption. The considerations were presented as a framework that could be used to compare the costs of using cloud computing with more conventional approaches, such as using in-house IT infrastructure. Their framework was very briefly evaluated using two case studies. However, no results were provided because the framework was at an early developmental stage and more conceptual than concrete.

Walker [41] also looked into the economics of cloud computing, and pointed out that lease-or-buy decisions have been researched in economics for more than 40 years. Walker used this insight to develop a model for comparing the cost of a CPU hour when it is purchased as part of a server cluster, with when it is leased (e.g. from AWS EC2). The model was demonstrated using two scenarios, one where the cost of leasing was compared with purchasing a 60,000 core HPC cluster, and another where it was compared with purchasing a compute blade rack consisting of 176 cores. Walker showed that in both scenarios it would be cheaper to buy than lease when CPU utilization is very high (over 90%) and electricity is cheap. However, as expected, this would be reversed if CPU utilization is low or electricity is expensive. Walker also developed similar models for comparing the costs of using cloud storage versus purchasing hard drives [42]. Walker's models are a good first step in developing models to aid decision makers, but they are too simplistic and narrow in scope as they focus only on the cost of a CPU hour or storage gigabyte. Further work towards financial decision support is required that includes other costs such as data transfer and database costs.

Assuncao et al. [43] investigated the use of clouds to extend the capacity of locally maintained clusters. They simulated the costs of using various strategies when borrowing resources from a cloud provider, and evaluated the benefits of doing this by using performance metrics such as the Average Weighted Response Time (AWRT) [43]. AWRT is the average time that user job-requests take to complete, shorter AWRTs means shorter waiting times. The investigation done by Assuncao et al. is potentially useful for organisations that have private clouds and are looking to use public clouds when their in-house resources are over-utilised. However, cloud users expect instant resource allocation and AWRT might not be the best metric to measure performance improvement. Further research is required to identify the right metrics for evaluating the benefits of using public clouds to complement private clouds. This is outside the scope of this thesis.

Deelman et al. [44] used simulation to calculate the cost of running a data-intensive scientific application on AWS. The focus of their study was to investigate the performance-cost tradeoffs of different execution plans by measuring execution times, amounts of data transferred to and from AWS, and the amount of storage used. They found the cost of running instances (i.e. CPU time) to be the dominant figure in the total cost of running their application. Others such as Kondo et al. [45] have found that the majority of the costs could be attributed to bandwidth usage.

The study done by Deelman et al. highlighted the potentials of using cloud computing as a cost-effective deployment option for data-intensive scientific applications. It assumed that the cost of running instances on AWS EC2 is calculated on a dollar-per-CPU-second basis (i.e. they normalized the costs). However, AWS charge on a dollar-per-CPU-hour basis and charge for a full hour even for partial hours. This means that launching 100 instances for 5 minutes would cost 100 CPU hours on AWS. This would result in significantly different costs from those shown by the simulations of Deelman et al. The most cost effective execution plans for AWS would be those that provision the right number of CPUs to ensure that jobs finish execution within full CPU hours and not partial hours.

Singer et al. [46] developed a method to model the load of a server cluster to make recommendations on which sourcing strategies should be used to obtain the same resources from Amazon EC2. The output of their model is a virtual machine usage function that indicates how many VMs (and what types of VMs) are needed, a reservation schedule that indicates how many VMs should be reserved using AWS's 1 or 3-year reservation scheme, and the total cost of using the cloud for the server cluster. Their work is focused on evaluating whether replacing scientific computing clusters with public clouds make economic sense. However, it needs to take into account a number of other factors including the costs of computer rooms, staff, electricity etc. to make their models more realistic.

Furthermore, as Armbrust et al. [4] argue, elasticity – the ability to quickly scale up or down one's resource usage – is an important economic benefit of using public clouds as it transfers the costs of resource over-provisioning and the risks of under-provisioning to cloud providers. Singer et al.'s models should also take into account these risks. Armbrust et al. provide a few theoretical examples to highlight the importance of elasticity with respect to costs. An often-cited real-world example of elasticity is Animoto.com whose active users grew from 25,000 to 250,000 in three days after they launched their application on Facebook[21]. This demonstrates that enterprises can now develop highly elastic systems that respond to user-demand without taking on the risk of under-provisioning infrastructure.

The issue of elasticity was also discussed by Suleiman et al. [47] who highlighted various economic factors that need to be investigated by an organisation that is thinking of deploying their applications on public IaaS clouds. Their aim was to highlight a number of open research challenges, which include modelling the costs and elasticity of an application, selecting appropriate scaling strategies, determining performance bottlenecks, and monitoring SLAs.

Misra and Mondal [30] suggest that the utilization patterns of IT systems (i.e. their elasticity) can be determined by their average usage, peak usage and the amount of data handling/transactions that are done. They propose the following five usage profiles:

1. Moderately variable workload with no surges
2. Highly variable workload with spikes
3. No variability constant workload
4. Moderately variable workload with occasional surges
5. Constant workload but workload pattern varies at different times of the year owing to different types of project undertaken

[21] http://blog.rightscale.com/2008/04/23/animoto-facebook-scale-up/

The above profiles are useful when describing the general usage pattern of systems. However, they cannot be used to describe the specific usage pattern of a system, and are not practically useful in simulation tools. The notion of *elasticity patterns* that is proposed in this thesis addresses these limitations (discussed in Chapter 5). Furthermore, the cost modelling approach and supporting tool presented in this thesis can be used in conjunction with traditional software size measurement techniques to provide a more holistic overview of cloud migration costs. Tran et al. [48] describes Cloud Migration Point (CMP), which is a methodology for estimating the size of cloud migration projects that was inspired by Function Point [49]. This approach assumes that a cloud has already been selected and the tools presented in this thesis could be used to select a cloud.

One of the cloud adoption challenges for enterprises is that the majority of their systems have been built on the assumption that increases in demand will be supported by 'scaling up' to more powerful servers rather than 'scaling out' to larger numbers of servers, as Animoto did. Changing the architecture of these systems to support scaling out will inevitably be very expensive and in many cases will simply be impossible.

Another challenge is dealing with performance variations between single-tenant and multi-tenant infrastructures, where for example in AWS EC2, the same underlying hardware can be shared between multiple user VMs. In such cases, the performance experienced by each VM (measured by running CPU, memory, disk I/O benchmarks etc.) can vary over time and even between different machines running the same instance type [50], [51]. Ideally, the expected performance of instances in a cloud would be measured and categorised in advance, to support decision makers while considering the costs. This area is an open research challenge as historical benchmarks are not always predictive of future performance, as the performance can be affected by other factors such 'neighbouring' VMs running on the same machine and the network load.

2.2.2 Tool Support

Most cloud providers have developed cost calculators for their websites as their utility pricing models are very different to traditional up-front hardware purchasing options, or the hardware leasing contracts offered by IT vendors. These calculators include:

- Rackspace Cost Calculator[22]: A primitive webpage with a slider at the top to resemble memory capacity and a few boxes allowing users to enter the number of servers, running hours, incoming and outgoing bandwidth. A single figure is then displayed at the bottom of the page showing the total monthly cost.
- Microsoft Azure Cost Calculator[23]: A single webpage with sliders that users can update to resemble changes in virtual machine instances, databases, storage, bandwidth etc. This results in a single monthly cost figure to be updated as the sliders change.
- AWS Simple Monthly Calculator[24]: A more comprehensive calculator that includes a number of pre-defined example deployments, and breaks-down user requirements into different AWS offerings. Once users have entered their requirements, a single figure at the top of the page provides the estimated monthly cost; the first month cost is also displayed if any reserved instances are required.

These calculators have a number of problems:

1. One of the key benefits of using the cloud is elasticity. However, these calculators completely ignore elasticity during their calculations. The notion of elasticity patterns presented in this thesis enables users to see the effects of elasticity on their costs.
2. Users cannot easily compare the cost of their system on different clouds or different deployment options using these calculators (for example, AWS's cost calculator does not include Azure's prices). Furthermore, different providers charge for different things, thus users have to describe their system in the cloud provider's terms when using their calculator.
3. Each cloud provider has its own calculator showing prices in the provider's billing currency, and users have to calculate their costs manually if they want to use multiple providers (e.g., deploying a system on the AWS EU cloud but backing it up on the Rackspace UK cloud).

[22] http://www.rackspace.co.uk/cloud-hosting/learn-more/calculator/
[23] https://www.windowsazure.com/en-us/pricing/calculator/advanced/
[24] http://calculator.s3.amazonaws.com/calc5.html

Aside from cloud provider cost calculators, there are return-on-investment and total-cost-of-ownership calculators provided by a number of vendors. These include the Cloud ROI tool[25] by eyeTask, and the Cloud Computing TCO Calculator[26] by RightScale. The extent to which these tools are useful in practice is not clear as the tools are based on proprietary data and the outputs of the tools are not entirely specific to the scenario being modelled. For example, the RightScale TCO cannot be used to obtain cost estimates of using different cloud providers or compare different deployment options. These tools also ignore the cloud's elasticity and cannot model a system's elasticity during their calculations.

Buyya's CLOUDS Lab has developed CloudSim [52], which is a useful toolkit for the modelling and simulation of cloud computing environments. As evidenced by the use-cases mentioned in [52], CloudSim is more suited to developers who are concerned about the performance of their applications, and cloud providers such as HP who are interested in modelling the properties and resource utilization of datacentres. In contrast, the tools presented in this thesis, and specifically the cost modelling tool, are targeted at decision makers in enterprises that are interested in deploying their IT systems on the cloud.

Li et al. [53] have developed CloudCmp[27], which is a framework for comparing cloud providers by running performance benchmarks on them to estimate the performance and cost of an application if it where deployed on a cloud. The framework includes performance benchmarks for CPU, memory and disk I/O, and the per-hour cost of the smallest VM type supported by the target cloud. As the authors mention in their work, they do not consider performance or cost differences between different types of instances within a cloud. Aside from the considerable performance differences between different types of instances in clouds such as AWS, CloudCmp's cost comparison is unrealistic as cloud providers have different prices for their various types of instances, storage, databases, and even multiple purchase options (e.g. reserved, on-demand, spot). The lack of support for these different prices makes CloudCmp unusable in practice.

[25] http://www.eyetask.com/simple-cloud-roi-model/
[26] http://www.rightscale.com/tco-calculator/
[27] http://cloudcmp.net/home

Overall, the cost modelling tool presented in this thesis fills a gap by enabling decision makers to model the deployment of IT systems on various clouds, including their deployment options and elasticity patterns as well as different cloud providers' pricing schemes and any future price changes. The modelling notations used in the cost modelling tool are an important enabling factor in making it a practically useful tool.

2.3 System Modelling

One approach that could support system deployment decisions on public clouds would be to develop a system modelling tool that can be used to model relevant attributes of IT systems for cloud deployment. These models could then be used to reason about and investigate migration decisions. For example, by modelling a system's hardware infrastructure and applications (at the executable level), it becomes possible to estimate the costs of running parts of that system in a cloud.

The Unified Modelling Language (UML) [54] is often used to model software systems in industry. The UML 2.2 includes various modelling notations and has 14 types of diagrams including ones to model use cases, sequences of actions, classes, component, and deployments. The UML deployment diagrams are the most relevant to the work presented in this thesis as they enable users to model their physical system infrastructure and how applications are deployed onto them. In its essence, a deployment diagram enables users to model the deployment of 'software artefacts' onto 'hardware nodes' [54].

The System Modelling Language (SysML) [55] is more general purpose than the UML and is mainly used in systems engineering domains such as defence, automotive, aerospace and medical devices. The SysML is actually specified as a UML profile, meaning that it uses UML's meta-modelling notations, and thus it is supported as a plugin in some UML modelling tools such as Eclipse[28]. The SysML does not include UML's deployment diagrams but instead allows the deployment of software/data onto hardware to be modelled using internal block diagrams. These diagrams provide similar functionality as UML's deployment diagrams but are more

[28] http://www.eclipse.org/modeling/mdt/papyrus/

general purpose and are usually used to describe requirements rather than just a deployment scenario.

The UK Ministry of Defence Architecture Framework (MODAF)[29] and its US counterpart The Open Group Architecture Framework (TOGAF)[30] are enterprise architecture frameworks that can be used manage entire enterprises. MODAF has a number of different viewpoints including Strategic, Operational, System, Technical, Acquisition, and Service Oriented, where each viewpoint contains a number of views. A view is a visualisation that shows an area of focus within the enterprise using either diagrams or text. Some of the MODAF views focus on structural aspects of enterprise architecture, however, none of them are directly focused on system deployments.

For example, the SV-1 view (Resource Interaction Specification) can be used to model system deployments but the notations are similar to the UML. In fact many of the views can be represented using UML diagrams but none of the view use the UML deployment diagram notations directly. Therefore, MODAF highlights a range of considerations and how they might be modelled, but UML provides actual modelling notations and is more relevant to the work presented in this thesis.

System modelling languages such as the UML are mostly utilised during a system's design phase, whereas infrastructure configuration management frameworks and tools are utilised during system deployment. Recently, open-source configuration management tools such as Puppet[31] and Chef[32] have gained popularity in industry, while HP Lab's SmartFrog has been the focus of on-going research for over a decade [56]. Such configuration management tools enable system administrators to declaratively describe the configuration of their systems at a high-level (e.g. services, packages, dependencies), and they take care of executing various scripts and programs to get the system to the described state.

Several companies now provide cloud management platforms that take the basic ideas of configuration management tools and provide customised functionality for the deployment and management of systems on IaaS clouds. One of the most advanced

---

[29] www.modaf.org.uk/
[30] http://www.opengroup.org/togaf/
[31] http://puppetlabs.com/
[32] http://www.opscode.com/chef/

cloud management platform is RightScale[33], which includes a range of features to automate the deployment and monitoring of systems on IaaS clouds. Overall, this thesis focuses on the decision making phase (prior to the actual deployment of a system on the cloud), thus the tools presented in this paper can be used in conjunction with configuration management tools to support organisations migrate their systems to the cloud.

2.4 Research Methods

The initial phase of our work was exploratory in nature, during which we aimed to discover the challenges that organisations face during cloud adoption. This phase involved case study research, which consists of a detailed investigation, within the context of a particular organisation, during which data is collected over a period of time to study a phenomena [57]. Case studies can include multiple data collection methods such as participant observation, ethnography and interviews. The main case study data collection method used in this thesis was semi-structured interviews. This involves the researcher preparing a set of questions and discussion points, and interviewing a small number of individuals to collect 'in-depth' data. In-depth, here, refers to the case where a researcher wants to investigate "how the apparently straight-forward is actually more complicated" [58]; which in this thesis refers to our finding that the adoption of cloud computing, initially seen as posing mainly technical challenges, was in fact more complicated as it posed numerous organisational challenges for enterprises (discussed in Chapter 3).

Once the cloud adoption challenges were understood and an approach (and supporting tools) was developed to support decision makers, further case studies were carried out to research questions such as "how effective are these tools in practice?" and "why would industry practitioners use these tools?" Such 'how' and 'why' questions are particularly suitable for case study research [59] as they involve direct feedback from participants and are performed in the context of their day-to-day job.

Aside from the drawback that case study research, and in particular interviews, can be time intensive and costly to perform, it can be difficult to generalise from case studies as they consist of the investigation of a particular context, which might not be similar

[33] http://www.rightscale.com/

to other scenarios. Furthermore, case study findings can be affected by researcher biases or previous experiences. Thus to address these limitations, we complemented our use of case studies with empirical investigations, where the tools were made available publicly on the web, and data was collected on how they were being used over a four-month period (discussed in Chapter 8).

This thesis also involved the use of simulation-based research, as the cloud cost modelling tool performs a discrete time simulation to estimate the cost of deploying a system on public IaaS clouds. The main issue around the use of simulation-based research is the accuracy of the simulation outputs, and simulation programs need to be verified and validated. The cost modelling tool was verified by comparing its output data, for a given set of input data, with what was expected from the input data. This was achieved by developing a set of automated test cases (described in Chapter 6). In addition, the cost modelling tool was validated by comparing its predicated estimates with the actual deployment costs of a test system during an experiment (described in Section 8.1).

3 Challenges of Cloud Adoption

The adoption of cloud computing is an emerging challenge that enterprises face as the economics of cloud computing become more attractive due to economies of scale and competition amongst providers. Companies such as Amazon, Google and Microsoft are investing vast sums in building their public clouds and they seem to be leading the way in the technological innovation of clouds by releasing new services and frequent updates for their services. For example, a quick look at AWS' news archive in 2011 shows that they rolled-out around 10 new services while adding many new features to their existing services[34]. In addition, they added four new clouds worldwide: two in the west coast of the US (US-West Oregon and US GovCloud), Tokyo and Sao Paulo. AWS also released a Security[35] and an Economic[36] centre on their website, which shows that there is user demand for advice about the implications of using cloud computing. There is an opportunity for the research community to address this demand by providing independent and impartial advice, tools and techniques to enterprise users who are interested in cloud adoption.

The adoption of cloud computing in enterprise environments is non-trivial. Understanding the organisational benefits and drawbacks is far from straightforward because the suitability of the cloud for many classes of systems is unknown or an open-research challenge; the adoption of cloud computing results in a considerable amount of organisational change that will affect peoples' work in significant ways; corporate governance issues regarding the use of cloud computing are not well understood; and cost estimation is complicated due to the number of variables comprising inputs to the pay-as-you-go billing model of cloud computing.

This chapter starts with a case study that identifies some of the major challenges that enterprises face during cloud adoption, which include socio-technical challenges and organisational change, concerns with security and compliance with regulations, and changes in how IT systems are supported and maintained. This case study was chosen as it enabled cloud migration decisions to be explored. The system involved in the case study was non-trivial (i.e. more realistic than a single web application) and three

34 http://aws.amazon.com/about-aws/whats-new/2011/
35 http://aws.amazon.com/security/
36 http://aws.amazon.com/economics/

organisations were involved. The IT systems integrator involved in the case study also provided access to staff members, system acquisition documentation and their support database, which were essential in enabling us to explore the factors involved in cloud migration.

The following chapter (Chapter 4) builds on the issues identified in this chapter, and describes how a system was developed to provide enterprises with a starting point for risk assessment. The shift towards pay-as-you-go pricing models also poses challenges in how organisations can obtain cost estimates of using a cloud. Chapter 5 describes how the cost estimation challenges described in Section 3.5 of this chapter were addressed by developing a cost modelling tool that provides organisations with cost estimates of deploying their system on public clouds.

3.1 A Case Study Highlighting Cloud Adoption Challenges

This case study illustrates the potential benefits and risks associated with the migration of an IT system in the oil and gas industry from an in-house datacentre to Amazon EC2 from a broad variety of stakeholder perspectives across the enterprise. Our results show that the system infrastructure in the case study would have cost 37% less over 5 years on AWS EC2, and using cloud computing could have potentially eliminated 21% of the support calls for this system. These findings seem significant enough to call for a migration of the system to the cloud. However, the case study revealed that there are significant risks associated with this. Whilst the benefits of using the cloud are attractive, it is argued that decision-makers should consider the overall organisational implications of the changes brought about with cloud computing to avoid implementing local optimisations at the cost of organisation-wide performance.

3.1.1 Introduction

Over the last few years, startup companies such as Twitter and Zynga Games have used clouds to build highly scalable systems. However, cloud computing is not just for startups; enterprises are attracted to cloud-based services as cloud providers market their services as being superior to in-house datacentres in terms of financial and technical dimensions e.g. more cost effective, equally or perhaps more reliable, and highly scalable [1], [4], [60]. Whilst the technological and financial benefits may be seductive, it is important that enterprise decision-makers consider the overall organisational implications of the change and thus implement local optimizations at the cost of organisation-wide performance.

This case study investigated the migration of an IT system from a company's in-house datacentre to Amazon EC2. The primary focus of the case study was on the socio-technical and financial issues that decision-makers should consider during the migration of IT systems to the cloud.

This case study identifies the potential benefits and risks associated with the migration of the studied system from the perspectives of: project managers, technical managers, support managers, support staff, and business development staff. The case study is based upon data collected from an IT solutions company considering the migration of one of their systems to Amazon EC2's IaaS cloud. IaaS is arguably the most accessible type of cloud to enterprises as they could potentially migrate their systems to the cloud without having to change their applications. In addition, Amazon Machine Images are readily available for many enterprise applications such as Oracle Database and Citrix XenApp[37].

The remainder of this case study section is structured such that: the proposed migration project is introduced next; the methodology section describes the approach used to collect and analyse data; the results section identifies the cost saving benefits of using cloud computing and its affect on the support and maintenance of the system under investigation; the organisational benefits and risks of the migration are also discussed in the results section; the section concludes by discussing the main points revealed by this case study.

3.1.2 Proposed Migration Project

The case study organisation is a UK based SME that provides bespoke IT solutions for the oil and gas industry. It comprises of around 30 employees with offices in the UK and the Middle East. It has an organisational structure based on functional

[37] https://aws.amazon.com/solution-providers

divisions: Administration, Engineering and Support of which Engineering is the largest department.

The project investigated the feasibility of the migration of one of the organisation's data acquisition and monitoring systems to AWS EC2. The following is an anonymized description of the situation:

Company C is a small oil and gas company who owns some offshore assets in the North Sea oilfields. Company C needed a data acquisition system to allow them to manage their offshore operations by monitoring data from their assets on a minute-by-minute basis. Company C's assets rely on the production facilities of *Company A* (a major oil company), therefore the data comes onshore via Company A's communication links. Company C does not have the capabilities to develop their own IT systems, hence they outsourced the development and management of the system to *Company B*, which is an IT solutions company with a small datacentre.

Figure 3 provides an overview of the system, which consists of two servers:

1. A database server that logs and archives the data coming in from offshore into a database. A tape drive is used to take daily backups of the database, the tapes are stored off-site.
2. An application server that hosts a number of data reporting and monitoring applications. The end users at Company C access these applications using a remote desktop client over the internet.

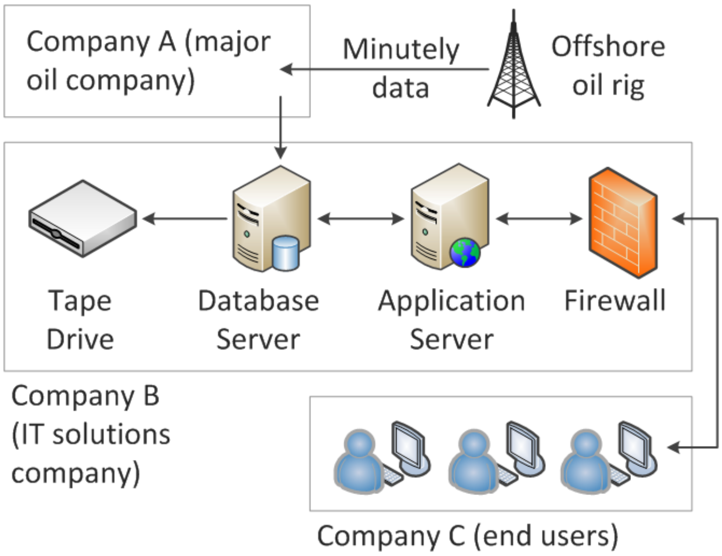


Figure 3: System overview

The system infrastructure was deployed in Company B's datacentre and went live in 2005. Since then, Company B's support department have been maintaining the system and solving any problems that have risen. This case study investigated how the same system could be deployed using the cloud offerings of AWS. Figure 4 provides an overview of this scenario, where Company B deploys and maintains the same system in the cloud.

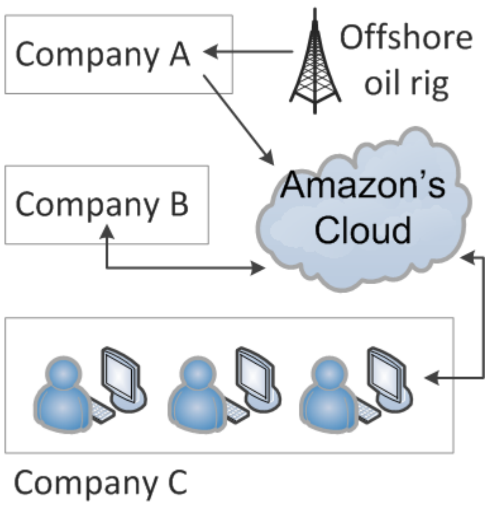


Figure 4: System deployed on a cloud

3.1.3 Methodology

This case study involved fieldwork at Company B's offices from May to July 2009. Initially, all documents relating to the system under investigation were gathered and studied. The fieldwork had three stages:

Stage 1: The infrastructure costs of the system were calculated from project reports and invoices. These costs were compared with the costs of a similar infrastructure setup on AWS EC2.

Stage 2: Company B has a database of all support and maintenance issues regarding the systems that they support. This database was manually researched and all of the support calls that would potentially be affected by the migration were identified and analyzed.

Stage 3: The results from the above two stages were used to produce a poster. The poster was presented to Company B's employees and six semi-structured interviews were performed at their offices. The interviews started by giving the interviewees an overview of AWS EC2 as they were only partially familiar with this technology. Each interview was recorded and a transcript of each interview was produced. The benefits and risks identified from the interviews are discussed in the next section.

3.1.4 Results

3.1.4.1 Infrastructure Costs

Company C paid £104,000 to Company B for the system in 2005, £19,400 of which was for the system's infrastructure; the rest of the costs were for system development and deployment. The infrastructure included two servers (each having two Intel Xeon 3.4GHz processors, 2GB RAM, 6 x 72GB hard drives in a RAID 10 array resulting in around 200GB of effective storage, Windows Server 2003 OS), a tape drive, network equipment, a server rack, shelf spares. In addition, Company C pays £43,000 per year to Company B for system support and maintenance, £3,600 of which is for the running costs of the system infrastructure.

Over a five year period, the total cost of the system infrastructure is therefore: £19,400 + (5 x £3,600) = £37,400 (shown in Table 1). We acknowledge that hardware performance has changed since 2005 and perhaps it may be perceived that costs should have reduced, however in reality they remain similar. For example, the servers used by Company C cost £4,525 in 2005, the ones used in a similar project in 2009 cost £4,445 (when the case study was performed).

AWS EC2 provides an option of using either small or large server instances depending on the amount of CPU power and RAM required. The system could initially run on two small instances as the application and database server do not seem to be under a heavy load. However, this could be changed for large instances if the performance is found to be unacceptable. This would not have been possible using the existing approach since all hardware must be purchased before the system is deployed, and cannot easily be changed afterwards.

Table 1 shows a comparison of the costs of the system infrastructure, the amounts have been rounded to the nearest £10 and are based on AWS' prices at the time when the case study was performed. The following specifications were used to calculate the costs of running the system on AWS: two Microsoft Windows On-Demand instance (AWS did not offer reserved instances for Windows during the case study period) in Europe running 730 hours per month (i.e. 24x7); 20GB data transfer in; 20GB data transfer out; 200GB EBS storage (i.e. amount of effective storage on existing servers), 100 million EBS I/O request; 30GB EBS snapshot storage (for daily backups); 10 snapshot GET requests (in case backups need to be retrieved); 30 snapshot PUT requests (for daily backups).

Table 1: Comparison of infrastructure costs between using AWS and Company B's datacentre

Period	**Amazon Server Instances**			**Company B**
	2 small	**1 small + 1 large**	**2 large**	
1 Month	£200	£390	£590	£620
1 Year	£2,400	£4,680	£7,080	£7,440
5 Years	£12,000	£23,400	£35,400	£37,400

From Company B's perspective, the cloud presents an opportunity to bid for new projects without having to worry about space in their datacentre as they are currently running out of rack space, and building a new data centre is an expensive venture. It also means that they could propose a cheaper alternative to deploying systems in their in-house datacentre for their clients.

From Company C's perspective (the end users), Table 1 shows that the cost of running their system in the cloud is cheaper than using Company B's datacentre. For example, it would be 37% cheaper to deploy the system in the cloud assuming that a small and a large server instance are used. Furthermore, no upfront capital is required for infrastructure in the cloud since users are charged on a monthly basis. The potential cost reductions certainly seem significant, but the affects of a migration on the support and maintenance of the systems must also be considered.

3.1.4.2 Support and Maintenance

The system is currently supported and maintained by Company B's support department who also perform regular health checks to ensure that the system is running as expected. The health checks involve checking error logs, backup logs, server load levels, communication links etc. The support and maintenance of the system would be affected if the system was migrated to the cloud since the support department would no longer have full control over the system infrastructure.

Company B maintains a database of all the support calls they receive by telephone or email either externally from end users or internally from support engineers doing regular heath checks. Since the system went live in 2005, 218 support calls have been made regarding the operation of the system. The majority of these calls were about software problems, however, the titles of all calls were studied and a shortlist of 112 calls was made for further investigation. It was found that the following 45 calls were related to the system's infrastructure:

- 38 calls were related to backup problems between the database server and the tape drive. Common problems included faulty tapes, failed backup attempts, and even loose cables presumably related to tapes being taken in and out of the drive on a daily basis. These problems were usually fixed by erasing the tapes, rebooting the tape drive or re-running backup scripts, but there were a few occasions when no backup was taken for that day.
- 5 calls were related to network problems, one of which required a router to be rebooted, and another that was caused by a power cable being unplugged accidentally.
- 2 calls were related to power outages at Company B's datacentre.

The previously mentioned calls could potentially have been eliminated if the system was deployed in the cloud since Amazon would be responsible for hardware related issues. As shown in Figure 5, this accounts for around 21% of the support calls but it should be noted that some additional calls might be introduced if the system was migrated to the cloud. These cloud related issues could include power outages at Amazon's datacentres or network latency issues; however, the important point is that these issues would be dealt with by Amazon. This could be seen as a big advantage for Company B's support department as it allows them to focus on software related issues, which are more important to the end users.

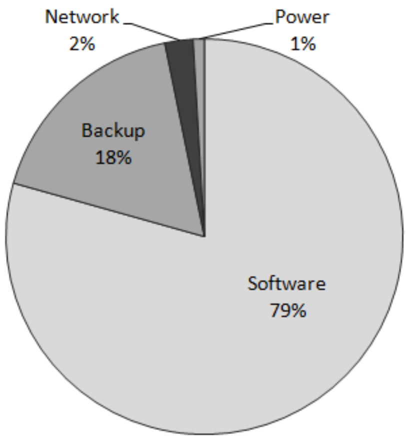


Figure 5: Overview of support calls

3.1.4.3 Stakeholder Interviews

Analysis of the interview data suggests that the proposed cloud migration would have a positive net benefit from the perspective of the business development functions of the enterprise and the more junior levels of the IT support functions. A perceived zero net benefit was perceived by the project management and support management functions of the enterprise. A negative net benefit was perceived by the technical manager and the support engineer functions of the enterprise.

The interview data suggests that there are numerous potential benefits but also risks associated with the migration of the system to the cloud. Table 2 and Table 3 summarize the benefits and risks of the proposed migration project. The second column in the tables refers to the number of specific benefits/risks identified, and hence indicates the distribution of benefit or risk across different areas. Twelve specific benefits were identified in contrast to fifteen specific risks. According to the analysis the largest source of benefit to be derived from using public clouds is the opportunity to manage income and outgoings in a new way, followed by the improved satisfaction of work and removal of tedious work and opportunities to offer new products/service. The largest source of risk will be derived from the potential deterioration of customer care and service quality due to increased dependence on 3rd parties, decrease in job satisfaction and departmental downsizing.

Table 2: Sources of benefits identified by stakeholder impact analysis

Benefits	Number of benefits
Opportunity to manage income and outgoings	3
Improve satisfaction of work & removal of tedious work	3
Opportunity to offer new products/services	2
Improved status	2
Opportunity to develop new skills	1
Opportunity for organisational growth	1

Table 3: Sources of risk identified by stakeholder impact analysis

Risks	Number of risks
Deterioration of customer care & service quality due to increased dependence on 3rd parties	6
Decrease of job satisfaction	3
Departmental downsizing	2
Uncertainty with new technology	2
Lack of supporting resources and understanding of the cloud	2

3.1.4.4 Benefits

3.1.4.4.1 Opportunity to manage income and outgoings

Introducing third party cloud infrastructure solutions presents itself as an opportunity to improve the management of income and outgoings for both finance staff and customers. Third party cloud infrastructure solutions facilitate the easing of cash-flow management for finance staff as the cloud's pricing model has minimal upfront cost and monthly billing, and it also minimizes variability of expenditure on electricity. These are a benefit, in contrast to an in-house datacentre, as upfront costs of buying hardware are high and clients can be slow to pay, resulting in cash-flow difficulties. Additionally energy costs are a significant outgoing and, in principle, by using an external provider they would benefit from the provider's ability to negotiate wholesale energy prices. However, as public cloud providers do not provide details of energy costs for systems deployed on their clouds, it is difficult to assess this.

Third party cloud infrastructure solutions also surface many opportunities for managing income for customers, sales and marketing staff, as new pricing models can be offered to them. This is a benefit, in contrast to setting up internal datacentres, which has a large upfront cost. The use of public clouds means that the finance department can directly bill their customer for their infrastructure usage, hence reducing the finance departments' administrative burden.

3.1.4.4.2 Improved satisfaction of work and removal of tedious work

Third party cloud infrastructure solutions present an opportunity for support engineers, sales and marketing staff to improve the satisfaction of their work. It is an opportunity for support engineers to shed unsatisfying routine and potentially time-consuming work such as hardware support, network support and switching backup tapes as well as being offered new challenges in terms of cloud administration. This is a benefit as support engineers can focus on more satisfying and value-adding work such as resolving customers' software support requests. This benefit is enabled by the migration to cloud infrastructure as the cloud provider would be responsible for the infrastructure maintenance.

Technical developers could also benefit from the migration as they are involved in systems support (e.g. performing regular system health checks), which is sometimes viewed as a chore. In small organisations, there is not usually a clear distinction between the roles of system administrators and technical developers, and different people have to be involved when there is a problem.

3.1.4.4.3 Opportunity to offer new products/services

Using a public IaaS cloud presents an opportunity for sales and marketing staff to create new product/service offerings that better fit customers needs in terms of scalability and cost effectiveness. This also provides staff with a new and potentially satisfying challenges that would not have existed without the migration to a cloud.

3.1.4.4.4 Improved status

Introducing third party cloud infrastructure solutions presents an opportunity for support management and support engineers to improve their status. Support managers can improve their status in the organisation by successfully championing the high profile migration that has strategic implications. This is a benefit to the support manager as by working with new and potentially prestigious technology it may lead to career progression and increased job satisfaction. Support engineers would also benefit by improving their status within their industry by developing sought after cloud administration skills and experience.

3.1.4.4.5 Opportunity to develop new skills

Migrating to a cloud presents an opportunity for support managers, engineers, sales and marketing staff to develop new skills. For support managers and engineers it is an opportunity to develop new system administration skills for applications deployed on a cloud. This is a benefit as the support engineers will expand their existing skill sets and experience with knowledge of managing a technology that will be in demand throughout the IT industry for years to come. For sales and marketing staff it presents an opportunity to develop skills in product/service creation and launching. This is a benefit to sales and marketing staff as it will expand their existing skill sets and experience enabling their career progression.

3.1.4.4.6 Opportunity for organisational growth

Using public IaaS clouds presents an opportunity for sales and marketing staff to create new product/service offerings that may appeal to a larger market due to a cloud's scalability and its cost effectiveness. This is a benefit as it may facilitate sales staff meeting targets by enabling them to target market segments previously not attracted by limitations of scalability.

3.1.4.5 Risks

3.1.4.5.1 Deterioration of customer care and service quality due to increased dependence on 3rd parties

Using public IaaS clouds presents a risk to customer care and overall service quality for support managers, support engineers and customer care staff. Support managers and engineers are at risk of becoming dependent upon a cloud service provider, which they have no control over. There is also a risk of requiring additional resources to deal with short-term issues that arise subsequent to the migration (e.g. shortfalls in cloud operations knowledge resulting in tasks taking temporarily longer to complete). Support managers and engineers specifically risk becoming dependent upon a cloud service provider for resolving hardware and network issues. This is a risk as it could result in the deterioration of service quality that the support manager would not be able to control. Support managers also risk temporarily requiring more resources to cope with migration and also the relative lack of knowledge and experience held by support staff regarding cloud systems. This is a risk because staff may initially require more time to perform the same tasks due to the time required to learn how to perform tasks in the cloud environment, which could compromise service quality and customer service.

Customer care staff are also at risk of not being able to offer the existing levels of customer service as it may take longer to resolve customer queries because cooperation with external service providers may become necessary. This is a risk because response times to deal with customer queries may increase resulting in backlogs and cascades of additional work as customer call back for progress updates and will result in customer care staff dissatisfaction.

3.1.4.5.2 Decrease in job satisfaction

Using IaaS clouds presents a risk of decreasing job satisfaction for support engineers, sales and marketing staff. Support engineers risk decreasing job satisfaction as work may shift from a hands-on technical role to reporting and chasing up issues with third party cloud providers. Support engineers will become dependent upon the responsiveness of third party service providers to resolve problems thus reducing the level of control support engineers have over resolving issues.

This is a risk to support engineer satisfaction as they derive satisfaction from technical aspects of work and rapidly resolving problems to customer satisfaction. Sales and marketing staff risk of decreasing job satisfaction if they are set unrealistic goals regarding the selling of the new cloud based services. This is a risk to sales and marketing's satisfaction as they derive satisfaction from meeting sales and market share targets. Customer care staff also risk decreasing job satisfaction because their ability to perform their job will be dependent upon third parties out of their control resulting in a greater lag between customer queries and resolution.

3.1.4.5.3 Departmental downsizing

The use of public IaaS clouds presents a risk of downsizing to IT support departments. IT support departments are at risk of downsizing if the majority of their work comprises hardware and network support. This is a risk because cloud providers will be responsible for maintaining these aspects of support making the capability unnecessary within the IT support department. Both support managers and support engineers will be impacted as support engineers may lose their jobs and the support managers may lose influence as they have a small department.

3.1.4.5.4 Uncertainty with new technology

Public IaaS clouds are relatively new and present a risk to the finance/business development staff as it may open the organisation to long-term volatility derived from market forces associated with the costs of using a cloud. Additionally, migrating a system to a cloud decreases the certainty of customer lock-in in terms of software support contracts as now the hardware is maintained externally and therefore the company can no longer make the case that it offers an 'all-in-one' maintenance contract which avoids having to deal with multiple contactors. Another consideration is the loss of in-house expertise resulting in additional barriers to bringing the system back in-house if the cloud provider is inadequate.

3.1.4.5.5 Lack of supporting resources and understanding of the cloud

The use of public IaaS clouds presents a risk in resource scarcity in IT support and sales/marketing departments. There is a risk of having to temporarily upsize the IT support departments to cope with migration and also the relative lack of engineering knowledge and experience of cloud-based systems. This is a risk because staff may initially require more time to perform the same tasks due to having to learn new skills for systems deployed in a cloud. There is a risk of temporarily upsizing sales/marketing to cope with the creation and launch of new cloud-based products/services. This is a risk because sales and marketing staff will need to develop appropriate strategies and materials to ensure the marketplace is aware of their product offering.

In summary, these results illustrate that whilst the financial and technological analyses are certainly important, the organisational dimension should also be considered. This should be particularly considered from service quality and customer care perspective, and the organisational governance and risk implications of being so highly dependent upon a third party for product/service delivery to customers. In some cases, the financial dimension may not even be the primary consideration for business-critical applications. These findings are reinforced by the fact that at present the majority of management at the organisation is reluctant to implement the change beyond a test environment despite the financial incentives as the risks are perceived to outweigh the lost savings.

3.1.5 Discussion

The findings of this case study show that cloud computing can be a significantly cheaper alternative to purchasing and maintaining system infrastructure in-house (for the system under investigation). Furthermore, cloud computing could potentially eliminate many support-related issues as there is no physical infrastructure to maintain. Despite these advantages, this case study showed that there are important socio-technical issues that need to be considered before organisations migrate their IT systems to the cloud.

The system infrastructure in the case study would have cost 37% less over 5 years on AWS EC2, and using cloud computing could have potentially eliminated 21% of the support calls for this system. These findings seem significant enough to call for a migration of the system to the cloud but interviews with various employees revealed that there are significant disadvantages tied to the promised benefits. The disadvantages include risks to customer satisfaction and overall service quality due to diffusion of control to third parties; decreased job satisfaction due to changes in nature of work; and opening the organisation to long term cost volatility in terms of cloud-usage costs.

This case study has practical implications for industrial practitioners assessing the benefits of external cloud infrastructures for their organisation. The generic benefits identified can be leveraged to gain buy-in from stakeholders whilst the generic risks identified should be adapted into a risk register and monitored to ensure their projects do not fall prey to common cloud infrastructure migration risks.

The limitations of this case study were that the cost analysis only focused on system infrastructure costs, and did not quantify: the cost of doing the actual migration work; how the support staff costs would be affected by the migration; the cost of a support contract that might be required with AWS Premium Support[38]. Support staff costs are difficult to quantify as they would first require the system to be migrated to the cloud and run for a period of time to study any issues that would arise. There are also longer-term costs associated with the migration of systems, such as the cost of migrating to another cloud provider if the current provider is inadequate or raises their prices, or even the costs associated with the loss of experience/knowledge if a company needs to re-deploy the system in-house.

Whilst it is clear that from a financial perspective, end-users could benefit from cloud computing it is unclear whether it will materialize for the majority, as many organisations outsource their IT to system integrators and it is at present unclear whether there is sufficient financial incentive for these system integrators to migrate systems to the cloud. On first impression it may appear that system integrators profits will not rise (assuming IaaS costs are passed directly to the end-user) and may even be marginally less due to loss of small profits associated with hardware sales. However this ignores the following facts:

1. That the system integrator will be able to focus their resources and effort on performing value-adding and more profitable activities (e.g. system software support) rather than hardware builds and hardware maintenance;
2. The system integrator will not be paying in-house hosting costs e.g. electricity, cooling, off-site tape archiving.

[38] http://aws.amazon.com/premiumsupport/

Both facts indicate that in the medium to long term the integrator will be more profitable per unit of work performed. These arguments however require empirical substantiation.

The remainder of this chapter provides a more in-depth discussion of the various issues that enterprises face during cloud adoption, including issues relating to organisational change (Section 3.2), security and compliance (Section 3.3), support and maintenance of IT systems (Section 3.4), and finally system deployment costs (Section 3.5).

3.2 Organisational Change

The benefits of new technologies are usually realized when enterprises change their structure and processes to take advantage of technological innovation and, for sure, the benefits of cloud computing can only be realized if such changes take place. There was a major organisational change when PCs became cheap enough to buy on individual departmental budgets and power shifted away from the IT department. Cloud computing is likely to result in a similar change but on a more significant scale, because power not only shifts away from the IT department, but also shifts outside the organisation to cloud providers such as Amazon.

Understanding the significance and the extent of the organisational changes associated with cloud adoption can be difficult for enterprises. However, the success of cloud adoption "is as much dependent on the maturity of organisational and cultural (including legislative) processes as the technology, per se" [10]. The process is likely to be prolonged and some predict that it could take 10 to 15 years before big enterprises makes this shift [61]. A number of issues need to be investigated as changes will arise throughout an organisation:

- What will be the changing role for the central IT department within organisations? Will their role change from "provider to certifier, consultant and arbitrator" as Yanosky [8] suggests, or will the complexity of IT systems and the lack of customized support from cloud providers and online support forums mean that organisations will still need central IT to provide and support most of their systems?

- What are the political implications for organisations that lose control over some aspects of their services? Will it mean that moving to a cloud-based system will be resisted by support personnel and system administrators who might either be worried about losing their jobs, or about no longer having complete control over a system? Would system administrators be happy to give up some of their control over systems and rely on cloud service providers for the support of end users?
- Project management is going to be affected because the authority of the IT department is going to be eroded by cloud computing. Cloud computing is increasingly turning "users into choosers" [8], and project managers can replace the services provided by the IT department with services offered in the cloud. This is already starting to happen, for example in BP, where a group bypassed the company's IT department by using AWS to host a new customer facing website[39].
- System acquisition and accounting is going to be affected because hardware and network infrastructure is not going to be procured upfront anymore; it will be consumed as a service and paid for just like a utility in the cloud.
- Security is going to be affected because virtualization introduces new vulnerabilities [62], and there could be conflicts between customers and cloud providers who are both attempting to harden their security procedures [18] (cloud providers for example do not allow port scanning, which is often used by organisations to verify security policies with regards to open ports).
- Compliance is going to be affected because the geographic location of data will not be exactly known in the cloud; this has long-term implications for enterprises concerned with data privacy [63], [22], [23], [64]. How would compliance departments react to the migration of applications and data to cloud service providers? They might not have the same level of access to a cloud as they currently do to their internal systems, so how would they have to change their working practices?
- Some enterprises have already started migrating desktop applications to the cloud [65]. Even though this seems to be one relevant use of cloud computing,

[39] http://a6.64.354a.static.theplanet.com/Individual-Case-Studies/bp-fuels-cloud-computing-interest

it can be challenging to integrate such desktop applications with legacy systems. For example, many organisations use Microsoft Excel as a front-end to access legacy systems, which often do the actual data processing. It is unclear how the migration of such desktop applications, for example from Microsoft Office to Google Docs, is going to affect the integration of these applications with legacy systems. Microsoft's cloud-based Office 365 product might provide simpler integration opportunities.

- System support is going to be affected because administrators will no longer have complete control of a system's infrastructure. Their work could increasingly involve contacting cloud providers and waiting for them to look into system problems. Such a scenario was recently reported by Jesper[40] whose application, which was running on Amazon EC2, came under a denial of service attack and had to wait over 16 hours before the problem was fixed.
- Finally, what about the work of end users? The cloud might help collaborative work but what can users do when the cloud goes down? They cannot tell Google or Amazon to prioritize their problem as they could before with their own IT department. Would end users care about this? And would they change their working practices when central IT no longer has complete control over a system?

The organisational changes will not be straightforward and will require a great deal of management effort due to the highly interconnected nature of legacy infrastructures, the political nature of IT facilitated organisational transformation, and the difficulties of aligning technical systems and organisations.

Large enterprises inevitably have highly interconnected infrastructures comprising a large number of computing systems that have been developed over a long period of time. These depend on different technologies, have different 'owners' within the enterprise and have complex dependencies both between the systems themselves, the data that they process, the middleware used and the platforms on which they run. Business processes have evolved to make use of the portfolio of systems available and these often rely on specific system features. Normally, there is no individual or group

[40] http://blog.bitbucket.org/2009/10/04/on-our-extended-downtime-amazon-and-whats-coming/

within the enterprise who knows about all of the systems that are in use, and dependencies are often discovered by accident when something simply stops working after a change has been made. For international companies, different jurisdictions mean that the same system in different countries may have to be used and supported in different ways.

Furthermore, IT provision is profoundly affected by political considerations [66], [67]. Senior management in the enterprise may set IT policies but these are left to individual parts of the enterprise to enact in their own way [68]. Managers naturally tend to adopt strategies that benefit their part of the company. Employees resist changes that originate from other parts of the organisation [69]. At the inter-group level, the tension between central IT provision and end users has been constant since the 1960s with complaints from users that central services are unwilling or unable to respond quickly to changing user requirements.

Cloud adoption decisions are challenging because of a range of practical and socio-political reasons. It is unlikely that all organisations will completely outsource their back-end computing requirements to a cloud service provider. Rather, they will establish heterogeneous computing environments based on dedicated servers, private organisational clouds and possibly more than one public cloud provider. How their application portfolio is distributed across this environment depends not only on technical issues but also on the previously described socio-technical factors, in addition to the impact on work practices and constraints derived from existing business models.

3.3 Security and Compliance

From an enterprise perspective, security and regulatory issues are critical. Inevitably and understandably, most enterprises are likely to be cautious in moving their applications to the cloud simply because they do not really understand the security and regulatory issues involved. The difficulties in understanding the issues are exacerbated by the complexity of their systems – some data may have to be maintained within a specific jurisdiction, some data may be transferable to the cloud. Even where models of existing systems, their data and their dependencies exist, such information has simply not been required up till now and collecting such information is likely to be costly.

Security is also about perceptions, many worry about security in the cloud because of a lack of control, while others argue that systems deployed in the cloud could be made as secure as systems deployed in local datacentres [2]. Some organisations might be over confident about their internal security policies, but liability issues need to be resolved before enterprises start migrating applications to the cloud. The addition of third parties into the provider-consumer relationship introduces liability issues. For example, who is going to be liable for an incident that occurs while an enterprise 'bursts out' of their private cloud onto a public cloud? Regulations relating to data protection may place constraints on the movement of data and the national jurisdictions where it may be maintained. Furthermore, the cost and time to move data around could be a major bottleneck in the cloud. Enterprises with large volumes of data may therefore wish to specify where that data should be made available, when it may be moved around, etc. The practical issues that affect data migration in the cloud need to be investigated.

Another issue with data migration is compliance, which is especially challenging to satisfy when sensitive data is involved. For example, companies in the financial and health sectors have many regulatory requirements that restrict data movement. Compliance departments are likely to be conservative in their interpretation of the regulations and will require very detailed evidence that any movement of data outside the enterprise does not have associated compliance risks.

Cloud providers are increasingly addressing the security and compliance concerns that organisations have by publishing whitepapers with details of their security practices and their compliance with different regulations. For example, AWS' security centre[41] publishes the following details:

- Certifications and accreditations including details of their audits
- Physical security practices of their large-scale datacentres
- High-level details of their system architecture and security processes
- Data privacy and backup procedures for their services

[41] http://aws.amazon.com/security/

There are also signs that cloud providers might build specialised clouds for specific industries that comply with that industry's security practices and regulations. AWS' US GovCloud[42] is one such cloud, which can only be used by government agencies and their contractors after being screened by AWS. Overall, the high-level of activity in this area indicates that these are critical issues and their urgency is such that cloud providers and users might develop negotiated solutions in the immediate future.

3.4 Support and Maintenance of IT Systems

Currently when there is a problem with computer systems, organisations have coping strategies and workarounds that often rely on local expertise and knowledge. Users are usually good at knowing who has the local expertise about a particular system in their organisation, and could ask them for help. One of the reasons organisations might move services to the cloud is to reduce IT support costs, so there will be less local expertise than before, and it is not clear who users could turn to once this expertise is lost.

In the case of IaaS, users have some control over their infrastructure and can design for failures from the outset. For example, AWS specifically advises users to develop their systems in such a way that they can deploy their system on multiple availability zones or even clouds if possible. Availability zones are distinct locations in AWS' clouds that are engineered to be independent from other zones; hence failures in one availability zone should not affect others. However, using multiple availability zones or clouds can have related compliance issues as mentioned in Section 3.3.

In the case of PaaS however, users have almost no control over the infrastructure and cannot design for infrastructure failures. This is the responsibility of the PaaS provider, and one common reason that users choose PaaS over IaaS; users want to focus on their application development and not worry about the support and maintenance of the underlying infrastructure. However, this loss of control and expertise leaves them with few options when there is an outage. Over time, users can be locked-in to the PaaS provider's platform as they lose the expertise to deploy and maintain their systems on IaaS clouds.

---

[42] http://aws.amazon.com/govcloud-us/

The deployment of applications on public clouds could mean that management will no longer be able to demand that their problem gets priority from their IT support staff. In fact, a manager's authority may be diminished because they are not part of the organisational interface with the cloud provider. The diversity in current work environments and the ability to work while disconnected, allows coping strategies to be developed that may be impossible if everyone relies on a single networked provider for their applications. This critical dependence on a cloud's availability is forcing providers to improve their communication strategy during outages. For example, after a major outage in April 2011, AWS published a detailed post mortem[43] explaining what had gone wrong, how they are planning to prevent the same issue from happening again, and how they are improving their communication strategy.

Most cloud providers have online forums as part of their support services. During failures, these forums are often used as a first point of contact and a place to discuss problems with other users as they enable user-user interaction [70]. In addition, large cloud providers such as AWS also provide different levels of 'premium support' that cost extra[44]. Organisations need to consider how the support and maintenance of their systems are affected by the adoption of cloud computing and select appropriate supporting services.

3.5 System Deployment Costs

Organisations have to consider several types of costs during cloud adoption, including IT infrastructure, system administration, software licenses and costs of 3rd party tools, systems engineering and software changes, staff costs etc. Most of these costs can be estimated with spreadsheets, and indeed cloud providers such as Amazon have created spreadsheets for cost comparisons. Furthermore, different projects will have different cost concerns. For example a cloud migration project might have significant staff costs due to software changes, whereas a company deploying a simple but scalable web application on the cloud might have significant infrastructure costs.

Understanding the operational costs of using public clouds is complicated because the cloud's utility billing model is a shift away from capital to operational budgeting. The

[43] http://aws.amazon.com/message/65648/
[44] http://aws.amazon.com/premiumsupport/

cloud's pay-as-you-go billing model has a certain degree of uncertainty that makes it non-trivial to estimate and forecast costs due to the following reasons:

1. There is uncertainty in cloud costs, which is due to actual resources consumed by a system at runtime (determined by its load). Elasticity, the ability to scale up/down the resources used by a system at runtime, is one of the key benefits of using cloud computing. However, elasticity has an associated infrastructure cost that needs to be considered during cost estimations.
2. The deployment options used by a system affect its cost as things like data transfer are more expensive between clouds compared to data transfer within clouds. There are many deployment options available in the cloud and their differences need to be considered during cost estimations.
3. There are many cloud providers worldwide, leading to thousands of prices being available for different types of infrastructure. CloudHarmony, a website that runs performance benchmarks on clouds, lists details of around 100 cloud providers worldwide[45]. Each provider can have hundreds, or even thousands, of various prices for its infrastructure services; for example AWS EC2 has at least 1,092 individual prices: 7 clouds x 13 types of instances x at least 3 OS choices (Linux, Windows, Windows with SQL Server) x at least 4 purchase options (on-demand, light/medium/heavy-utilization reserved).
4. Cloud providers can charge differently for similar types of infrastructure. For example, AWS charges for read/write requests to their Simple Storage Service (S3) while Rackspace does not charge for this in their CloudFiles storage services (similar to S3).
5. Cloud providers can change their pricing schemes (what they charge for) and their prices at any time. For example, Microsoft switched to a new pricing model for their SQL Azure database service in February 2012[46], and AWS have changed their prices 19 times since they launched EC2 in 2006[47].

[45] http://cloudharmony.com/clouds
[46] http://blogs.msdn.com/b/cbiyikoglu/archive/2012/02/16/the-new-pricing-model-for-sql-azure-explained.aspx
[47] http://aws.typepad.com/aws/2012/03/dropping-prices-again-ec2-rds-emr-and-elasticache.html

The consequence is that decision makers are faced with uncertainty regarding their cloud cost estimates and whether cloud adoption is more cost effective than other, more traditional, forms of IT provisioning such as co-location. In many organisations, procurement costs have to be known in advance before approval can be gained. Furthermore, specific signatories may be required to approve procurement and this mitigates against the use of on-demand systems. Organisations need to know their costs to manage their cash flow and cost uncertainty is often not regarded as an advantage.

The following chapter (Chapter 4) builds on the issues identified in this chapter, and describes how a system was developed to provide enterprises with a starting point for risk assessment. Chapter 5 describes how the cost estimation challenges described in this section were addressed by developing a cost modelling tool that provides organisations with cost estimates of deploying their system on public clouds.

4 Benefits and Risks of Cloud Migration

Public clouds are often marketed as scalable, reliable, secure and cost-effective deployment options for IT systems. However, there is much hype surrounding cloud computing[48] and uncertainty regarding the actual realization of the marketed benefits. For example, the cost savings of using public clouds that are often cited by cloud providers have to be examined in the wider context of other benefits and risks. It can be difficult or meaningless to quantify indirect cost savings of, say, the improved time-to-market or flexibility provided by using public clouds.

The work presented in this chapter is mainly aimed at enterprises that are considering the migration of their IT systems to the cloud rather than start-ups with new systems. From an enterprise perspective, costs are important but so too are customer relationships, public image, flexibility, business continuity and compliance. Therefore, the various benefits and risks need to be identified and brought together, so that decision makers can be informed and supported in starting a discussion regarding the benefits and risks of using the cloud in their organisation.

This thesis uses the following definition of a benefit:

an advantage to an organisation over its status quo provided by using public clouds.

The ISO/IEC Guide 73:2002 defines risk as the "*combination of the probability of an event and its consequence*". In this thesis, risks are seen as undesirable events that might occur if public clouds are used.

The general benefits and risks of using public clouds were identified by reviewing over 50 academic papers and industry reports. The identified benefits and risks were categorised as organisational, legal, security, technical or financial and were then listed in a spreadsheet. The categories were taken and adapted from [71]. This spreadsheet provides a starting point for risk assessment as it identifies the main risks and describes their potential consequences as well as some mitigation approaches.

[48] The Gartner Hype Cycle (http://www.gartner.com/it/page.jsp?id=1763814) recently put cloud computing in the "Peak of Inflated Expectations" phase.

The spreadsheet includes benefits and risks that were initially identified in the IT outsourcing literature but could equally apply to cloud adoption as using public clouds can be seen as a form of IT outsourcing. There is even similar marketing material being used to encourage organisations to use cloud computing, for example Loh & Venkatraman [72] mention that in 1990 when Martin Marietta Information Systems Group (now called Lockheed Martin) ran an advert to promote their IT outsourcing offerings they used the phrase "You don't own a power plant for your electricity… Why own a data centre for your information systems?"

The IT outsourcing literature has shown that IT outsourcing does not always meet expectations. Lacity and Hirschheim [73] analyzed the extent to which the expectations of fourteen Fortune 500 companies engaged in IT outsourcing were met. The study revealed that many of the outsourcing success stories portrayed in the literature painted an inaccurate picture. For example, outsourcing contracts that promised clients cost savings of 10 to 50% of their IS costs over the life-time of the contract were often anticipated savings that were not actually achieved. Lacity and Hirschheim [73] found that, in some cases, in-house IT departments could make similar cost savings through standardization, consolidation and internal charge-back mechanisms, but were prevented from implementing such strategies by internal politics and organisational culture.

Dibbern et al. [74] wonder if IT outsourcing is "nothing more than a pendulum" that started with organisations creating internal IS departments only to realize that outsourcing can be more beneficial, but after going through several outsourcing contracts, discovered that it is unsatisfactory. Therefore, they are bringing their outsourced IT systems back in-house due to poor service levels, changes in strategic direction and failed cost saving promises [75].

Cloud computing can be seen as another swing in the IT outsourcing pendulum, where organisations are seeking to outsource their IT infrastructure due to the potentially cost effective, scalable and reliable services provided by cloud providers. However, there is a key difference between cloud computing and IT outsourcing: the cloud's self-service model and the lack of fixed long-term contracts give more control and flexibility to clients compared to traditional IT outsourcing. Nevertheless, for cloud computing not to result in another swing-back of the IT outsourcing pendulum, it is important that organisations consider the risks of cloud migration before committing to external cloud service contracts.

This chapter starts by describing the Cloud Suitability Checklist, a simple checklist that can be used to assess the suitability of public clouds for IT system deployment (Section 4.1). The benefits and risks of cloud migration (described in Sections 4.2 and 4.3) can be further investigated by organisations if there are no major barriers identified by the Cloud Suitability Checklist. Section 4.4 describes how the identified benefits and risks were used to create a system to provide organisations with a starting point for risk assessment.

4.1 Cloud Suitability Checklist

The purpose of the Cloud Suitability Checklist is to support decision makers in determining whether public clouds provide a suitable deployment option for their IT system. Understanding the characteristics of cloud computing is important as it has radically different properties to those of traditional enterprise datacentres. This is mainly due to the cloud's highly scalable nature, physical resource sharing between virtual machines, potential issues to do with data transfer over the internet and insufficient guarantees regarding up-time and reliability of processing and data storage services. For example, typical public IaaS clouds give no reassuring guarantees about server uptime or network performance. This has important implications for the viability of certain classes of software architectures and business-critical systems.

The Cloud Suitability Checklist comprises a simple list of questions to provide a rapid assessment of the suitability of public IaaS clouds for a specific IT system. The checklist, partially shown in Figure 7, is available online as a web-based form that can also be printed and used offline (fits on one page[49]). The complete checklist is shown in Table 4 and covers eight main areas of potential concern. The outcome of the analysis is a recommendation of whether or not to proceed with further analysis of the benefits and risks of using cloud computing.

[49] https://docs.google.com/open?id=0B6Xeif9FIjfLazgyTFcxSXFBRFU

ShopForCloud.com - Cloud Suitability Checklist

Overview
ShopForCloud.com provides tools to support the assessment of the benefits, risks and costs of using cloud computing. This checklist is one of the tools; it comprises a simple list of questions to provide a rapid assessment of the suitability of public clouds for your IT systems. The outcome of the analysis is a recommendation of whether or not to proceed with further analysis of the benefits and risks of using cloud computing.

1. Elasticity
- Does your software architecture support scaling out to multiple servers?
- If not, does your cloud provider support a suitability server size such that you can scale-up your application to a bigger server?
Notes:

2. Network
- Is the bandwidth between your systems/users and the cloud sufficient for your system?
- Is the latency of data transfer to the cloud acceptable?
Notes:

3. Processing
- Is the performance of servers appropriate for your application at the expected operating load?
- Do the servers have enough memory for your application?
Notes:

4. Access to hardware
- Does your system require special access to hardware components? If so, does your cloud provider support this?
Notes:

5. Availability and dependability
- Does your cloud provider provide an appropriate SLA?
- Are you able to create the appropriate availability by mixing geographical locations or cloud providers?
Notes:

Figure 6: Screenshot of the cloud suitability checklist

Although the Cloud Suitability Checklist is not exhaustive, it does cover the main issues that the users should focus on during their decision-making process. For example, questions 1–3 cover technical properties of the cloud and if answers to these questions are negative, then it would be challenging to take full advantage of cloud computing [76]. Negative answers to questions 5–7 would increase the risks of using the cloud [18], whereas negative answers to questions 4 and 8 might inhibit the use of cloud computing altogether. This is because most cloud providers do not provide access to hardware components, and not complying with regulatory requirements could lead to legal disputes.

Table 4: Cloud Suitability Checklist

Category	**Questions**
1. Elasticity	- Does your software architecture support scaling out to multiple servers? - If not, does your cloud provider support a suitable server size

	so you can scale-up your applications to a bigger server?
2. Network	- Is the bandwidth between your systems/users and the cloud sufficient for your system? - Is latency of data transfer to the cloud acceptable?
3. Processing	- Is the performance of servers appropriate for your application at the expected operating load? - Do the servers have enough memory for your application?
4. Access to hardware	- Does your system require special access to hardware components? If so, does your cloud provider support this?
5. Availability and dependability	- Does your cloud provider provide an appropriate SLA? - Are you able to create the appropriate availability by mixing geographical locations or cloud providers?
6. Security	- Does your cloud service provider meet your security requirements? (e.g. do they support multi-factor authentication or encrypted data transfer)
7. Data confidentiality and privacy	- Does your cloud provider provide sufficient data confidentiality and privacy guarantees?
8. Regulations and compliance	- Does your cloud provider comply with the required regulatory requirements of your organisation and systems?

The checklist is aimed at roles in an organisation that are responsible for system infrastructure procurement decisions, such as IT managers or CTOs. Depending on the type of organisation, such people might need information and input from others when investigation the issues. For example, they might need input from system administrators about system performance and access to hardware, or input from security directors or engineers about compliance and security requirements when discussing questions 6-8. Overall, the checklist is designed to inform decision makers about the issues rather than act as strict questionnaire with a mandated methodology for use.

4.2 Benefits of Using the Cloud

The following sections describe the benefits of using cloud computing. Each section starts by highlighting the main benefits in its category before describing the complete list in a table.

4.2.1 Technical

Elasticity and the ability to address volatile demand patterns is a key technical benefits of using public clouds (B3). This shifts the risks of over/under provisioning of infrastructure from organisations to cloud providers, and enables new businesses to start with a limited amount of infrastructure and grow when required. Public clouds have APIs that can be used by organisations directly; this is another key technical benefits of using public clouds as it gives organisations quick access to additional computation resources (B1) and simplifies their system deployment processes (B7).

Table 5: Technical benefits of using cloud computing

ID	Benefit	References
B1	Quicker access to additional computational resources and specialized skills (e.g. IT specialists who build and maintain clouds).	[74], [77–79]
B2	Access to temporary computational resources that can be used for prototyping of new systems or parallel operations of systems. Systems can be installed on a public cloud for experimentation then migrated to an internal server for operation.	[80]
B3	Ability to address volatile demand patterns and the flexibility to scale-up/down resource usage without discontinuity or service interruption. Reduced risk of over/under provisioning infrastructure resources.	[2], [4], [74], [76], [81]
B4	Reduced run time and response time due to the ability to acquire vast computational resources for short time periods, e.g. a batch job that takes 1000 hours could, in principle, be done in 1 hour using 1000 servers for the similar costs. This can lead to a reduced time to market.	[4], [76], [81], [82]

B5	Anywhere/anytime/any device (desktop, laptop, mobile etc.) access to computational resources and applications can be setup without too much effort. This in turn simplifies collaboration amongst users and simplifies application support and maintenance.	[83], [84]
B6	Increased system security due to more investment into security by cloud providers (e.g. specialised security teams, greater resilience, protection against network attacks, and quicker disaster recovery procedures).	[85], [86]
B7	Simplified and faster system deployment due to automation of resource provisioning APIs and auto-scaling.	[1], [76], [84]
B8	Simplified and cheaper provisioning of disaster recovery and business continuity plans due to geo-distribution and replication facilities provided by cloud providers.	[76], [87]
B9	Simpler implementation of variable access control policies through simpler partitioning of systems; related to the benefits of simplified sharing with partners and collaborators (B4).	

4.2.2 Organisational

One of the key organisational benefits of using public clouds is that it frees organisations from having to worry about setting up and maintaining reliable IT infrastructure (B11). This in turn enables management and IT personnel to focus on value-added activities such as improving their IT systems and gives organisations the opportunity to trial new products to gauge customer interest (B13).

Table 6: Organisational benefits of using cloud computing

ID	Benefit	References
B10	Vehicle for organisational change (e.g. to eliminate a troublesome function in the organisation).	[74], [78]
B11	Ability to focus on core business activities and free-up management and IT personnel from mundane tasks (such as hardware support activities) so that they can focus on value-added activities.	[2], [74], [77], [79]

B12	Better satisfaction of work and an improved status for some people in the organisation due to removal of tedious system administration work such as hardware maintenance.	[13]
B13	Opportunity to offer new products or services or trial products to gauge the level of interest from customers.	[13]
B14	Devolution of decision making on IT infrastructure requirements to operational units. Variable provision in different parts of the organisation (this could also be a risk)	[80]
B15	Reduced risks of technological obsolescence as cloud providers update the infrastructure.	[74]

4.2.3 Financial

The key financial benefit of using public clouds is the transformation of capital expenditure (spent on IT infrastructure) to operational expenditure (B18), and the potential to reduce costs due to the economies of scale that are achieved by cloud providers (B16). It should be noted that potential cost savings are dependant on the usage patterns of the cloud infrastructure, which is discussed in Chapter 5.

Table 7: Financial benefits of using cloud computing

ID	Benefit	References
B16	Reduced costs due to more efficient operations and less infrastructure maintenance costs but also due to economies of scale that can be achieved by cloud providers.	[74], [77], [86]
B17	Reduced energy consumption due to IT infrastructure being moved to cloud providers, can lead to greener organisations.	[82], [85], [86]
B18	Reduced need for capital investment and the ability to transform fixed costs into variable costs. This can simplify cash-flow management.	[13], [74], [76], [81], [82]
B19	Reduction in physical space requirements leading to lower real estate costs.	
B20	Competitive advantage from the use of on-demand pricing of IT infrastructure, this can lead to financial benefits as it enables	

	organisations to create on-demand pricing models for their products.	

4.3 Risks of Using the Cloud

4.3.1 Technical

The major technical risk of using public clouds is major outages (R1), which have occurred on numerous occasions in the past. During such outages, there is very little that can be done by organisations, as they do not have control or visibility of what is happening inside the cloud. Using multiple clouds or providers can mitigate this risk; however, this can be technically challenging and issues around interoperability and compatibility arise in such scenarios (R5 and R6). Public clouds are shared resources and can exhibit variable performance, thus risks around performance (R3, R8 and R9) might also cause issues for systems where performance is key.

Table 8: Technical risks of using cloud computing

ID	Risk Mitigation approaches	References
R1	Major service interruption resulting in extensive outages and unavailability of services or loss of data.	[4], [13], [83], [88], [89]
	Use multiple cloud providers, monitor applications from outside the cloud. Replicating the system across multiple clouds has associated costs and technical challenges.	
R2	Data lock-in for SaaS/PaaS and system lock-in for IaaS.	[4], [18], [78]
	Mitigate IaaS lock-in risk by using middleware that is compatible with multiple clouds (e.g. RightScale). See [18] (p.26) for other mitigation approaches. Indicator: lack of interest from providers to participate in standardization efforts.	
R3	Performance is worse than expected (e.g. CPU clock rate, I/O and network data transfer and latency rates). It might be difficult to prove to the cloud provider that their system performance is not as good as they promised in their SLA as the	[4], [78], [90–92]

	workload of the servers and the network can be highly variable in a cloud. This might lead to disputes and litigation.	
	Mitigate risk by investigating performance of the cloud under investigation before making decision (using performance benchmarks). Rent more VMs or higher spec ones to deal with slow CPU clock rates, and where suitable, use physical disk shipping to reduce effects of network latency/transfer rates. Use third party monitoring tools to independently verify the system performance. Switch to other cloud providers that offer better performance.	
R4	Lack of resources in the cloud that might lead to the inability of the provider to serve current or future resource demands.	[18]
	Reserve instances up-front, use multiple cloud providers.	
R5	Cloud APIs and software features could change over time resulting in incompatibilities between the customer's system and the cloud.	[89]
	For IaaS, use cloud middleware such as RightScale or software libraries such as Fog.io to reduce the number of cloud APIs you need to deal with.	
R6	Interoperability issues between clouds as there are incompatibilities between cloud providers' platforms.	[18]
	Use cloud middleware to ease interoperability issues.	
R7	Cloud providers failing to keep-up with technological improvements.	[74]
	Monitor system performance over time and switch to other cloud providers that offer better performance. Indicator: Performance getting gradually worse.	
R8	Network performance could degrade over time as more and more users start to use the cloud.	[90]
	Duplicate system on another cloud and keep as a stand-by node, but this has associated costs. Monitor network performance and response times from outside of the cloud.	
R9	Uncontrollable sources of data transfer delay or bottlenecks.	

	National broadband differences in prices and services becomes important for multinational enterprises.	
	Where suitable, use physical disk shipping for large data transfers, but not all cloud providers support this. Use cloud providers with local datacentres.	

4.3.2 Organisational

One of the main organisational risks in using public clouds is the loss of governance and control over computing resources (R10) as users can bypass their central IT and use cloud-based services. There are also a number of risks related to the support and maintenance of systems as using public clouds increases the dependence of organisations on third parties (R15, R16, R17 and R22). The usual risks of dealing with new technologies and resistance to change should also be considered by organisations (R18 and R20).

Table 9: Organisational risks of using cloud computing

ID	Risk **Mitigation approaches**	References
R10	Loss of governance and control over resources (both physical control and managerial), might lead to unclear roles and responsibilities, for example, users can purchase computing resources using their credit cards without explicit approval from central IT.	[13], [18], [74], [78], [80]
	Clarify roles and responsibilities before cloud adoption.	
R11	Loss of IT expertise (e.g. system admins), which can in the long-term limit an organisations ability to grow and acquire new systems due to the lack of local expertise and erosion of knowledge. This will also make it difficult to bring the system back in-house or to migrate the system from one cloud to another if the provider's service levels are unsatisfactory.	[78], [80], [91–93]
	Perform an expertise audit and assess the expertise that should be retained.	

R12	Reduced staff productivity during the migration as changes to staff work (e.g. staff getting less satisfying work) and job uncertainty (or departmental downsizing) leads to low staff morale and anxiety spreading in the organisation.	[13], [79]
	Ensure that experts are not dismissed and involve them in the migration project so that they get a sense of ownership. Provide training in cloud technology and enable staff to learn new skills. Indicator: Rumours spreading in the organisation about future job uncertainty.	
R13	Lack of organisational learning. For example, by using SaaS, the organisation might miss the opportunity to learn from commodity software applications that might over time become strategically important. There's also a risk that the organisation relies upon the cloud provider to innovate and loses its innovative capacity.	[92], [93]
R14	Managing a system deployed on several clouds might take extra management effort compared to deploying systems in-house (e.g. to manage the relationship with the cloud providers, deal with problems, deal with changes to cloud services). This is one of the hidden costs of deploying systems on the cloud.	[74], [92], [93]
	Make management aware of the extra effort that might be required. Indicator: Using several cloud providers for a system, cloud providers having different types of support mechanisms.	
R15	Loss of business reputation due to malicious activities carried out by co-tenants (e.g. spamming, port scanning, crashing servers).	[18]
	When using IaaS, follow Amazon Web Service's security guidelines.	
R16	Changes to cloud providers' services (e.g. they terminate a service) or they are acquired by another company that changes services.	[18]
	Use multiple cloud providers.	

R17	Deterioration of customer care and service quality due to an increased dependence on third parties and loss of governance and control over systems (related to risks R10 and R16).	[13]
	Monitor the organisation's service levels. Purchase premium support services from cloud providers.	
R18	Uncertainty with new technology and a lack of supporting resources to resolve technical problems.	[13], [88]
	Investigate support mechanisms of the cloud provider, purchase premium support services.	
R19	Cloud provider could go out of business.	[18]
	Use multiple cloud providers, backup data outside of the cloud.	
R20	Resistance to change resulting from organisational politics and changes to people's work.	[2], [13], [80]
	Use insights from organisational change management and involve key stakeholders in the adoption process	
R21	Private data being exposed due to a change of responsibility for users who have a lack of awareness about where to put different types of data.	[94]
	Avoid putting sensitive data in public clouds. Increase awareness and train users about which types of data they can put on public clouds as they ultimately will be responsible for their decisions (even if firewall protection is in place).	
R22	Mismatch between existing incident handling procedures and cloud providers' procedures. Lack of information or no access to a cloud's vulnerability information or incident report data. Leads to limited responses from an organisation in case of incidents.	[27]
	Check cloud provider's SLA and ensure that it has well-defined incident classification schemes and reporting procedures (e.g. what is reported, how fast it is reported, to whom it is reported). Indicator: Lack of incident handling information in the SLA.	

4.3.3 Financial

The major financial risk of using public clouds is that the actual costs might be higher than estimates (R23) due to the pay-as-you-go pricing models used by cloud providers, and hidden costs related to system integration and unrealistic expectations of support staff reductions (R25). The cost modelling tool presented in the next chapter can be used to obtain estimates and make organisations aware of their infrastructure costs.

Table 10: Financial risks of using cloud computing

ID	Risk / Mitigation approaches	References
	Risk	
	Mitigation approaches	
R23	Actual costs may be different from estimates, this can be caused by inaccurate resource estimates, cloud providers changing their prices, or inferior performance (e.g. due to over-utilised servers) resulting in the need for more computation resources than expected.	[89], [92], [95]
	Monitor existing resource usage and use estimation tools to obtain accurate cost estimates of deploying IT systems on the cloud. Check results of performance benchmarks.	
R24	The costs of switching from one cloud provider to another provider could be quite high. This could also take a long time depending on the data volumes involved and be complicated due to incompatibilities between the cloud provider's platforms.	[91], [92]
	Use cloud management systems to reduce migration efforts. Such systems help to de-couple a system from the clouds that it uses, and hence, reduce the effort required to migrate a system.	
R25	Increased costs due to complex system integration problems between existing systems and cloud-based systems. Inability to reduce costs due to unrealisable reductions in the number of system support staff (e.g. due to their knowledge of existing systems).	[77], [84]
	Investigate system integration issues upfront, avoid migrating highly interconnected systems initially and, where possible,	

	integrate system incrementally.	

4.3.4 Security

Most of the security risks of using cloud computing are not particularly new or unique to cloud computing (e.g. denial of service attacks, R26), however, the widespread use of public clouds could make them more significant. One example of this is the risks posed by browser vulnerabilities (R30) as most clouds provide web dashboards that can be used to control the infrastructure. Cloud providers have the resources and expertise to invest in high levels of security and monitoring. Smaller organisations that manage their own servers may not have access to such resources, and so may be running infrastructure with security vulnerabilities. In such organisations, using the cloud may therefore improve overall security.

AWS has published a set of security best practices for users of its public clouds [96], which can also be applicable for other public IaaS clouds. These guidelines include advice around protecting data in transit (e.g. using SSL) and at rest (e.g. using file encryption), protecting AWS credentials (e.g. using multi-factor authentication) and general guidelines around securing applications (e.g. using security groups appropriately).

Table 11: Security risks of using cloud computing

ID	Risk	References
	Mitigation approaches	
R26	Denial of service attacks. Leads to unavailability of resources and increases cloud usage bills.	[15], [18], [62], [89]
	Use network monitoring tools (although some providers do not allow this) or monitor applications from outside of the cloud. Armbrust et al. [4] argues that although denial of service (DOS) attacks can be a problem in the cloud, cloud providers are likely to have DOS protection as a core competency.	
R27	Private data could be accessed by other customers or cloud providers. Various security vulnerabilities that can be used to	[4], [19], [63]

	extract information from a target VM are discussed in the academic literature. However, such attacks are unlikely to be seen in practice as cloud providers take such vulnerabilities very seriously.	
	Use technologies such as encryption. Use large instance sizes to ensure the physical server is not shared with other instances (e.g. use Extra Large Instances in AWS) however, this could significantly increase costs. Follow AWS's security guidelines.	
R28	Interception of infrastructure management (API) messages and data in transit. This could lead to the infrastructure being manipulated by third parties.	[18]
	Use secure communication protocols and multi-factor authentication. Follow AWS's security guidelines.	
R29	Insecure or ineffective deletion of data when scaling down resource usage or when changing providers.	[18]
	Use encrypted data storage to reduce risks, use special procedures to delete data. Encryption has associated costs and risks, such as the need for a key management solution and loss of encryption keys.	
R30	Browser vulnerabilities become more significant, specially when using SaaS.	[15], [21]
	Put in strict browser update policies to ensure security patches are deployed in a timely manner.	

4.3.5 Legal

The main legal risks of using public clouds are related to compliance with regulations that are applicable to an organisation and its IT systems (R32-R35 and R37). This is affected the by the jurisdiction that is applicable for a given cloud and its users (R34).

Table 12: Legal risks of using cloud computing

ID	Risk	References
	Mitigation approaches	

R31	Unusable software license on the cloud due to the license using traditional per-CPU licensing agreements or needing physical software locks etc.	[18], [97]
	Check that all of the required software has appropriate license agreements.	
R32	Non-compliance with regulations that require informed consent from users when dealing with personal data (e.g. the first principle in the UK's Data Protection Act 1998).	[22]
	Avoid risk by explicitly getting users to give consent. This is not a cloud-specific risk.	
R33	Non-compliance with regulation that requires strict access mechanisms such as using more than a username/password to restrict access when dealing with personal data (e.g. the seventh principle in the UK's Data Protection Act 1998).	[22]
	Avoid risk by using additional mechanisms to restrict access (e.g. AWS offers multi-factor authentication). This is not a cloud-specific risk.	
R34	Lack of information on jurisdictions used for data storage and processing. Leads to non-compliance with regulations that require certain types of data to be kept in national boundaries (e.g. the eighth principle in the UK's Data Protection Act 1998 that requires personal data to be kept within the EEA).	[22]
	Avoid risk by using data centres within the required jurisdiction. Indicator: Cloud providers that don't disclose the country of their data centres.	
R35	Non-compliance with data confidentiality regulations. Unauthorised access to data by cloud providers.	[4], [22], [63], [89]
	Avoid risk by checking cloud providers' service agreements and using encrypted data transfer and storage. Follow AWS's security guidelines. Indicator: Cloud providers not supporting encrypted data transfer.	
R36	Losing some of the intellectual property (IP) rights over a system due to the use of a cloud (e.g. in the UK there are	[18], [22], [74]

	Database right and Copyright issues). Increased risk of IP rights being violated.	
	Investigate IP rights with cloud provider. Indicator: Lack of clarity in providers' terms of use regarding ownership of IP rights.	
R37	Non-compliance with industry regulations, such as the Financial Services Authority regulations in the UK, and the following regulations in the US: Health Insurance Portability and Accountability Act (HIPAA), Federal Information Security Management Act (FISMA), Payment Card Industry Data Security Standards (PCI DSS), Sarbanes-Oxley (SOX) and Statement on Auditing Standards No 70 (SAS 70).	[18], [22], [74]
	Check compliance with the auditors and cloud providers. When using IaaS and dealing with data that is protected by HIPAA, follow AWS's guidelines. Indicator: Lack of regulated customers using the cloud.	
R38	Obsolete contractual protections that do not apply to cloud computing.	[22]
	Use customised contractual agreements with providers. Refer to Bradshaw et al. [98] for a detailed review of standard contracts that are often used by cloud providers. Indicator: Using outsourcing contractual agreements for cloud projects.	
R39	Private data stored on the cloud can be accessed by foreign governments due to differences in jurisdictions.	[98]
	Check the terms of service of the cloud provider to find out which jurisdiction applies. 'Safe Harbour' policy agreements between countries such as the US and European Union also need to be considered.	

These risks highlight the importance of compliance with regulations when using clouds; these general issues need to be taken into account in service agreements between cloud providers and user organisations.

4.4 Supporting Risk Assessment

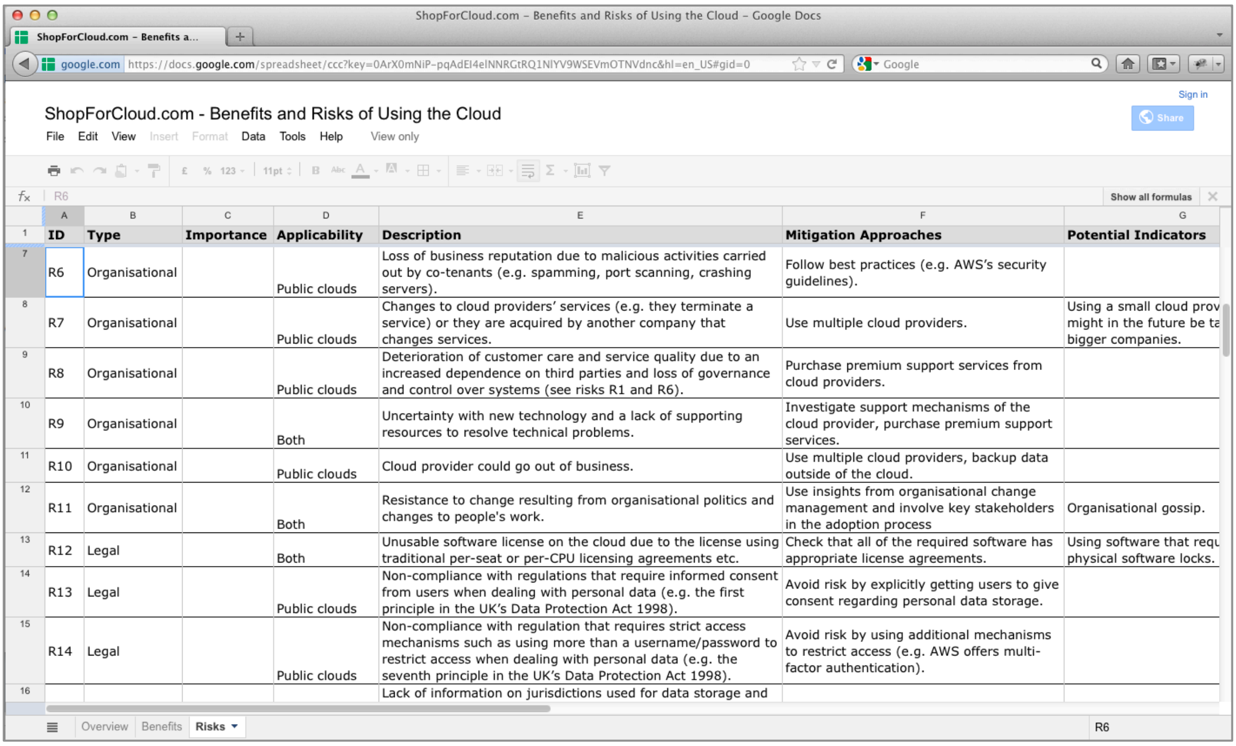
As part of a comprehensive analysis and survey of the IT outsourcing literature, Dibbern et al. [74] point out that initially, researchers in IT outsourcing were concerned with *why* organisations outsourced their IT (e.g. major factors involved in the decisions, advantages and disadvantages of outsourcing). Later on, researchers focused on *how* organisations outsourced their IT (e.g. evaluating different vendors and structuring IT outsourcing contracts). The work presented in this chapter is similar in nature to the early IT outsourcing research in identifying the advantages (benefits) and disadvantages (risks) of using public clouds.

Public clouds are platforms for IT system deployment, and the benefits and risks of using them vary depending on the organisations that use them and their systems. Therefore, organisations need to have a discussion about the benefits and risks of using public clouds. The aim of the benefits and risks spreadsheet is to support organisations in starting those discussions by informing decision makers about the generic benefits and risks of using public clouds.

The spreadsheet was used in the case studies presented in this thesis, which are discussed in Chapter 7. Each stakeholder involved in a case study was asked to make a copy of the online spreadsheet, read through the benefits/risks and set their importance (described next) from their perspective. The stakeholders were then asked to arrange a meeting to discuss the top 5 or 10 benefits and risks from their perspectives.

The spreadsheet is not meant to be used to directly drive cloud adoption decisions; rather it serves to highlight the important issues that should be considered by organisations during their risk assessment process. The benefits and risks spreadsheet was made available online on the ShopForCloud blog[50] and has so far been downloaded more than 230 times. Figure 7 provides a screenshot of the spreadsheet, which has been implemented using Google Docs.

[50] The spreadsheet is available from https://sites.google.com/site/alikhajeh1



ID	Type	Importance	Applicability	Description	Mitigation Approaches	Potential Indicators
R6	Organisational		Public clouds	Loss of business reputation due to malicious activities carried out by co-tenants (e.g. spamming, port scanning, crashing servers).	Follow best practices (e.g. AWS's security guidelines).	
R7	Organisational		Public clouds	Changes to cloud providers' services (e.g. they terminate a service) or they are acquired by another company that changes services.	Use multiple cloud providers.	Using a small cloud prov might in the future be ta bigger companies.
R8	Organisational		Public clouds	Deterioration of customer care and service quality due to an increased dependence on third parties and loss of governance and control over systems (see risks R1 and R6).	Purchase premium support services from cloud providers.	
R9	Organisational		Both	Uncertainty with new technology and a lack of supporting resources to resolve technical problems.	Investigate support mechanisms of the cloud provider, purchase premium support services.	
R10	Organisational		Public clouds	Cloud provider could go out of business.	Use multiple cloud providers, backup data outside of the cloud.	
R11	Organisational		Both	Resistance to change resulting from organisational politics and changes to people's work.	Use insights from organisational change management and involve key stakeholders in the adoption process	Organisational gossip.
R12	Legal		Both	Unusable software license on the cloud due to the license using traditional per-seat or per-CPU licensing agreements etc.	Check that all of the required software has appropriate license agreements.	Using software that requ physical software locks.
R13	Legal		Public clouds	Non-compliance with regulations that require informed consent from users when dealing with personal data (e.g. the first principle in the UK's Data Protection Act 1998).	Avoid risk by explicitly getting users to give consent regarding personal data storage.	
R14	Legal		Public clouds	Non-compliance with regulation that requires strict access mechanisms such as using more than a username/password to restrict access when dealing with personal data (e.g. the seventh principle in the UK's Data Protection Act 1998).	Avoid risk by using additional mechanisms to restrict access (e.g. AWS offers multi-factor authentication).	
				Lack of information on jurisdictions used for data storage and		

Figure 7: Screenshot of the spreadsheet

Risk assessment typically includes risk identification, analysis, planning and monitoring [71]. Risk analysis involves assessing each of the identified risks and assigning them probability, impact and exposure values:

- **Probability**: Probability of the occurrence of the risk, which can be one of the following:
    - Very Unlikely = Very Low or <10%
    - Unlikely = Low or 10-25%
    - Possible = Medium or 25-50%
    - Likely = High or 50-75%
    - Very Likely = Very High or >75%
- **Impact**: The loss to the organisation or its impact on the organisation if the risk occurs, which can be one of the following:
    - Insignificant = Very Low
    - Noticeable = Low
    - Tolerable = Medium
    - Serious = High
    - Catastrophic = Very High
- **Exposure**: Exposure is estimated based on the probability and loss of a risk using tables similar to Table 13 (taken and adapted from [18], [71]). Low risks

(yellow) have exposure values of 1-3, medium risks (orange) have exposure values of 4-6 and high risks (red) have exposure values of 7-9. The risks can be sorted using their exposure values and the organisation should focus on the high-exposure risks first.

Table 13: Risk exposure

	Probability				
Impact	**Very Unlikely**	**Unlikely**	**Possible**	**Likely**	**Very Likely**
Insignificant	*1*	*2*	*3*	*4*	*5*
Noticeable	*2*	*3*	*4*	*5*	*6*
Tolerable	*3*	*4*	*5*	*6*	*7*
Serious	*4*	*5*	*6*	*7*	*8*
Catastrophic	*5*	*6*	*7*	*8*	*9*

Traditional risk analysis involves assigning values to the probability and impact of risks to calculate their exposure. The risks are then ranked using their exposure and organisations focus on mitigating high-exposure risks. However, for the purposes of the spreadsheet, a single *Importance* column provides a similar prioritisation mechanism, and is simpler for users as they do not have to guess the probability and impact values. Decision makers can go through the spreadsheet and rate each item as unimportant, marginally important, moderately important, important, or very important from their perspective. This type of scaling is called the Likert scale and is often used in survey-based research.

In larger organisations, several stakeholders from different departments might be involved in cloud adoption decisions (e.g. central IT, accounting, compliance departments). Such stakeholders are likely to have different perspectives on which benefits/risks are important and in these scenarios, each stakeholder would use the spreadsheet individually before meeting the others to discuss their views. In such cases, to provide a holistic picture of the benefits and risk from an organisation's perspective, the weighted average of the benefits/risks can be calculated and charted on a radar graph (or spider chart). The weighted average can be calculated by multiplying the number of benefits/risks in each category (organisational, legal, security, technical or financial) by the weight of each benefit/risk (unimportant = 1 …

very important = 5), and dividing the result by the total number of benefits/risks in that category. This calculation is simple enough to be done manually and the tool does not perform it automatically.

The benefits and risks spreadsheet has been used in several case studies, which are described in Chapter 7.

5 Elastic Cost Modelling

Estimating the infrastructure costs of deploying IT systems on the cloud presents a number of challenges, as the clouds utility billing model is very different from the traditional infrastructure procurement methods that enterprises normally use. These challenges (described in Section 3.5) can be summarised as the need to take into account the following considerations during cost estimation: elasticity of the system's infrastructure; the deployment options used by the system; the differences between the cloud providers pricing schemes, and changes that they make to their prices.

A major contribution of the work presented in this thesis was to address these challenges by developing a cost modelling tool, and support enterprises in answering the following questions:

1. How much does it cost to deploy IT systems on public clouds?
2. How can a breakdown of the costs be obtained (useful for cost optimizations)?
3. How would different elasticity scenarios affect the costs?
4. How can different deployment options and design alternatives be compared in terms of their infrastructure costs?
5. How can the costs of using different cloud providers be compared for a given system?
6. How would the costs be affected by changes in the cloud provider's prices, and how can the costs be re-estimated quickly when they do change their prices?

The technical nature of cloud cost estimation means that any tool has to be mainly targeted at technical roles in enterprises, such as software architects, engineers, IT managers and consultants. However, cost estimates produced by such tools can also be used by project managers and finance departments as part of other tasks, such as cash flow management or budget estimation. Figure 8 shows an overview of the steps involved in cloud cost modelling:

1. Users describe the infrastructure requirements of their system using a tool (e.g. the type and number of servers they need).
2. The tool runs a simulation that produces a cost report for the system described in Step 1; this simulation uses the tool's database of the prices from cloud

providers. A ‘simulation’ here refers to a program that takes the model produced in Step 1 and the live prices from cloud providers as input, and produces a monthly cost report by stepping through each month between the required report dates to calculate the monthly costs of the system.

3. Users can repeat Steps 1 and 2 and investigate the previously mentioned questions; for example they can change their requirements described in Step 1 to compare different deployment options or cloud providers.

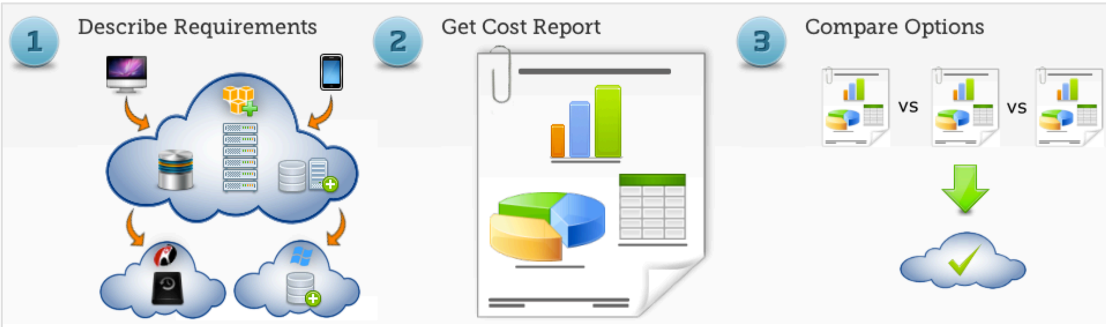


Figure 8: Overview of cloud cost modelling

This chapter starts by describing the main modelling notations that were developed as part of the cost modelling tool (Section 5.1). A unique feature of this tool is its ability to take into account the effects of elasticity as part the cost estimation. This is enabled using the concept of *Elasticity Patterns*, which is introduced in Section 5.2. A working prototype of the tool was developed as part of the tool’s early design stages. This prototype, described in Section 5.3, was used during a number of industrial case studies to gather feedback (Chapter 7), and inform the development of ShopForCloud.com (Chapter 6).

5.1 Modelling Notations

The cost modelling tool focuses on estimating deployment costs (i.e. the infrastructure costs) that have to be paid to cloud providers. To do this, the tool has to enable users to create a model of their system deployment at a level of detail that is suitable for cost modelling. Figure 9 shows an overview of the modelling notations that were developed as part of the tool:

Deployment: a deployment is simply a container and represents a group of servers, storage, databases etc. that belong to a system. Deployments can be cloned and changed to compare different cloud providers and infrastructure options. The tool generates a cost report for each deployment.

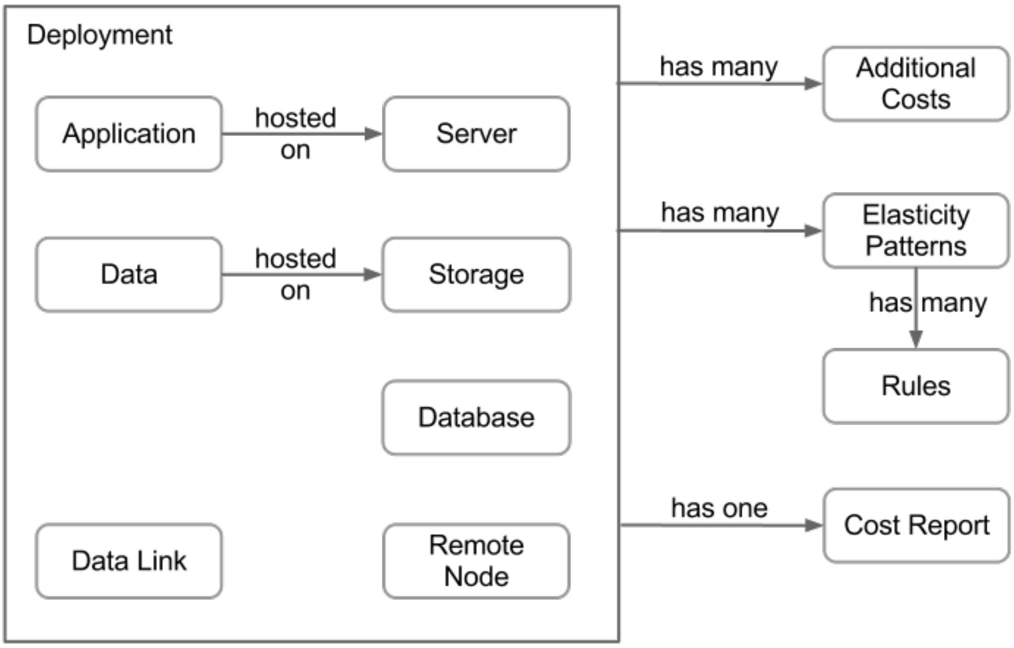


Figure 9: Modelling notations developed as part of the tool

Server: represents a virtual machine (instance in AWS terminology). Users can choose from the various server types offered by different cloud providers. The server type specifies the server's operating system and specifications such as CPU clock rate, RAM and local disk size. Users can also specify the quantity, and the number of hours that the server will be running every month.

Storage: represents persistent storage that can either be attached to a server (e.g. AWS EBS) or used standalone (e.g. AWS S3). Users select a storage type when creating storage nodes, in addition to the quantity, size (e.g. 100GB) and the number of read and write requests that are expected per month. Depending on which type of storage is being used, read requests can represent things like GET operations on files (e.g. for S3), or disk read request (e.g. for EBS). Similarly, write requests can represent PUT/POST operations (e.g. for S3) or disk write request (e.g. for EBS). The Unix `iostat` command can be useful when obtaining estimates of read/write requests.

Database: represents hosted databases such as the AWS' Relational Database Service (RDS) or Microsoft's SQL Azure. Users select a database type when creating databases, in addition to the quantity, number of hours that the database runs each month, size (e.g. 100GB) and the number of database transactions that are expected per month.

Application: represents software applications or programs in a system. Applications are deployed on servers, where their running hours are added to for cost calculations. For example, if there are 2 applications and the first one runs for 8 hours per day (09:00-17:00) whilst the second one runs for 1 hour every night, then the server hosting the applications will need to run for 8 + 1 * 30 = 270 hours per month. The data transfer to and from applications is also mapped to the server hosting them.

Data: represents application data in a system. Application data is deployed on storage nodes, where their storage size and read/write requests are added to for cost calculations. For example, if there are 3 applications and each uses 2GB of data, then the storage node that hosts the application data will need to be 6GB. The data transfer to and from application data is also mapped to the storage node hosting them.

Remote Node: represents external nodes and is useful when modelling the data transfer between resources in the deployment and other nodes. Remote nodes can be anything that requires data transfer to and from a deployment such as users, or applications outside of the deployment.

Data Link: represents regular data transfer between a source and destination node in a deployment. A data transfer link can be created between any of the previously mentioned resources; users specify the data transferred in either direction of the data link in Gigabytes per month.

Additional Cost: deployments often have additional costs such as staff, software licences or the costs of using third party services. These general costs are represented as additional costs and can be added to different deployments. Users specify the amount of the cost per month.

Elasticity Pattern: patterns can be used to describe interesting events or regular routines in a system, during which the resource consumption of the system changes. Patterns describe the elasticity properties of systems. For example, an online shopping system can use a pattern to describe how its resource needs change during the busy shopping season. Patterns can contain many rules (to describe the changes), and are further described in the next section.

Report: each deployment has its own cost report that shows the estimated cost of that deployment during the selected reporting period. Cost reports are the main output of the tool.

These modelling notations include the basic components of any system being deployed in the cloud. To summarise, the cost modelling tool can include the following costs in its estimates: running hours of servers/databases, storage costs, read/write requests of storage, transaction costs of databases, costs of data transferred in/out of a cloud. Furthermore, additional costs can be created by users to include other items in the estimates. There can be other costs associated with deploying a system in the cloud, e.g. the cost of a static IP address; however, these costs are usually insignificant.

Some systems use special services in the cloud, for example AWS' CloudFront[51] service that provides fast multimedia content delivery over the web. The tool does not currently support such services as they are specific to each cloud provider and do not generalize well across various providers. However, support for these special services could be added to the tool in the future.

The wide range of deployment options available from cloud providers presents cost estimation challenges for organisations as they have to choose from hundreds of different types of servers, storage and databases. One of the benefits of the cost modelling tool is that it enables users to generate cost reports for different deployment options and perform 'what-if' style analysis relatively easily.

5.2 Elasticity Patterns

Public IaaS clouds enable users to develop highly elastic systems that can scale up/down their resource consumption to address changes in demand. Start-up companies use IaaS clouds as they do not know, in advance, the demand for their applications; hence, the cloud's elasticity enables them to address any unexpected peaks in demand, if their applications become popular.

---

[51] http://aws.amazon.com/cloudfront/

However, most enterprise applications are likely to have a fairly stable set of users with predictable usage patterns. For example, in the retail industry, sales reporting applications are likely to have peaks in demand just before the shops close when managers check their sales figures for the day. They are also likely to have peaks on the last day of each month when sales reports have to be generated for regional managers. Such patterns also occur over longer time periods in different industries; for example taxation systems in the accounting industry are likely to have peaks before the yearly tax deadline (which is April in the US), whereas retail systems are likely to have peaks leading up to Christmas in December. Figure 10 illustrates such patterns by showing the web traffic of two sites, one from the retail industry and one from the accounting industry in the US. In addition to these systems that have a predictable yearly usage period, there are also systems that have one-off usage periods, such as websites setup for the Olympics or FIFA worldcup.

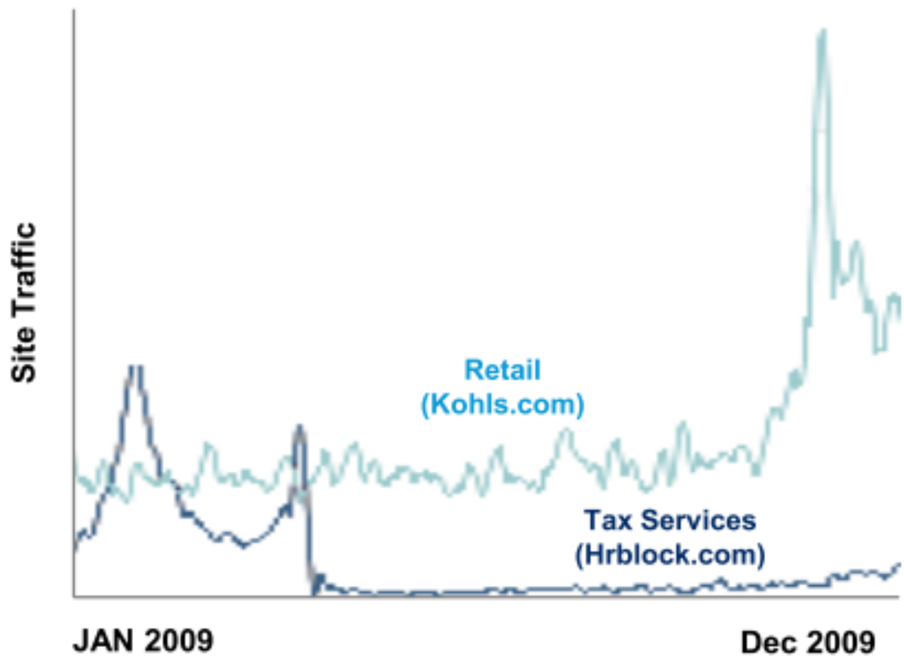


Figure 10: Web traffic of two websites from different industries, image taken from [99]

A simple notation was developed to enable patterns to be expressed in natural language. These patterns can be discovered from historic data (if available), or can be created heuristically to do what-if style analysis of the infrastructure costs for different scenarios. An *elasticity pattern* consists of a *baseline* and a set of *rules* that describe how the baseline changes over time. For example, the storage requirements of a system can be described by saying that the *baseline is 100GB* and *every month it is increased by 5GB*. There are two types of rules that can be created:

- **Permanent rules**: changes made by permanent rules persist (Figure 11), hence they change the baseline after they have been applied. Permanent rules are useful when describing a pattern that sets a new baseline. For example, a user might create a storage unit that uses 100GB per-month, and this value needs to be increased by 5GB every month. The user can create a single permanent rule to describe this pattern (every month +5), rather than create individual rules that set the baseline every month (January: 100GB, February: 105GB, March: 110GB...)
- **Temporary rules**: changes made by temporary rules only apply for the duration of the rule (Figure 12), hence the baseline is changed back to its original value after the rule has been applied. Temporary rules are useful when describing a pattern that is one-off. For example, a user might create 5 web servers to serve visitors every month, but it needs to double this to handle a peak that the site gets during the summer due to an upcoming promotion. The user can create a single temporary rule to describe this pattern (jun-aug *2). This means that the number of servers will be set to 10 in June, July and August, and go back to 5 in September.

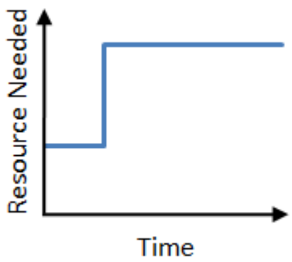


Figure 11: Permanent rule

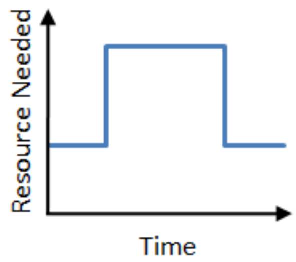


Figure 12: Temporary rule

Rules are defined as follows, where the underlines words are parameters:

```
Type: during year in month on day at hour operation value
```

The month, day and hour parameters are optional and can be used to express fine-grained changes. Table 14 describes the parameters that can be used when creating rules, where YYYY is a 4-digit year, MMM is a 3-letter month (e.g. jun), and DDD is a 3-letter day (e.g. tue).

Table 14: Rule parameters

Type	Year	Month	Day	Hour	Operation	Value
Permanent	every.X.years	every.X.months	every.X.days	Every.X.hours	Add (+)	A number greater than or equal to 0
Temporary	year.X	MMM	every.DDD	[0-23]	Subtract (-)	
	year.X-year.Y	MMM-MMM	every.DDD-DDD	[0-23]-[0-23]	Multiply (*)	
	YYYY		[1-31]		Divide (/)	
	YYYY-YYYY		[1-31]-[1-31]		Raise to power of (^)	
			first.DDD		Set to value (=)	
			first.DDD-DDD			
			last.DDD			
			last.DDD-DDD			

The following scenario illustrates how the amount of storage needed by a retail ordering system can be expressed with elasticity patterns: initially, 200GB of storage is needed but every month this is increased by 5GB, and at the end of every month this needs to be doubled as a full backup of the data is created. However, backups only create a temporary peak as the backup data is transferred off the cloud. During November and December the amount of storage is increased by 15GB per month due to the busy shopping season. Figure 13 shows the outcome of these patterns (the temporary peak at the end of each month is not shown):

```
Baseline: 200GB
Permanent: during every.1.year in every.1.month +5;
Temporary: during every.1.year in every.1.month on 30 *2;
Permanent: during every.1.year in Nov-Dec +15;
```

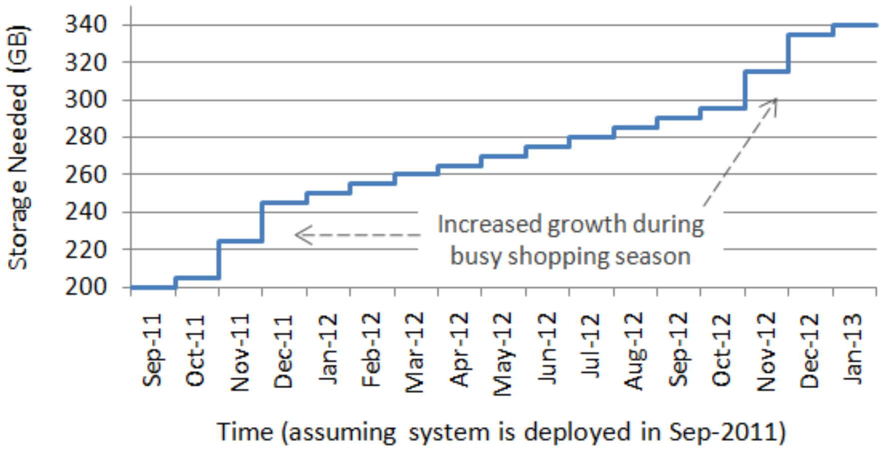


Figure 13: Elasticity patterns of storage needed by a retails ordering system

When a pattern has multiple rules, the position of its rules can make a difference to the resulting values. For example, consider the following two rules and a baseline of 100:

```
Permanent: every.1.year in every.1.month +5
Permanent: every.1.year in every.1.month *2
```

If the rule positions are kept as above, the first three monthly usage values will be 210 *((100 + 5) * 2)*, 430 *((210 + 5) * 2)*, and 870 *((430 + 5) * 2)*. However, if the second rule is applied first, the first three monthly values will be: 205 *((100 * 2) + 5)*, 415 *((205 * 2) + 5)*, and 835 *((415 * 2) + 5)*. Unfortunately it is not possible to avoid this complexity, however, based on our experience with how users used elasticity patterns (discussed in Chapter 8), there are very few occasions where this corner case applies. Figure 14 shows the outcome of using different arithmetic operators when creating rules. Note that a relative percentage change to a baseline can be expressed with the multiplication operator, e.g. *1.5 would represent a 50% increase.

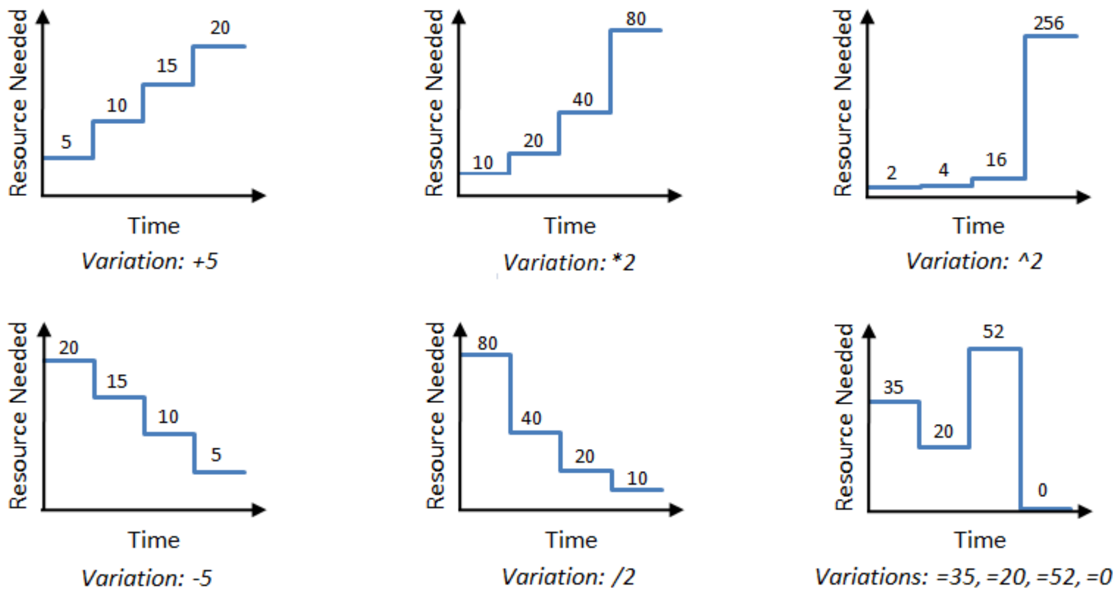


Figure 14: Outcome of using different arithmetic operators when creating rules

Elasticity patterns describe how a numeric value changes over time. A discrete time simulation was developed to process elasticity patterns between a start and end date (described in Section 6.5). In IaaS, cloud cost modelling, elasticity patterns can be used to describe computational resources such as the number of hours that a virtual machine is running for, number of virtual machines, size of storage, number of input and output requests to storage, number of database transactions, and data transferred in/out of a server.

The concept of elasticity patterns is generic and can be used in other modelling tools. For example, elasticity patterns are being used by the ServicesToTheCloud research project[52] to model the number of users a service might have over time. This enables the tool developers to simulate the revenue of a service and evaluate different pricing models for new cloud-based software services. Furthermore, elasticity patterns have been extended by Johnson et. al [100] to support probabilistic rules, where the user can specify the probability of a rule being applied.

Elasticity patterns could, in the future, perhaps be automatically derived from a system's operational and managerial parameters that users are often familiar with. For example, a photography business might experience high workloads on their photo processing systems during special occasions such as graduations or weddings in the

[52] http://www.servicestothecloud.com

summer. Specialised monitoring agents could be developed to gather data from such systems and generate patterns based on historical data. Such patterns would relate to the user's system, e.g. the number of photographs taken in a day, and can be used to generate specific patterns for the cost modelling tool, e.g. the growth rate of the storage system or the data transfer required between systems. This could save users time and effort when specifying elasticity patterns. In addition, users are more likely to be familiar with such high-level patterns and more comfortable when tweaking them.

5.3 Prototype of Cost Modelling Tool

A prototype of the cost modelling tool was developed in Python to experiment with graphical modelling tools and gauge the practicality of using such a tool. The prototype was based on and extended the capabilities of UML deployment diagrams [54], which enable a system's deployment to be modelled. In its essence, a UML deployment diagram enables users to model the deployment of software artefacts onto hardware nodes.

The cost modelling tool extends UML deployment diagrams by including the new modelling notations that were described in Section 5.1. This was achieved by creating a UML Profile with the Eclipse Model Development Tools[53]. Once a user has imported the UML Profile into the Eclipse IDE, they can create graphical models of their systems, and how they could be deployed on IaaS clouds. Figure 15 shows an example model, where the servers, storage, application and data transfer between applications are modelled for a given system. This particular system belonged to a university's Computer Science department and is further described in the case study in Section 7.1.

The main advantage of using the UML is that most practitioners are familiar with its modelling notations and there is good tool support from vendors. However, such tools are usually quite rigid in the interface that they provide, and it can be difficult to customise their behaviour. Furthermore, graphical modelling notations, such as those provided by the UML, generally become difficult to visualise after the models grow beyond a certain size.

[53] http://www.eclipse.org/modeling/mdt/?project=uml2

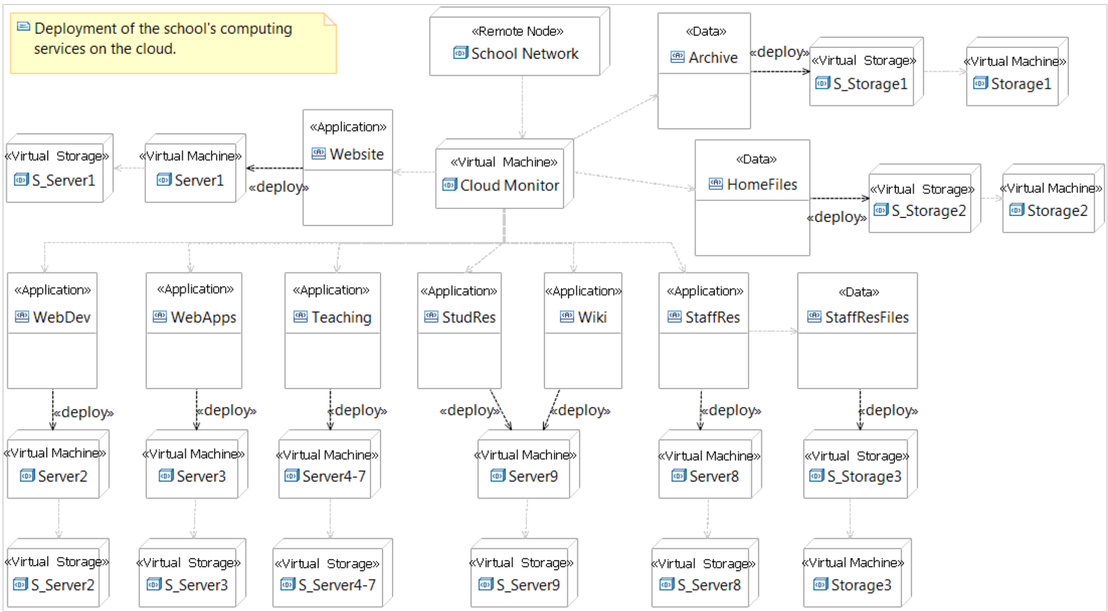


Figure 15: Cloud deployment diagrams can be created using the Eclipse IDE by installing the UML Profile

Once the user has created a model of their system, they can select the cloud they wish to use for each of their servers, storage nodes or databases. The prototype supports Amazon Web Services, Microsoft Azure, FlexiScale, Rackspace, and GoGrid (other providers can easily be added). The various infrastructure prices of the cloud providers could have automatically been added to the tool if the providers had created web services that provided the prices; however, they do not currently provide such web services and the prices had to be manually entered into the tool from the providers' websites.

After a cloud deployment model has been created and any elasticity patterns have been defined, the user has to set a start and end date for the cost simulations to be performed. Once the simulation starts, the tool converts the graphical deployment model into an XML file that is then used to create a directed cyclic graph representing the model. The elasticity patterns of each node and edge in the graph are processed for each month between the start and end date of the simulation.

The total resource usage of each node is then multiplied by the per-unit cost of that resource, depending on which cloud is specified by the user. The per-unit price is retrieved from an XML file that stores the prices from the cloud providers (Figure 16 shows a small section of this file). This file contains over 600 prices from the previously mentioned cloud providers.

```
  <?xml version="1.0" ?>
  <!--  Updated on 02-Nov-2010  -->
- <providers>
  - <VirtualMachine provider="Aws.UsEast" type="Aws.OnDemand.Standard.Small" operatingSystem="Linux/UNIX">
    - <validFrom date="2010-05-01">
        <DataIn monthly="0.1" />
        <DataOut monthly="0.15" />
        <RunningHours monthly="0.085" />
      </validFrom>
    </VirtualMachine>
  - <Database provider="Aws.UsWest" type="Aws.Rds.Reserved1Year.Standard.ExtraLarge">
    - <validFrom date="2010-11-01">
        <DataIn monthly="0.1" />
        <DataOut monthly="0.15" />
        <RunningHours initial="1820" monthly="0.471" />
        <StorageSize monthly="0.1" />
        <InputRequests monthly="0.0000001" />
        <OutputRequests monthly="0.0000001" />
      </validFrom>
    </Database>
```


Figure 16: Cloud prices are stored in an XML file

Finally, the tool generates a detailed cost report showing how the cost of the system changes over time. Figure 17 shows a screenshot of an example report (the screenshot is provided to illustrate the tool's UI and need not be read in detail). The report is a webpage with embedded graphs and tables as well as a zoomable version of the model, which can be very useful when dealing with systems that have a large number of nodes. The graphs show how the cost of the system would change over the reporting period, as well as a breakdown of the costs that can be used as a starting point for optimisation (e.g., if the major costs are for the server running hours, then simply switching the servers off at night when they are not used might be a simple but effective cost optimisation technique). The system can also export the full costing details as a CSV table for further analysis in Microsoft Excel.

The model can be divided into different groups, and the report provides a detailed breakdown of the costs of each group. A group can represent a department, an organisation or an entire system. This enables architects to evaluate different deployment options of a system and see which is the cheapest. For example, system architects can investigate the costs of duplicating parts of the system on a different cloud for increased availability.

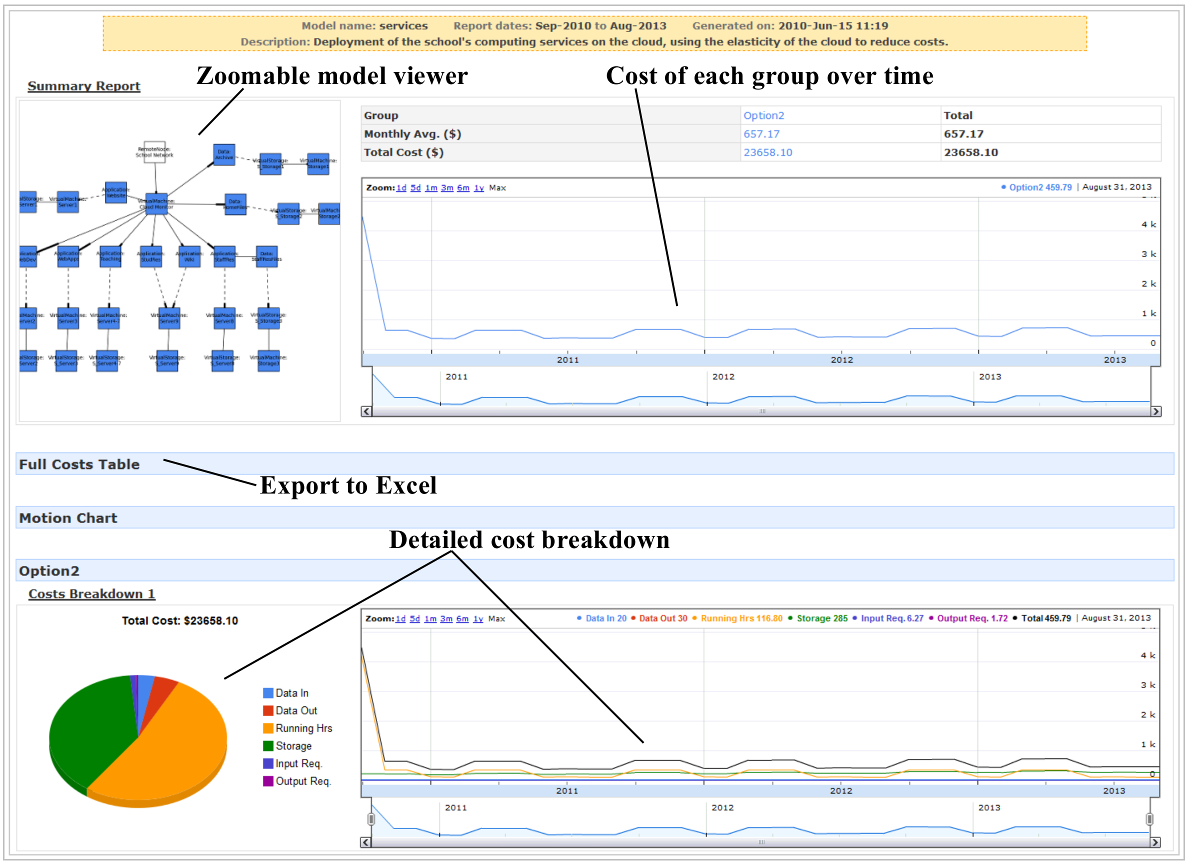


Figure 17: Screenshot of an example cost report showing how the cost of a system could vary over time

Graphical modelling notations are often used in software engineering, as they are easier to understand than detailed textual descriptions, and can be particularly useful during the shift from the analysis to design stages. Graphical representations of systems (i.e. system models) can be useful for end-users since they "leave out details" and provide an abstraction that simplifies and communicates the most relevant characteristics of a system from the user's perspective [71]. Larger systems can lead to larger models with more details; hence, more advanced browsing tools are required to enable users to grasp the main concepts and focus on what they are interested in without being overwhelmed by the 'messiness' of the models.

The following issues were discovered during the development of the prototype:

1. The tool installation process had to be simplified. To use the tool, a user would have to download and install the Eclipse IDE, install the Model Development Tools plugin, import our UML Profile and enable it from the Eclipse options menu.

2. Graphical modelling tools are attractive as they are visual and non-technical users can relate to them, however, they do not scale very well. For example, models containing over 30 nodes can be messy to layout on the screen. A tabular interface would be more scalable.
3. The Eclipse IDE has usability issues. For example, when a user performs a copy/paste operation on a node (or group of nodes), Eclipse creates two copies of the node but ties both of them to the same data model underneath, meaning that when the user changes one node's attributes, the other node is also changed without the user noticing it.
4. Cloud pricing information cannot be obtained manually as there are too many prices and they change over time. None of the cloud providers that were supported in the prototype had an API for pricing information; therefore, a screen-scraping approach would be one way to obtain this information automatically. Some clouds have tiered pricing and a more flexible method needs to be used to store such prices. Also, a relational database would be more appropriate for storing pricing information (compared to XML), as it would provide better search functionality and more flexible storage of prices using relations and join-tables.
5. The cost report produced by the prototype contained too much information for the average user. Users simply need to know what the monthly costs are, and a breakdown of the costs. The concept of grouping parts of the model, to produce sub-reports, is redundant as users can just create separate models.

Developing a customised web application for cloud cost modelling would address issues 1, 2 and 3 above. Using a relational database (such as MySQL or PostgreSQL) along with the web application would address issue 4, and simplify the process of keeping pricing information up-to-date. Issue 5 could also be addressed during the web application development by removing unnecessary information from the report. The next chapter describes the development of ShopForCloud.com, a web application developed for cloud cost modelling, which has been developed to address the problems identified in the prototype system.

6 ShopForCloud.com

ShopForCloud.com is a web application that enables users to estimate and forecast the costs of deploying their systems on public IaaS clouds. It supports users in making system deployment decisions by providing them with a simple tool that can be used to obtain cost estimates for alternative deployment options, cloud providers and usage scenarios (using elasticity patterns). ShopForCloud is one of the major contributions of this thesis and aims to make cloud cost modelling simple enough such that organisations could use the tool in a self-service manner. It used insights gained during the development of the cost modelling tool prototype (described in Section 5.3), and was designed to be user-friendly and robust.

ShopForCloud was launched publicly in February 2012 and ran as a hosted service for four months. A simple website was setup to attract new users (see Figure 18); this website was supplemented with a blog[54] and Twitter[55] feed. Anyone could sign-up and create an account during this time period, and 270 users signed-up. Chapter 8 provides an evaluation of ShopForCloud and discusses how users used the tool. A *Getting Started Guide* was also put online to demonstrate the functionality of the tool; see Appendix A for details. A typical user would:

1. Login and have a look at the clouds that were supported in the tool
2. Create a deployment representing a system that they wanted cost estimates for
3. Create any required elasticity patterns and attach them to the resources in their deployment
4. Generate a cost report for their deployment
5. Clone the deployment and try alternative deployment options or clouds.

The Scrum process was used during the 6-months development period of ShopForCloud and the work was carried out in 2-week sprints, which were followed by a 1-hour demo to interested parties to gather early feedback (mostly on the user interface). The tool consists of around 14,500 lines of code, which is available online[56] under the New BSD license. The `git` revision control and source code

[54] http://blog.shopforcloud.com
[55] https://twitter.com/#!/shopforcloud
[56] https://github.com/alikhajeh1/cloud_cost_modelling

management system was used, and all of the code was hosted on Github.com. Around 270 test-cases were developed for the tool, which covered 91% of the source code.

This chapter describes the design and implementation of ShopForCloud; it starts with an overview of the system architecture and describes the main components of the tool (Section 6.1). These components include the classes that were used to model clouds and their resources (Section 6.2); classes that were developed to scrape cloud provider websites to obtain pricing information (Section 6.3); classes that were developed to enable users to model their system deployments (Section 6.4); details of the main algorithms that performed the cost simulations (Section 6.5); and an overview of how the cost reports were generated (Section 6.6). Finally, Section 6.7 describes how ShopForCloud was deployed and maintained on a cloud itself.

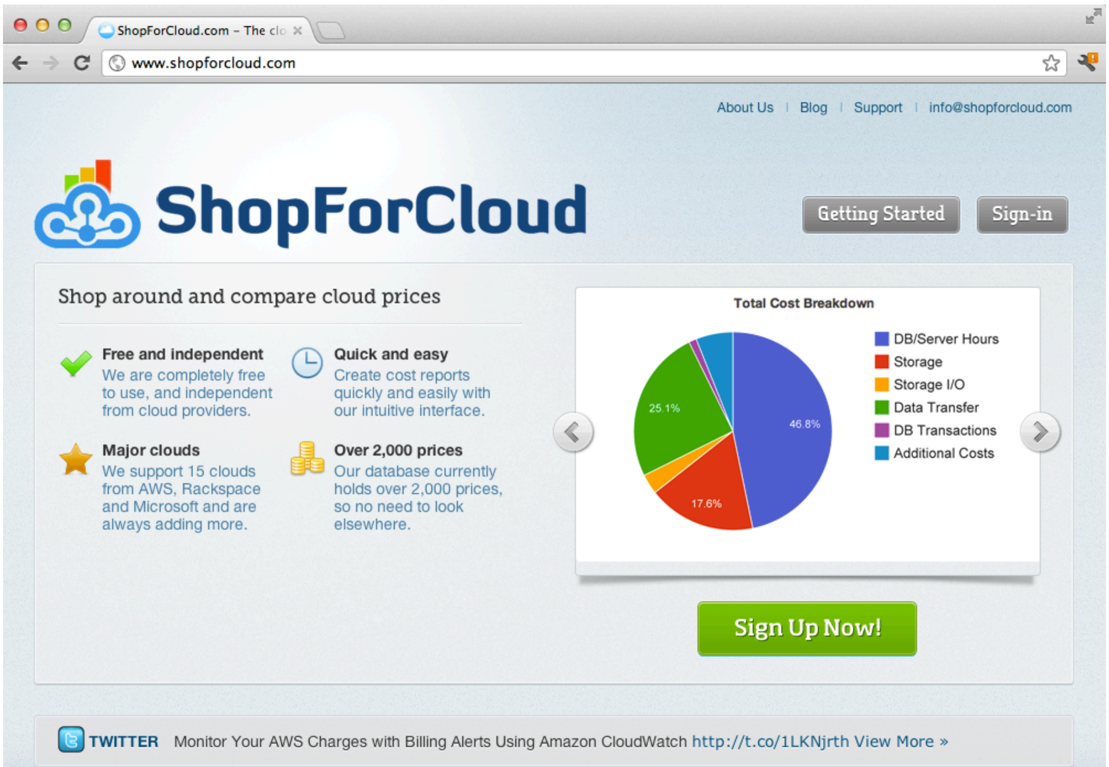


Figure 18: ShopForCloud.com

6.1 Architecture

Most modern web applications are developed using a Model-View-Controller (MVC) framework. Ruby on Rails (often called Rails) is one of the most popular MVC frameworks used by web developers, and there are many open source projects that can plugin into it to provide extra functionality such as user authentication or time/date validation. These open source projects are called *gems* in the Ruby terminology. Rails is database agnostic and can work with most popular databases as there are various gems that handle the actual database calls on behalf of the developer.

ShopForCloud was developed using the Rails framework. Figure 19 shows a high-level overview of the Rails architecture. When the browser makes a request, the web server receives it and uses `routes` to find the corresponding controller that should deal with the request (e.g. `GET servers/3` invokes the `show` action of the `server` controller passing it the ID 3). The dispatcher creates a new instance of the required controller and sends all parameters to it. Controllers perform the required actions by using models, which are Ruby classes that hold the business logic to find, update, validate and save data to the database etc. Finally, controllers invoke views to render the data; views usually use a mixture of HTML, CSS and JavaScript to create the web pages that users see.

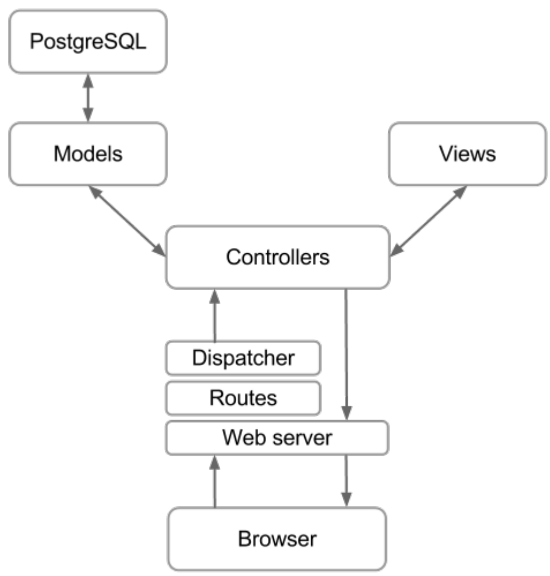


Figure 19: Rails MVC architecture

There are 23 Rails models and around 15 other classes in ShopForCloud; these can be divided into six main components as shown in Figure 20 (described in the following list).

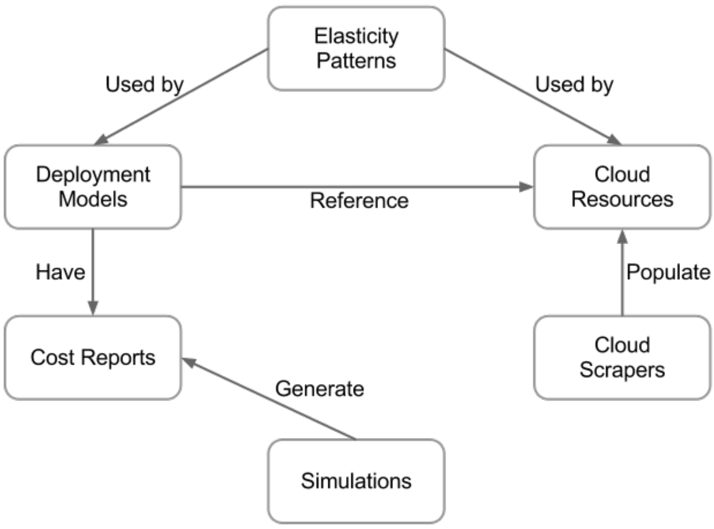


Figure 20: ShopForCloud system architecture

- Elasticity patterns, which were introduced in Section 5.2, are implemented in a generic way such that a user can define them independently and attach them to any component that has numeric attributes. For example, they can be attached to the *storage size* attribute of a *database*. They can also be used by developers to describe pricing information of cloud resources; for example, AWS reserved instances use a pattern to describe the recurring reservation fee (`every.3.years +276` for a Large Linux instance in the US-East cloud).
- Cloud resources include the types of servers, storage and databases that are supported by each cloud and their pricing information. These resources are analogous to classes in object-oriented programming, where users can create instances of these classes when describing their deployments.
- Cloud scrapers fetch the hundreds, or even thousands, of prices that each cloud has for its resources. A scraper creates cloud resources and updates their prices when they change, thus a scraper needs to be developed for each cloud provider that is added to the tool. ShopForCloud currently supports 3 cloud providers, namely AWS, Microsoft Azure and Rackspace. The functionality of the scrapers is further described in Section 6.3.
- Deployment models enable users to describe their system deployments. Deployments contain servers, storage, databases and the other resources that were described in Section 5.1, some of which refer to cloud resources. For example, when a user creates a server, they select the cloud and server-type of the server.

- A cost report is generated for each deployment. The cost reports are a simplified version of the ones that were developed during the prototype stage.
- Simulations are used to process elasticity patterns, and calculate the various values and raw data that are required to generate cost reports.

6.2 Modelling Clouds

Figure 21 shows the Rails models that were developed to represent clouds, their resources and pricing information. A cloud provider can have many clouds, which are often located in different regions of the world. Each cloud charges its users in a pre-defined currency, and has many types of resources, such as servers, storage and databases. For each of these resources, cloud providers usually charge for a number of items, thus there are many cloud cost structures (e.g. a server has `instance_hour`, `data_in` and `data_out` costs). Furthermore, each cloud cost structure can have many cost tiers. Therefore, each cloud resource type has many cloud cost structures in each cloud, which is represented by the ternary association named Cloud Cost Scheme.

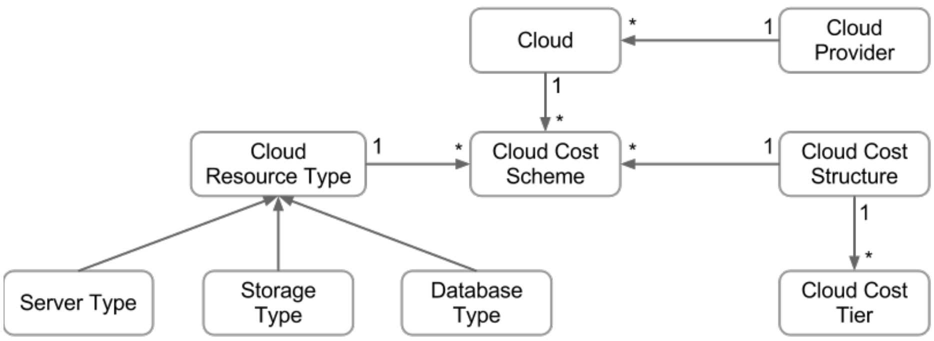


Figure 21: Modelling clouds, their resources and prices

This modelling notation is flexible enough to deal with differences in pricing schemes between cloud providers; for example, AWS charges for read/write requests for S3 storage, whereas Rackspace does not charge for this in its CloudFiles service. The notation also avoids storing duplicate information; for example, AWS has a *small* server type available in all of its clouds but they are charged at different rates (presumably due to different electricity costs). However, only one AWS small server type is created in the database, which has different cost structures in different clouds. The following list provides the attribute details of the Rails models presented in

Figure 21, and shows an example of how S3 storage is modelled in this notation (example data shown in brackets):

- **Cloud Provider**: name (Amazon Web Services), description (AWS Cloud), website (http://www.aws.amazon.com)
- **Cloud**: name (AWS EU-Ireland), description (Main European cloud), billing_currency (USD), location (Ireland)
- **Cloud Resource Type**: name (S3 Reduced Redundancy), description (Simple Storage Service). This model uses the Rails single-table-inheritance technique such that there is only one table and all sub-classes (Server Type, Storage Type, Database Type) store all of their data in the same table. The other attributes in this model are cpu_architecture (X86), cpu_speed (1.0 GHz), cpu_count (2), local_disk_count (1), local_disk_size (160 GB), memory (1.7 GB), operating_system (Linux/UNIX), and software (MySQL).
- **Cloud Cost Structure**: name (storage_size), units (per.1.gbs.per.1.months; this syntax is parsed and processed during the cost simulation), valid_until (null if it's a valid price, otherwise the date until it was valid), recurring_costs_monthly_baseline (used for reserved instances, this would be 0 for S3 storage as it is not applicable), custom_algorithm (used to override the default algorithm that is used during simulations; this is further explained in Section 6.5). AWS S3 also has a cost structure for data transferred out, read and write requests.
- **Cloud Cost Tier**: name (free tier), upto (1024 GB), cost (0.093, the currency is stored in the cloud model as all cost tiers use the same currency in a cloud). AWS S3 has six cost tiers for its storage_size cost structure, five cost tiers for its data_out cost structure, and one cost tier for each of its read and write request cost structures.
- **Cloud Cost Scheme**: cloud_resource_type (S3 Reduced Redundancy), cloud_cost_structure (storage_size), cloud (AWS EU-Ireland). AWS S3 has a total of 28 records in this table (4 cost structures x 7 clouds).

6.3 Cloud Scrapers

The prototype cost modelling tool demonstrated that it is feasible to create a database of prices from cloud providers. The prototype had a list of around 600 prices from AWS, Microsoft Azure, FlexiScale, Rackspace, and GoGrid. Out of these providers, AWS, Microsoft Azure and Rackspace were added to ShopForCloud as they were more popular (FlexiScale and GoGrid have a similar pricing scheme as Rackspace). Furthermore, the prototype demonstrated the need for an automated mechanism to add prices to the database and keep them up-to-date. Thus a scraper is developed for each cloud provider that is added to ShopForCloud.

AWS has over 2,000 prices for its various types of servers, storage and databases. For example, AWS EC2 has 7 clouds x 13 types of instances x at least 3 OS choices (Linux, Windows, Windows with SQL Server) x at least 4 purchase options (on-demand, light/medium/heavy-utilization reserved) => 1,092 individual prices. These prices are stored in around 30 JSON files and used by AWS web pages. The AWS scraper fetches these JSON files and parses them to create the required database records in ShopForCloud (as described in the Section 6.2). When a scraper is re-run, it invalidates the old prices by setting their valid_until date.

Microsoft Azure and Rackspace have much fewer prices for their clouds (around 30 prices for each cloud), and their scrapers are much simpler. Their scrapers rely on the developer to enter the prices into a template, which is then used to create the necessary database records. These scrapers could be changed to be similar to AWS in the future if these cloud providers increase the complexity of their pricing schemes.

Amazon changed their prices twice and Azure changed its prices once during the four months that ShopForCloud was running, and the scrapers were manually re-run on each change to update the prices. Cloud providers publicise any changes to their prices via Twitter, and this is how the developer could be notified of changes. However, more advanced mechanisms could be developed in the future to pickup any changes to the prices.

6.4 Modelling System Deployments

The modelling notations described in Section 5.1 enable users to model their system deployments at a level of detail that is suitable for cloud cost modelling. However, to simplify the tool further, the Application and Data notations were left out of ShopForCloud as they were somewhat redundant (they simply added their values to the server/storage hosting them).

Figure 22 shows the Rails models that were developed to represent system deployments. Each deployment has one cost report and can contain many servers, storage, databases, remote nodes, and data links between any nodes in the deployment. Users create additional costs independently of deployments and can add them to any deployment in their account using a join table. Patterns can also be created independently and attached to any component in the user's account that has numeric attributes. For example, a pattern can be attached to the *quantity* attribute of *servers*.

As shown in Figure 20, deployments and their related models are scoped to the user account, meaning that each database record has a user_id attribute that specifies the user who owns that object. This is a common practice when developing multi-tenant software-as-a-service and has to be strictly enforced to avoid security issues. As mentioned in the previous sections, a user has to specify the cloud and server/storage/database type when they create servers/storage/databases in their deployments. This enables the cost simulation to find the relevant pricing information for each resource in the database.

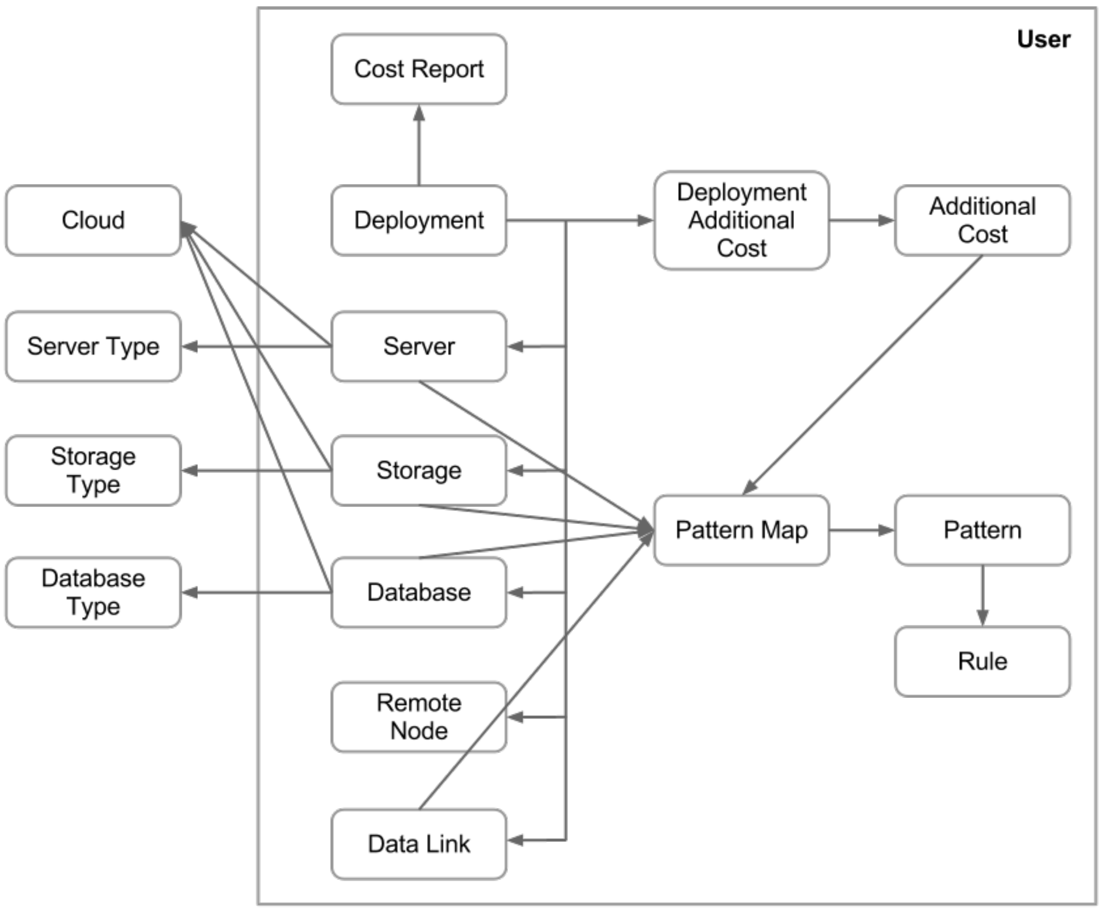


Figure 22: Modelling system deployments

Figure 23 shows a screenshot of the page for creating servers in a deployment. The following list provides the attribute details of the Rails models involved when creating deployments.

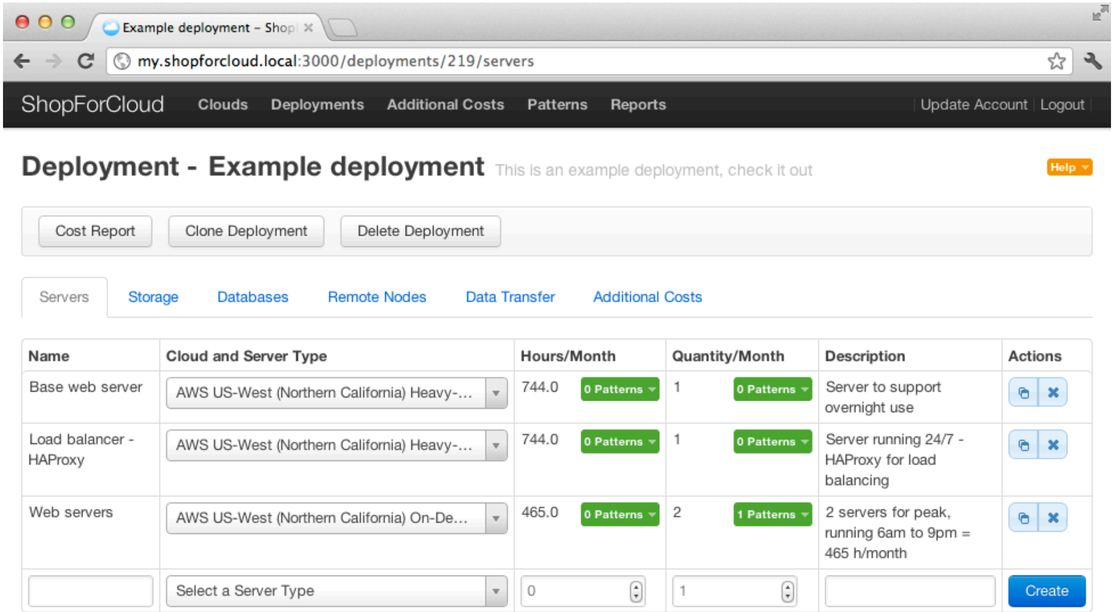


Figure 23: Creating servers in a deployment

- **User**: email, encrypted_password, sign_in_count, first_name, last_name, company, currency (all generated reports are converted into this currency)
- **Deployment**: name, description
- **Cost Report**: name, description, start_date, end_date, status (simulation status, which can be Pending, Processing, Completed, Failed), xml (raw xml data used to generate report), xslt_file (name of XSLT file that was used to generate report), html (generated HTML from combining XML and XSLT). The report generation mechanism is further described in Section 6.6.
- **Server**: name, description, cloud, server_type, instance_hour_monthly_baseline (baseline number of hours that the server will be running every month), quantity_monthly_baseline (baseline number of servers that will be required every month).
- **Storage**: name, description, cloud, storage_type, storage_size_monthly_baseline (baseline size of storage required every month in GB), read_request_monthly_baseline, write_request_monthly_baseline, quantity_monthly_baseline
- **Database**: name, description, cloud, database_type, instance_hour_monthly_baseline (baseline number of hours that the database will be running every month), storage_size_monthly_baseline (baseline

database size in GB), transaction_monthly_baseline (baseline number of transactions to database per month), quantity_monthly_baseline

- **Remote Node**: name, description
- **Data Link**: name, description, source_node, destination_node, source_to_destination_monthly_baseline (baseline number of GBs that will be transferred from source to destination every month), destination_to_source_monthly_baseline
- **Additional Cost**: name, description, cost_monthly_baseline (baseline monthly cost of the additional cost)
- **Pattern Map**: a pattern has a name and description, and the various rule attributes were described in Section 5.2. Patterns can be attached to any model attributes using the Pattern Map join table that contains: pattern_id, model_id, model_type (name of the Rails model is needed so its ID can be looked-up), model_attribute (the name of the attribute in the model that requires the pattern), and position (the pattern's position is important as it can make a difference during the calculations as mentioned in Section 5.2).

Figure 24 shows a screenshot of the page used to create pattern rules. The day and hour rule attributes were left out of ShopForCloud to make the concept of elasticity patterns simpler for users (based on our experience from case studies, these detailed patterns were rarely used). Each page on ShopForCloud has an expandable help section at the top, which is expanded by default on the first login of a user. The help section describes the features of each page and links to supporting material in the Getting Started Guide where necessary. As shown in Figure 24, popovers are also used to explain non-trivial form fields.

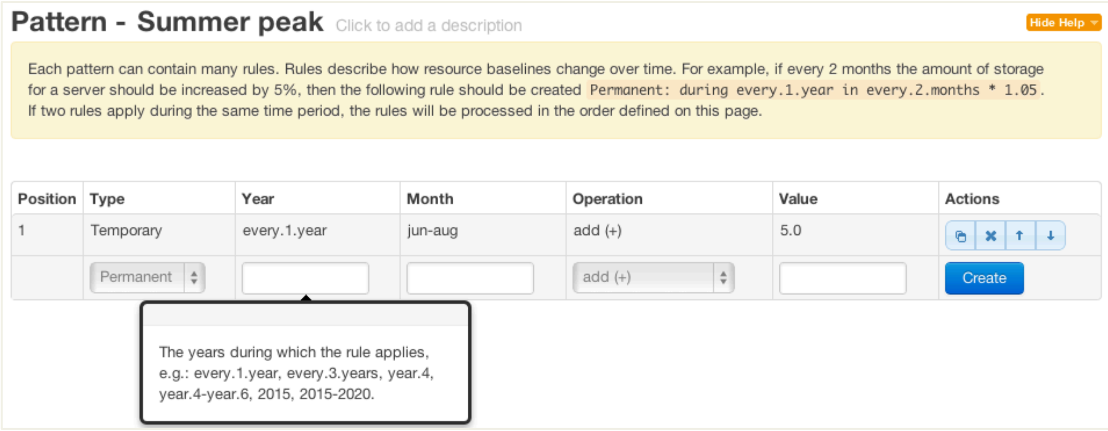


Figure 24: Creating pattern rules

Once patterns have been created, they can be attached to any attribute that has a pattern attachment button, as shown in Figure 25. The pattern attachment user interface section is hidden by default and is only shown when the pattern attachment button is clicked. Once shown, this section shows all of the available patterns on the left-hand side, and allows users to attach any patterns to the attribute by double clicking on it or dragging it to the right-hand side. The order of the attached patterns can be altered by dragging the patterns up or down.

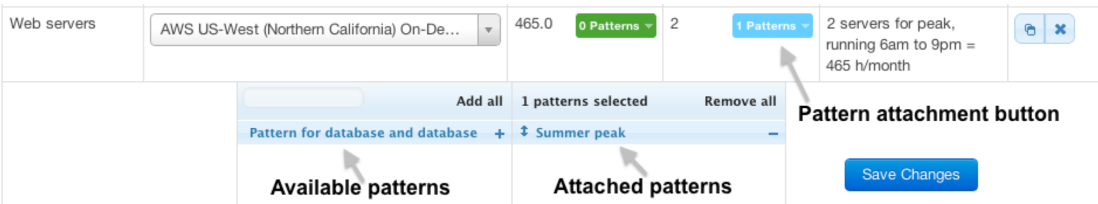


Figure 25: Pattern attachment user interface

The ShopForCloud user interface was developed using Twitter Bootstrap[57], which is a simple and flexible HTML/CSS/JavaScript UI component library. The UI also has a number of customised features such as edit-in-place of form fields and pagination to make it more user friendly. Another customised feature is the ability to find a server/storage/database type by simply entering parts of its name, which makes it much simpler to select a cloud and server/storage/database type when creating deployments.

[57] http://twitter.github.com/bootstrap

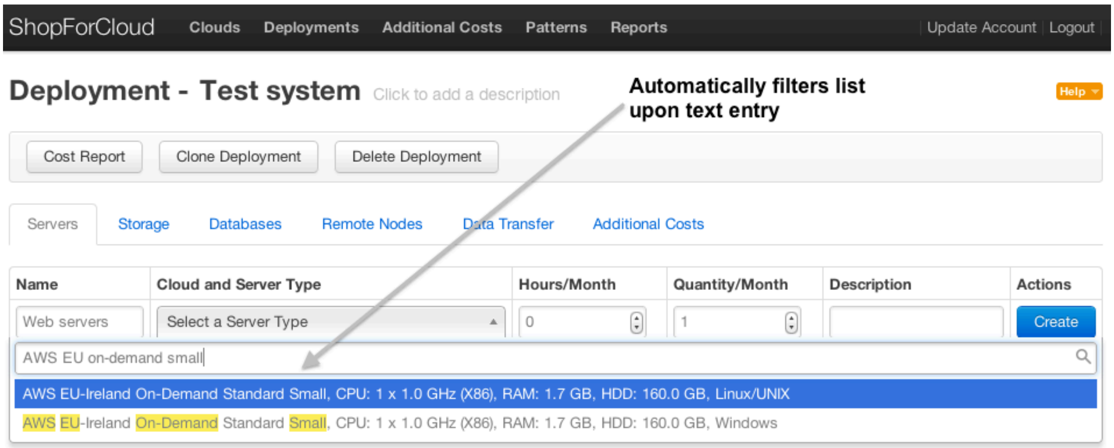


Figure 26: Automatic filtering of cloud and server type list upon text entry

6.5 Simulations

The key components of ShopForCloud are the simulations performed to evaluate elasticity patterns and calculate a deployment's costs. Once a user has created a report, where they also define the start and end dates of the report, a discrete time simulation is performed between the dates in monthly steps. Figure 27 provides an overview of the steps involved in the simulation.

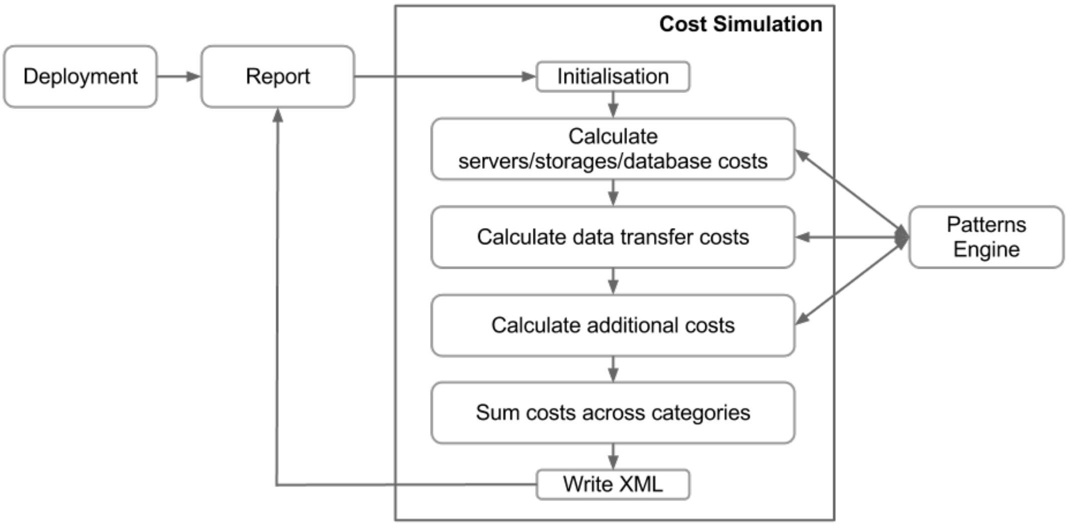


Figure 27: Cost simulation overview

The initialisation step creates an array (called the `results` array) for the months between the report start and end dates, where each array item contains a hash of the cost categories and their total cost for that month (the categories are instance_hour, storage_size, read_request, write_request, transaction, data_in, data_out, and additional_cost). The simulation is then started and the costs of the servers, storage, databases, data transfer and additional costs are calculated. The Patterns Engine is used to process any elasticity patterns in the deployment. The algorithms for these steps are described in subsequent sections. Finally, the simulation creates an XML file containing the data in the array that was created during the initialisation step. The XML data is saved in the report's database record and used to generate the final cost report (described in Section 6.6).

6.5.1 Patterns Engine

The Patterns Engine contains a static method called `get_monthly_results` that takes in a baseline value, a list of patterns that are to be applied to the baseline and the start/end dates of the simulation. As shown in Figure 28, this method returns an array of numbers, where each number represents the total usage for that month. The `get_monthly_results` method calls the helper `apply_rule` method, which in turn calls the `rule.applicable?` method to decide whether the rule should be applied or not. This method (not shown in Figure 28) uses the `time_diff` gem and a set of 8 regular expressions to parse the rules; it returns true if a rule is applicable for a given month. For example, this method would return `true` if it is called with the rule `every.1.year on jun-aug +5` and a `start_date` of 1 January 2012 and a `current_month` of July 2012.

```
def self.get_monthly_results(
                     monthly_baseline, patterns, start_date, end_date)
  results = []
  current_month = start_date
  # Loop through the months between start and end date
  while current_month <= end_date
    # Set the current month's usage to the baseline
    current_result = monthly_baseline

    # Process the patterns by applying their rules to the current month
    patterns.each do |pattern|
      pattern.rules.each do |rule|
        current_result = apply_rule(
                           start_date, current_month, current_result, rule)
        # Update the monthly_baseline if it's a permanent rule
        monthly_baseline = current_result if rule.rule_type == 'permanent'
      end
    end

    results << current_result
    current_month += 1.month
  end

  return results
end
```

```
def self.apply_rule(start_date, current_month, current_result, rule)
  if rule.applicable?(start_date, current_month)
    case rule.variation
    when '='
      current_result = rule.value
    when '+'
      current_result += rule.value
    when '-'
      current_result -= rule.value
    when '*'
      current_result *= rule.value
    when '/'
      current_result /= rule.value
    when '^'
      current_result **= rule.value
    end
  end

  return current_result
end
```

Figure 28: Patterns Engine Ruby code

6.5.2 Calculating Server, Storage and Database Costs

Servers, storage and databases have different cloud cost structures but these can all be calculated using the same method. Collectively these resources have the following cloud cost structures: instance_hour, transaction, storage_size, read_request, and write_request. Figure 29 describes the algorithm that was developed to calculate the costs of these resources.

Step 1: Create a multi-dimensional array of the monthly total usages (value * quantity) and quantities of each cloud cost structure for each cloud resource type in each cloud (e.g. an AWS US-East small on-demand Linux server, or a Rackspace UK CloudFiles storage node). During this step, measures are also taken to protect users from applying patterns that set invalid values; for example, a pattern that sets the monthly running hours of a server to more than 744 hours or sets the storage size to less than 0 GBs. The structure of the multi-dimensional array is:

```
[CloudResourceType.id][Cloud.id][CloudCostStructure.name][:values] =
    array of total usages (one array item for each month)
[CloudResourceType.id][Cloud.id][CloudCostStructure.name][:quantities] =
    array of quantities (one array item for each month)
```

Step 2: Use the array created in Step 1 to perform the cost calculations using the following algorithm:

```
Loop through all CloudResourceTypes
  Loop through all Clouds
    Loop through all CloudCostStructures
      If custom algorithm is needed
```

```
        Call custom algorithm
      Else
        Loop through all monthly values
          Calculate tiered cost for that month using values array
          Calculate recurring cost (e.g. reserved instances) for
              that month using quantities array
          Sum usage cost and recurring cost
          Convert cost from cloud’s billing currency to user’s currency
          Add cost to appropriate category in results array for that month
      End
    End
  End
End
```

Figure 29: Server, storage and database cost calculations

This algorithm is also flexible enough to allow custom algorithms to be called from within Step 2 to perform calculations for clouds that have a pricing scheme that cannot be evaluated using the default algorithm. Microsoft SQL Azure is one such service, where the storage cost of the database has to be calculated by amortizing the storage size over the days that the storage was used in that month:

```
storage_cost = get_tiered_cost(storage_size, pricing_units, pricing_tiers) *
((running_hours / 24).ceil / days_in_month)
```

The pricing unit of a cloud cost structure is stored in the database using a simple syntax and used accordingly during the simulations. For example, AWS S3 costs $0.01 `per.10000.requests` for reads, whereas AWS EBS costs $0.11 `per.1000000.requests` for reads, and Rackspace CloudFiles costs $0.15 `per.1.gbs.per.1.months` for storage. The tiered cost of a resource is calculated using the method shown in Figure 30. Essentially this method loops until the resource usage value is reduced to 0, and at each iteration it find the appropriate cost tier and multiplies the usage for that tier by its cost.

```
def get_tiered_cost(total_usage, units, ordered_tiers)
  cost = 0.0
  i = 0
  while total_usage > 0
    usage = total_usage
    if ordered_tiers[i].upto
      # If total_usage is less than upto then use that, otherwise calculate
      # the diff between this tier and the last tier (if there is one)
      usage = [total_usage, ordered_tiers[i].upto -
                                (i > 0 ? ordered_tiers[i-1].upto : 0)].min
    end
    total_usage -= usage
    cost += (usage / cost_units).ceil * ordered_tiers[i].cost
    i += 1
    # Deal with special case when there is no catch-all
    # tier (upto nil), shouldn't happen in practice
    return cost if i == ordered_tiers.length
  end
```

```
  return cost
end
```

Figure 30: Calculating the tiered cost of a resource (Ruby code)

6.5.3 Calculating Data Transfer Costs

Data transfer is free within a cloud and is usually charged using a tiered pricing scheme when data is transferred in/out of a cloud. For example, the AWS EU-Ireland cloud does not charge for data in, but charges using the following tiered pricing scheme per GB for data out: from 0GB up to 1GB: $0.0, up to 10240GB: $0.12, up to 51200GB: $0.09, up to 153600GB: $0.07, and above: $0.05. Thus to calculate the data transfer costs of a cloud, all data in and out of it must first be calculated. Data transfer links can be created between any nodes in a deployment, which leads to the cost calculation algorithm shown in Figure 31.

Step 1: Calculate the total data in and out of each cloud for each month, and store this in a hash where the cloud ID is the hash key and the monthly data in/out array is the hash value. During this step, measures are also taken to protect users from applying patterns that set the data transfer to less than 0 by accident.

```
data_transfer = {'data_in' => {}, 'data_out' => {}}
Loop through all data links in deployment
  Get the source and destination cloud of the nodes in the data link
  Set source/destination cloud to 'remote' if either node is a remote node
  If source cloud != destination cloud
    Add src_to_dest data to the source cloud's data_out
    Add src_to_dest data to the destination cloud's data_in
    Add dest_to_src data to the destination cloud's data_out
    Add dest_to_src data to the source cloud's data_in
  End
End
Delete the 'remote' cloud from the data_transfer array as
    it was only a placeholder
```

Step 2: Calculate the monthly data transfer costs by using the hash created in Step 1:

```
Loop through all data_in/data_out clouds in the data_transfer hash
  Find the data_in/data_out pricing details from database
  Loop through all monthly values
    Calculate tiered cost for that month
    Calculate any recurring cost for that month
    Sum usage cost and recurring cost
    Convert cost from cloud's billing currency to user's currency
    Add cost to appropriate category in results array for that month
  End
End
```

Figure 31: Data transfer cost calculations

6.5.4 Calculating Additional Costs

Additional costs are created by users and are expressed in the user's currency; hence they do not need a currency conversion. Figure 32 shows the algorithm that was developed to calculate the additional costs.

```
Loop through all additional costs in deployment
  Get the monthly values by evaluating any attached patterns
  Loop through all monthly values
    Add cost to 'additional cost' category in results array for that month
  End
End
```

Figure 32: Calculating additional costs

6.6 Reporting

The cost simulation saves its result as an XML document (shown in Figure 33) in the report table in the database. The XML document includes the user's currency, the total deployment cost between the report start and end dates as well as one row for each month in the reporting period. Each row contains a breakdown of that month's costs.

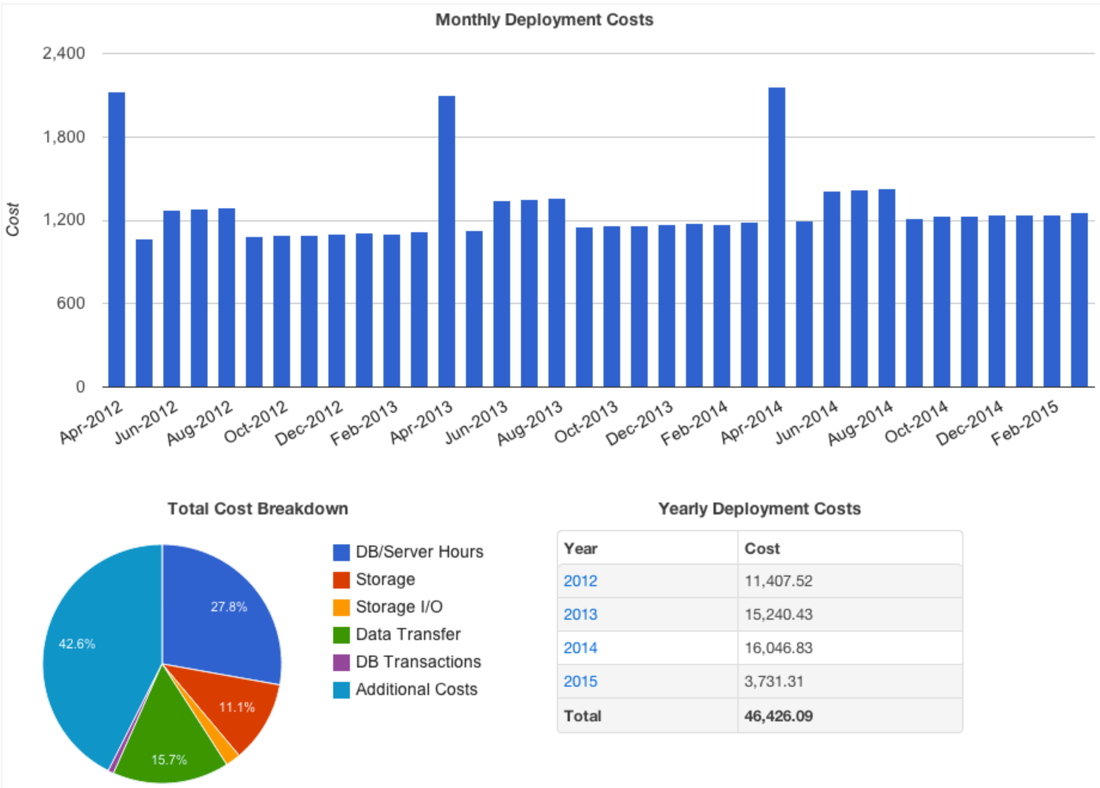
```
<deployment>
  <user_currency>United States Dollar (USD)</user_currency>
  <cost>26542.84</cost>
  <row>
    <year>2012</year>
    <month>Apr-2012</month>
    <instance_hour>1295.5</instance_hour>
    <storage_size>97.1</storage_size>
    <read_request>23.1</read_request>
    <write_request>3.3</write_request>
    <transaction>9.9</transaction>
    <data_in>0.0</data_in>
    <data_out>150.48</data_out>
    <additional_cost>49.0</additional_cost>
    <total>1628.38</total>
  </row>
  <row>
  ...
  </row>
</deployment>
```

Figure 33: XML produced by cost simulation

Extensible Stylesheet Language Transformations (XSLT) is a common tool used to process XML documents to produce HTML web pages. An XSLT was developed to process the report XML and produce the cost report shown in Figure 34. The report shows a bar chart of the monthly costs, a pie chart showing the cost breakdown, and an expandable table showing the monthly costs as well as the total deployment cost.

This example report shows the 3-year costs from the current month; hence, the yearly total for 2015 is lower as it only includes the first few months. The graphs are generated using Google Chart Tools[58], which is a JavaScript library that can take raw data (in JavaScript arrays) and produce customisable graphs.

The peaks in Figure 34 are caused by the yearly recurring fees for reserved instances that the deployment is using, while the steady cost increase and the summer peaks are caused by elasticity patterns being used in the deployment. The costs are shown in US Dollars by default but users can change their currency preference and generate reports in any currency. The currency conversion is done during report generation using live currency exchange rates from Google Currency[59].



Year	Cost
2012	11,407.52
2013	15,240.43
2014	16,046.83
2015	3,731.31
Total	**46,426.09**

Figure 34: Deployment cost report

6.7 Deploying ShopForCloud

The ShopForCloud website (used to attract new users) was deployed on AWS S3 as this was found to be the cheapest way to host static websites. An S3 bucket was setup

58 https://developers.google.com/chart/
59 http://www.google.com/finance/converter

and the necessary DNS configurations were made to map the www.ShopForCloud.com domain name to the S3 bucket. The website hosting costs were on average $0.12 per month, which was for the storage, read/write requests to S3 and data transfer (i.e. people viewing the website).

The ShopForCloud Rails application was deployed on Heroku and had the URL https://my.ShopForCloud.com. Heroku is a Platform-as-a-Service offering from Salesforce that provides a simple and scalable environment for hosting applications. Heroku itself is deployed on Infrastructure-as-a-Service and uses AWS EC2. An SSL certificate was used to secure communications between users and the application as a user's system deployment details and costs are quite sensitive information. Furthermore, the Rails security guidelines[60] were followed during the application development phase to ensure that best practices were implemented.

New Relic[61] monitoring was setup once the application was deployed on Heroku. This monitoring service pinged the application once every 30 seconds and sent an email notification if the application was unavailable for more than three minutes. The availability of the application over a three-month period was found to be 99.867%, which is satisfactory given that this is a non-business-critical application. The Airbrake[62] error monitoring service was used to capture any code exceptions that were thrown in the production environment, and an email notification was sent for each new exception. No major exceptions were reported during the four months that ShopForCloud was operational for; this is a sign of effective unit testing. Heroku's automated database backup service was used to keep seven daily backups, five weekly backups, and ten manual backups, which were used before any major database updates such as re-running the scrapers.

Heroku provides two types of processes, web and worker processes. Web processes are designed to run web server software and were used to run the ShopForCloud Rails application. Worker processes are designed to run background tasks that are managed through a queueing library. The DelayedJob Ruby queueing library[63] was used with

---

[60] http://guides.rubyonrails.org/security.html
[61] http://newrelic.com
[62] http://airbrake.io
[63] https://github.com/collectiveidea/delayed_job

one worker process for ShopForCloud. Two types of background jobs were created; one to run scrapers, and one to run the cost report simulations. The scrapers took two minutes to run when updating all prices, whereas a cost report simulation took around one second to run for a deployment containing ten nodes and two patterns.

Overall, as this chapter has shown, quite a substantial engineering effort was spent on developing ShopForCloud and ensuring that it was usable in a production environment with actual users. The positive effects of this effort were visible during the collection of case study data and user feedback, which are discussed in the next two chapters.

7 Case Studies

A number of case studies were carried out as part of the work presented in this thesis. The initial case study, described in Chapter 3, explored the challenges of cloud adoption. The cloud suitability checklist, the benefits and risks assessment tool and the prototype cost modelling tool were evaluated using four other case studies, which are described in this chapter. The next chapter evaluates ShopForCloud by comparing its features to other tools, measuring its accuracy using an experiment, and discussing the use cases performed by actual users without our involvement.

The first case study presented in this chapter, described in Section 7.1, involves the ageing computing infrastructure of a university department and investigates alterative deployment options for them. This case study focuses on cost modelling and investigates the costs of using public clouds versus buying servers. It shows that running systems on the cloud using a traditional 'always on' approach can be less cost effective, and the elastic nature of the cloud has to be used to reduce costs. Therefore, decision makers have to be able to model the variations in resource usage and their systems' deployment options to obtain accurate cost estimates.

The second case study, described in Section 7.2, involves an academic digital library and search engine, called CiteSeer^X, that indexes a continually growing set of documents from the web. This case represents a highly technical and automated system that is managed by a small team, which can be likened to a small enterprise that is free from the organisational hierarchy and overheads of large enterprises. The case study investigates the benefits, risks and costs of deploying CiteSeer^X on different public IaaS clouds. It shows that there can be significant differences between the costs of using different clouds, and thus highlights the importance of cost modelling during the decision making process.

The third case study, described in Section 7.3, involves the IT systems of the European R&D department of a large media corporation that has over 20,000 employees worldwide. This case represents a typical enterprise division with its own independently-managed systems that are part of a large inter-connected corporate IT environment. Similarly to the previously mentioned case study, this one also investigates the benefits, risks and costs of deploying an IT system on a public cloud.

However, it focuses on comparing alternative deployment options within a cloud, and shows that there can be significant cost differences between alternative deployment options.

An independent case study was carried out by Rongen [101] to investigate the feasibility of deploying of an enterprise application on public clouds. This case study involves BiZZdesign[64], which is a medium-size consultancy firm in the Netherlands providing modelling tools for enterprise architectures and business process management. Rongen used the tools presented in this thesis and provided a critical evaluation of them [101], which is summarised in Section 7.4.

This chapter ends with a summary of the findings from these case studies, and discusses the extent to which the tools developed as part of this thesis were effective in supporting decision makers (Section 7.5). The case studies also show that the benefits and risks of using public clouds need to be investigated from different stakeholder perspectives due to their different priorities and responsibilities. This issue is also discussed in Section 7.5.

7.1 A University Department's Computing Services

The School of Computer Science at the University of St Andrews has around 60 members of staff and 340 undergraduate and postgraduate students. The school provides a number of computing services to its staff and students including:

- Common services such as email, calendar, blog, and web hosting for student projects.
- Storage services such as home directories, backups, and storage of teaching materials.
- Network services such as DNS, VPN, wireless internet and user authentication.

The school has 5 full-time system administrators that maintain its relatively complex IT infrastructure. Some of these systems are interconnected and interact with wider university systems, such as those provided by the university registry and admissions departments. Therefore, the school can be likened to a medium-sized enterprise whose

[64] http://www.bizzdesign.com/

individual systems have evolved over the years to form a mesh of interconnected systems that serve its employees and customers (i.e. the students). The school's computing services are currently deployed on 28 application servers and 5 storage servers in an in-house machine room. There are around 200 desktop machines in the school's computer labs. Some of the school servers are 4 years old and the school is considering upgrading these servers in the near future. The school is considering 3 options:

1. Purchasing new servers to replace the existing servers.
2. Leasing the equivalent amount of resources from the cloud, and migrating its systems but maintaining their existing setup to keep things simple.
3. Leasing resources from the cloud and migrating its systems but changing their architecture to take advantage of the cloud's elasticity to reduce costs.

The tools developed as part of this thesis were used to support the school in investigating the feasibility of migrating some of its computing services to the cloud. A review of the school's computing services was carried out to find out which services would be suitable for migration. The cloud suitability checklist was used as part of this review and the following services were selected as possible candidates for migration:

- Archive: this service is used by all of the school's storage services and has 560GB of data at the moment.
- StaffRes: this service enables staff to store and manage teaching materials that are used for taught courses.
- StudRes: this service provides read-only access to a subset of the StaffRes files for students to access. StaffRes and StudRes have bursty elasticity patterns at the beginning and end of the academic year but are not frequently used during the rest of the year.
- Website: the school is thinking of re-building its website as it is outdated. The site sometimes suffers from slow loading times that could be caused by the university network being over utilized.
- WebDev: this service is used for testing the website when it undergoes major updates, but can also be used as a backup if the main web server fails. The

website rarely receives major updates; therefore, this service has a small usage.

- WebApps: this service includes blogs, public wikis, and software downloads. These applications are deployed on virtual apache hosts within one of the school servers as they have a very small usage.
- Home directories mirror: this service mirrors the home directories service that provides network storage for all school members and applications. The actual home directories service was not considered suitable for migration as the network latency between the school network and the cloud is too high.
- Teaching: this service is used to host student projects for various courses that require server-side technologies such as MySQL or Apache. This service is only used during term time, which is 24 weeks per year.

Collectively, the above services are currently deployed on 9 application servers and 3 storage servers. The remaining school services are unsuitable for migration as they either control the school network (e.g. the DNS server), or they need low network latencies that make them unsuitable for access over the internet (e.g. day-to-day network storage). A few services require access to hardware or network infrastructure; therefore, they are also unsuitable for migration (e.g. the network monitoring service).

Figure 35 shows the deployment model, built using the prototype cost modelling tool, that was created to represent the school's services that are being considered for migration. The model represents the school's network as a remote node that communicates with a monitoring server on the cloud. The previously mentioned services are modelled as applications and data, which are deployed on servers and storage nodes. Interdependencies between the applications and other services have been deliberately left out of the diagram to keep things simple and understandable. In practice, these interdependencies do not need to be taken into account during cost modelling as they do not affect costs. The main connections that affect costs are data transfer links, which have been included in the model.

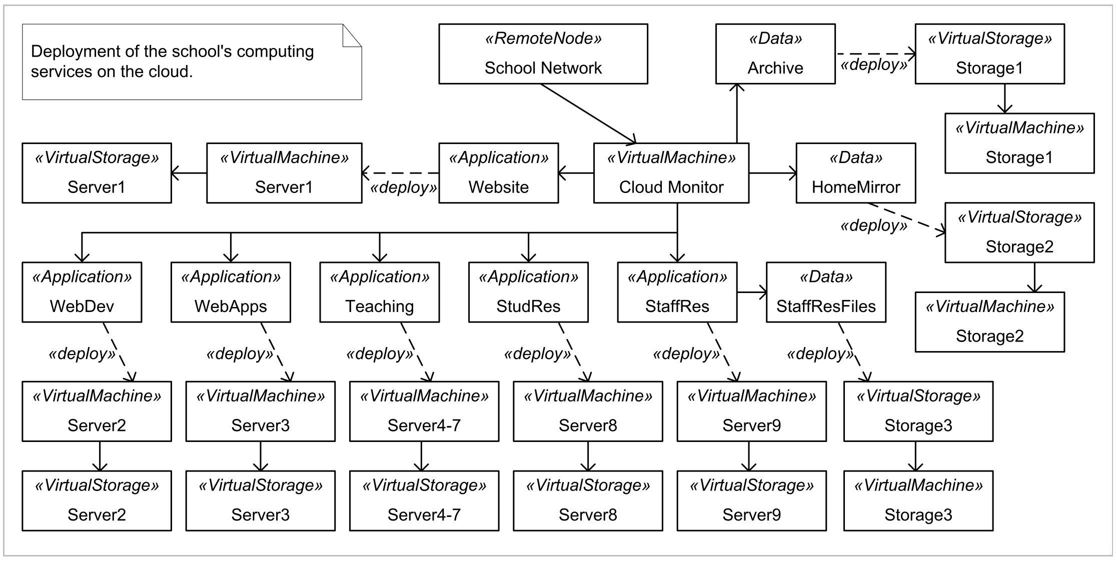


Figure 35: Overview of the school's systems being considered for migration

7.1.1 Results

The cost modelling tool was used to compare the costs of the school's options over a 6 year period starting from September 2010 (i.e. the start of the academic year). The school is considering using the AWS EU cloud; therefore, the cost estimates presented in this section are based on AWS's prices.

Option 1 - Purchasing physical servers: 9 application servers and 3 storage servers would be required to replace the existing servers. A mid-range application server costs around $1550 in the UK (e.g. a Dell PowerEdge R410 with an Intel Xeon 2GHz quad-core CPU, 2GB RAM and two 250GB hard drives configured in RAID1 to give 250GB usable storage). A mid-range storage server costs around $2500 (e.g. a Dell PowerEdge R510 with an Intel Xeon 2GHz quad-core CPU, 4GB RAM and five 250GB hard drives configured in RAID5 to give 1TB usable storage, with an extra disk as a hot spare). Electricity costs would be $106 per year for each application server and $155 per year for each storage server (based on energy usage estimates from Dell[65], assuming a 10% CPU load and a cost of $0.1 per kWh, which is what the school pays). Cooling and network infrastructure costs do not need to be considered as the school already has these facilities in its machine room for the existing servers. The costs of purchasing physical servers were calculated using a 3-year upgrade cycle where the school would pay the same upfront capital to upgrade to new servers in

[65] http://solutions.dell.com/DellStarOnline/DCCP.aspx

year 4. This is a reasonable upgrade cycle for the purposes of cost comparison as any server failures during this time are covered by Dell's basic 3-year guarantee.

Option 2 - Leasing equivalent resources from the cloud: Leasing the equivalent amount of resources in option 1 from the cloud would require 12 High-CPU Medium instances from the AWS EU cloud (using 'reserved 3-year' instances to reduce costs). The reserved instances option would have to be renewed in year 4 to keep the instance costs low. Similarly to option 1, each application server would have a 250GB EBS volume, and each storage server would have a 1TB EBS volume. The number of I/O operations were measured on the existing servers and these values were input into the cost modelling tool. In addition, it was estimated that 200GB of data would be transferred into the cloud each month, and 200GB would be transferred out each month.

Option 3 - Using the elasticity of the cloud: The resource usage of the existing servers was reviewed and the cost model that was created for option 2 was modified to include the school's actual resource usage. This involved defining patterns to switch-off instances when they were not in use. For example, the baseline number of instances for the teaching service was set to 0, and the pattern was set to `[temporary: every.1.year on sep-nov +4, temporary: every.1.year on feb-apr +4]` to show that 4 servers would be required during term-time. Three of the school's services did not require the High-CPU Medium type of instance as they had a small usage; therefore they were deployed on Standard Small instances. In addition, the storage servers were replaced by Amazon's S3 service, and elasticity patterns were defined to show how the school's storage demands increase over time. For example, the baseline storage of the archive service was set to 560GB, and the pattern was set to `[permanent: every.1.year on every.1.month +15]` to show that 15GB of extra storage would be required every month.

AWS has previously changed their pricing scheme for some their services, for example in November 2009 they lowered the price of all on-demand instances by

15%[66]. Therefore, it is useful for decision makers to consider the cost of their systems if cloud providers change their pricing scheme in the future. The school was interested to see how the cost of their system would change, if in 2 years time, AWS:

1. Increases instance-hour and storage prices by 15% due to rising energy costs.
2. Decreases instance-hour and storage prices by 15% due to Moore's Law and more powerful hardware coupled with increasing competition from other cloud providers.

Figure 36 shows how much the school would be paying for each option over the 6-year period that is being investigated. At the start (i.e. year 0), it would either cost $22,800 to buy physical servers (includes electricity usage for first year) or $23,300 to lease equivalent resources in the cloud. However, if the system is modified to use the elasticity of the cloud, then the starting cost would be $9,900. Figure 36 shows how the costs vary over the remaining years, for example in year 1, the elastic option would cost $6,700 compared to $1,400 for the electricity usage of the buy option.

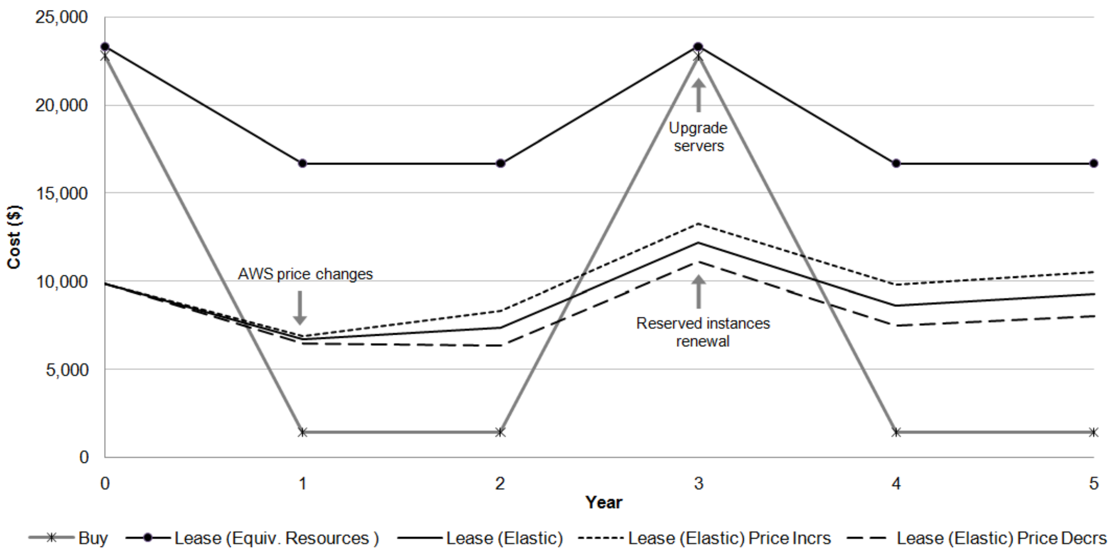


Figure 36: Yearly cost of different options that the school could take

To compare the school's options financially, their *net present values* (NPV) have to be calculated over the 6 years. NPV is often used by organisations to compare the overall value of different investment options by taking into account their incoming and outgoing cash flows [102]. Since the school does not make explicit profits from

[66] http://aws.amazon.com/about-aws/whats-new/2009/10/27/announcing-lower-amazon-ec2-instance-pricing/

its computing services, the incoming cash flow can be ignored. NPV calculations take into account the *cost of capital*, which is the return rate that capital could earn in an alternative investment option [102]. For example, the school could put the upfront capital into a bank savings account and earn interest if they choose a lease option. Assuming a 5% return rate, each cost, C, at year Y in Figure 36 has to be set to: $C = C / (1 + 0.05)^Y$. These costs then have to be summed to give the NPV of each scenario, which is shown in Figure 37. The percentage differences between the buy option and all other options are also shown in the figure.

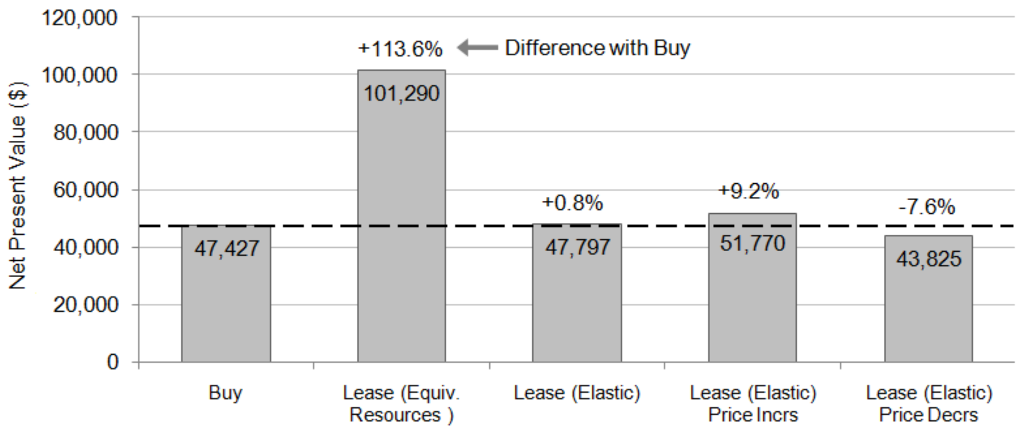


Figure 37: Net present value of different options that the school could take

It should be noted that a higher return rate favours the cloud option as future costs become more rewarding than upfront costs. Surprisingly, it can be seen that the elastic option is slightly more expensive than buying physical servers for the school. Leasing equivalent resources from the cloud and leaving them running 24x7 therefore makes no financial sense as it costs more than twice the buy option. However, if AWS reduces prices by 15% in 2 years time, then the elastic option becomes the cheapest option.

We did not explore the possible option of buying fewer physical servers and using virtualization to run several servers on one machine. This would certainly have reduced the overall costs of purchase but would incur additional local setup costs. Nor did we take account of any changes to staffing required – in practice, we do not think that there would be any significant reduction in support costs.

7.1.2 Discussion

The results demonstrate that the output of cost modelling helps to inform decision makers during the migration of IT systems to the cloud. The output recommended that the school should buy physical servers if they have the upfront capital. If not, then they should lease resources from a cloud provider but re-architect their system to use the cloud's elasticity, otherwise the costs would be higher than buying physical servers. However, the cost of this re-architecting would also have to be taken into account.

We found that due to situation specific factors, the results of the tool needed to be supplemented. For example, the *opportunity cost* of the buy option's upfront capital should also be considered. That is the benefit that the school would have received if they had used that capital to take an alternative action [102]. For instance, as the elastic option needs 60% less capital upfront, the remaining capital could be used for other investments such as improving facilities or increasing the publicity budget to recruit more students.

Another factor that should also be considered is the cost of infrastructure support and maintenance. Servers that are used for business critical applications often require expensive support and maintenance contracts with hardware suppliers that guarantee response times to support calls. Cloud providers are beginning to address this demand as well, for example AWS has a premium support package[67] that guarantees a one-hour response time for urgent issues. The individual services offered by each cloud provider and their service level agreements (SLA) should also be considered, in addition to the compensation that is provided if they fail to meet their SLA.

An important limitation to any cloud cost estimation approach (including cost modelling) is the need to have fairly accurate estimates of resource usage, as the estimated costs are sensitive to inaccuracies. For example, in this case study we identified that if a summer school is being run and it requires the use of 4 servers, then these would have to be leased from the cloud. In contrast, if the school already has 4 teaching servers that are not utilized during the summer, then those could be used. With traditional infrastructure provisioning enterprises do not have to worry too much

[67] http://aws.amazon.com/premiumsupport/

about usage patterns when they buy servers as they are often underutilized [103] and can accommodate temporary peaks. However, public IaaS clouds could be perfect for situations where usage patterns are unknown, and resource needs cannot be assumed to be met by underutilized servers.

For example, if the school is considering introducing distance-learning courses but they do not know the level of demand for such courses, then using a cloud makes financial sense as buying servers would be too risky. If eventually there is enough demand that is fairly constant and continuous in nature, then it could be case that the school could actually save money by migrating the courses into in-house servers.

The results of this case study can be understood within the context of our existing work as re-enforcing our arguments that decision makers should not rely solely upon financial data when making decisions pertaining to the adoption of cloud. Our previous case study that investigated the migration of system infrastructure to Amazon EC2 in an Oil & Gas IT company showed that the system infrastructure would have cost around 37% less over 5 years on EC2 compared to the in-house data centre (Section 3.1).

In contrast, the results of this case study showed that despite popular beliefs of cost savings in the cloud, there is not much difference between the costs of buying physical servers and the costs of deploying some of the school's IT systems on the cloud. The difference between the two case studies is that the system mentioned in Section 3.1 was a green-field development project; therefore new network infrastructure had to be purchased. It could well be that the cloud is a cheaper option for an enterprise that needs more than say 30 servers, due to the extra costs of racks, cooling, and network infrastructure that would be required for physical servers. The difference between these two case studies highlights the importance of cost modelling to enable decision makers to investigate the costs of deploying their specific systems on the cloud.

However, it should be noted that despite the favourable cost analysis in our previous case study (with the Oil & Gas IT company) it was decided not to migrate to the cloud due to benefits and risks relating to organisational change. These benefits can also be observed in this case study. The system administrators would be freed from maintaining hardware and can focus on supporting applications. The load on the school's internal network could potentially be reduced as requests would be sent to the cloud.

However, there are also barriers to using the cloud, mainly the migration of data and applications, which requires the system to be re-designed to use elasticity. System administrators will require some training for this but it should not require too much effort as only the management of the infrastructure is affected (e.g. AWS' APIs need to be used).

Using elasticity for the school's systems should be fairly straightforward as it will involve switching-off instances that are not in use, and using Amazon S3 is inherently elastic as storage does not need to be provisioned beforehand. In contrast, using elasticity can be challenging and expensive to achieve for interconnected enterprise systems that rely on other systems being available all the time, or for systems that use relational databases that cannot be easily scaled out. There are also the usual security that are often raised (discussed in the literature review), but such issues are not significant for this case study as data would be encrypted before being transferred to the cloud for storage. Other data, such as teaching material, are already available on the web and therefore the implications of storing them in the cloud would be no different.

7.2 An Academic Digital Library and Search Engine

This case study represents a highly technical and automated system that is managed by a small team. The team can be likened to a small enterprise that is free from the organisational hierarchy and overheads of large enterprises. The system under investigation in this case study is a digital library and search engine, called CiteSeer^{X}, that indexes academic papers, and enables users to search and access papers via the web. This system uses a service-oriented architecture and has the following components: a web application that is the main interface between the system and its users; document crawling and metadata extraction; ingestion, which processes the crawled documents and updates the system repository; a maintenance service that updates the indexes and generates relevant statistics; and finally a data backup and replication service. These components are deployed on 15 servers that are housed in a university datacentre. The system receives around 2 million hits per day and contains over 1.5 million documents requiring around 2TB of storage. The applications and infrastructure of the system were modelled using the prototype cost modelling tool, which, as shown in Figure 38, represents the model as a graph internally.

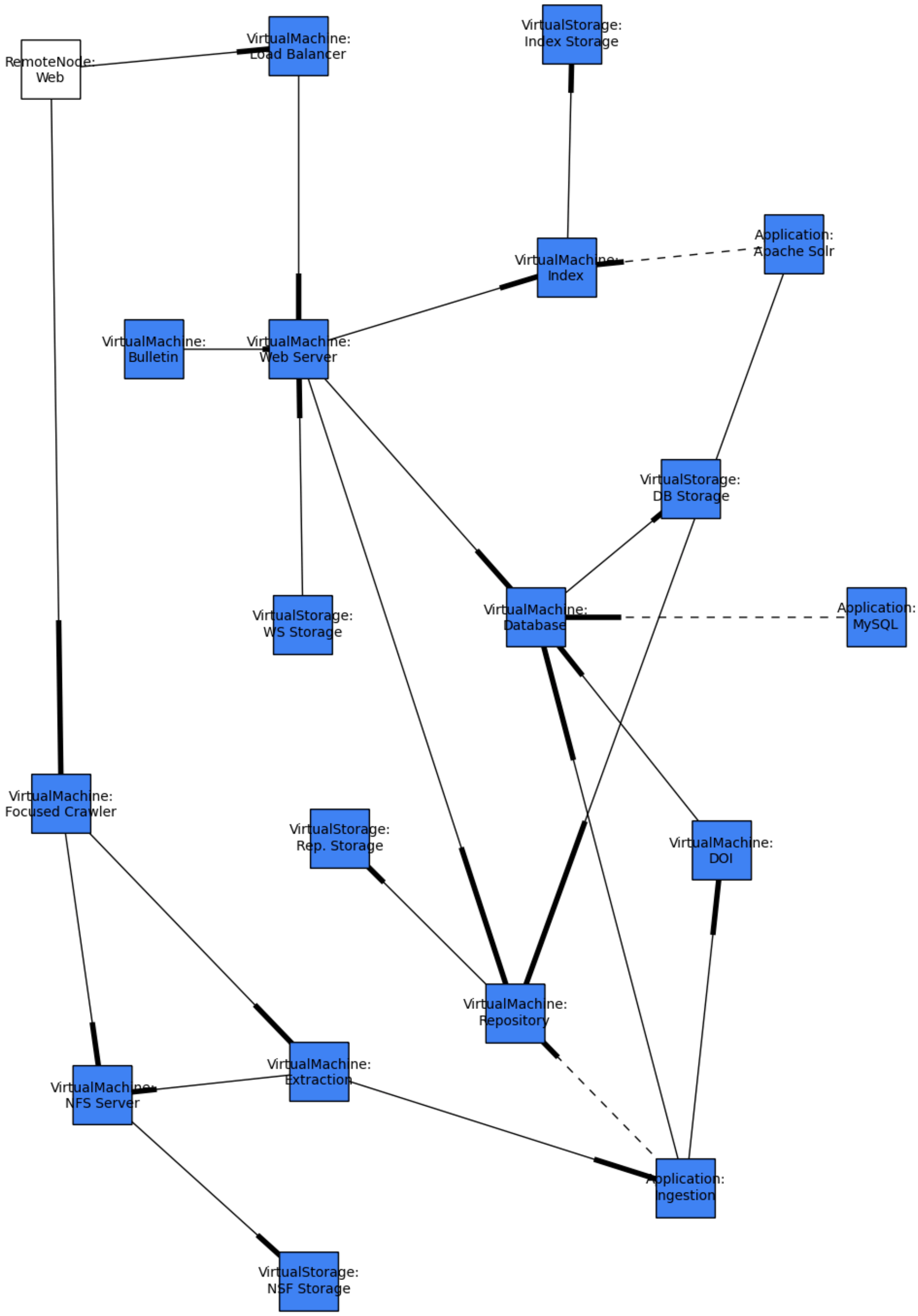


Figure 38: The cost modelling tool represents the UML diagrams as a graph internally (described in Chapter 5).

Based on historic data, a number of elasticity patterns were also created to model the growing resource needs of the system. These patterns included a 15GB/month increase in data transferred out of the system (caused by increasing number of visitors), and a 17GB/month increase in the size of the document repository.

The cost of using different cloud providers was investigated, and as shown in Table 15, AWS was found to be the cheapest. The significant price difference between the providers was unexpected as it is argued that cloud providers compete on price [90].

Table 15: Cost of using different public IaaS clouds

Cost ($)	***AWS US-East***	***FlexiScale***	***Rackspace USA***
First month	18,980	5,060	6,550
Monthly average	1,916	5,151	6,732
Total for 3 years	85,950	185,345	242,170
Difference with AWS		**+2x**	**+3x**

Cost modelling highlighted some interesting points:

1. Performance must also be considered but the performance vs. cost trade-off of IaaS clouds is an open research challenge as traditional benchmarking approaches are unlikely to be adequate since their results depend on the applications running and network-load of the cloud at that time; hence they are not easily generalizable.
2. The cost breakdown of AWS showed that 66% was for VMs, 20% was for data transferred out of the system, 10% was for storage, and 4% was for storage I/O requests. This information can be used to optimize the system for cost, e.g. switching off VMs when they are not in use is a simple but effective cost cutting technique.
3. Using AWS's S3 storage would cost around $4,000 more than using EBS (over the 3 years). This is due to the increased cost of storage I/O requests from S3. However, this estimate might be inaccurate as it is difficult to calculate the storage I/O requests for S3 by looking at disk I/O figures from Linux's `iostat` command.
4. Storage I/O is priced differently by providers: AWS charge for number of I/O requests to storage, FlexiScale charge for the amount of data transferred

to/from storage, Rackspace do not charge for input requests for files over 250KB in size, and GoGrid do not charge at all.

The benefits and risks assessment tool was also used as part of this case study; this is discussed in Section 7.5.

7.3 The R&D Department of a Large Media Corporation

This case study represents a typical enterprise division with its own independently-managed systems that are part of a large inter-connected corporate IT environment. The systems under investigation in this case study belong to the European R&D division of a large media corporation that has over 20,000 employees worldwide. This division is responsible for research and development in software applications that are used by the corporation. The division has around 40 office and management personnel, each with a laptop that is used for office applications and occasional consultation of software development functionality. There are also around 120 engineers, each with a workstation that is mainly used for software development and testing.

The R&D division uses a range of applications including IBM ClearQuest (software change management), IBM DOORS (requirements management), IBM ClearCase (software configuration management), Klocwork (static source code analysis) as well as a number of databases (SQL Server) and custom-made websites. These applications are deployed on 9 heavy-duty servers and 2 network storage systems that are housed in the local office.

The R&D division is currently thinking about its future infrastructure strategy and is interested in finding out the infrastructure costs of using AWS so that it can compare it with other options (e.g., Eucalyptus private cloud). As shown in Figure 39, the current setup of their applications and infrastructure was modelled using the prototype cost modelling tool.

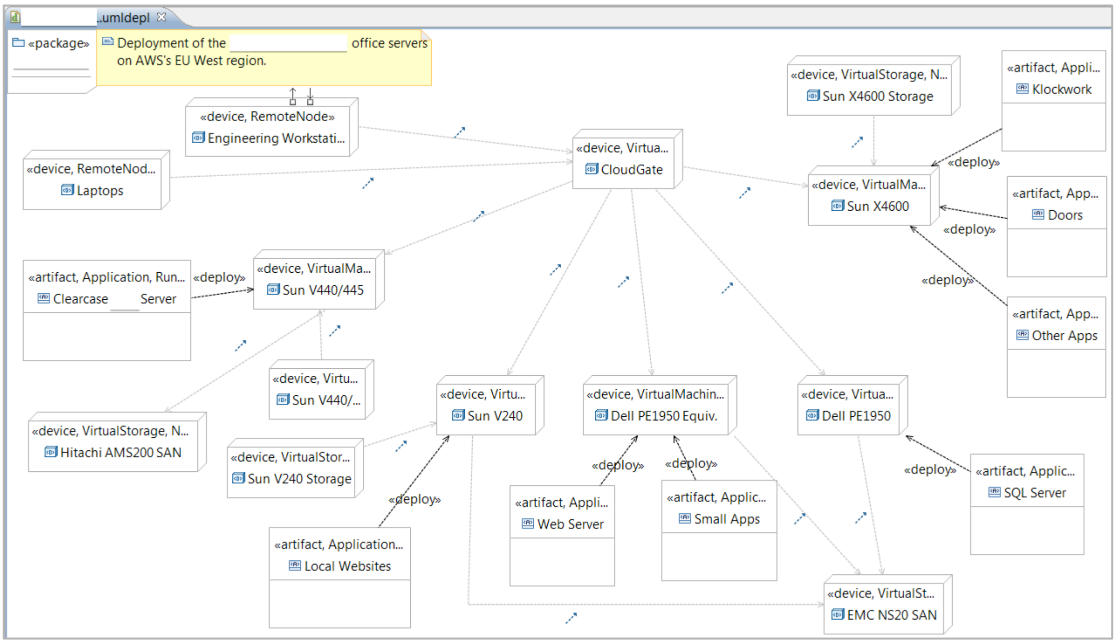


Figure 39: The cost modelling tool was used to create a model of the R&D division's infrastructure that was considered for cloud migration

The current setup is based on a client-server architecture, where users have powerful machines that are used for most of their work including code compilation. This setup results in the servers having a moderate load. The workstations are not usually under a heavy load but at times, users have to enlist the workstation of their neighbours to speedup code compilation.

The division is considering moving to a thin client setup where each engineer would have a standard office laptop from which they could connect to centralized servers from their home, office or a customer site. This new setup would include the current servers as well as new session servers for users to login to and work, and compute servers that are used to off-load heavy processing or host applications that have dedicated resource needs. The costs of three options were investigated:

1. Using AWS instances in a non-elastic manner by leaving them on 24x7.
2. Using AWS instances in an elastic manner where on working days, the session servers would be switched on during day-time and compute servers would be switched on during night-time.
3. Using AWS instances in an elastic manner as in option 2 but giving each user a small instance rather than using a few large instances to host all user sessions.

Using option 3 would mean that a user could switch on/off their instance at anytime without worrying about who else is using that instance, hence the small instances would be on for 8 hours per working day. The costs of the different deployment options are shown in Table 16.

Table 16: Cost of different deployment options on AWS-EU cloud

Cost ($)	***Non-elastic***	***Elastic***	***Elastic, small instances***
First month	67,350	65,430	75,260
Monthly average	6,259	4,344	4,175
Total for 3 years	**286,415**	**217,470**	**221,385**

Cost modelling highlighted some interesting points:

1. Typical enterprise divisions do not currently need to monitor their actual computational resource consumption in the level of detail that is required for cost modelling. Since servers are either procured upfront or leased from an IT vendor under a provisioning and support contract, the actual resource consumption does not affect costs in a significant manner. This issue highlights the need for monitoring tools that provide resource usage estimates that can be fed into cost modelling tools.
2. In contrast to popular belief, using AWS can have a fairly high start-up cost, which is used to reserve instances that reduce monthly costs in the long-term. In this case, it was around 30% of the total costs over the 3 years.
3. The elasticity of the cloud can be used to reduce costs. In this case, the difference between the non-elastic and elastic setup would be around $70,000 over the 3 years.
4. Different deployment options have to explored to find the cheapest. In this case, using 4 high-memory quadruple-extra-large instances costs around $4,000 less than using 120 small instances (one for every engineer). However, using small instances is more flexible as instances can be switched on when individual engineers need them, whereas a large instance has to be switched on even if only one engineer is using it. Whenever possible, enterprises should use small instances that enable them to increase or decrease their resource consumption by small chunks to avoid having under-utilized servers.

The benefits and risks assessment tool was also used during this case study; this is discussed in Section 7.5.

7.4 Dutch IT Consultancy

BiZZdesign is an IT company with around 100 employees; they are headquartered in the Netherlands but they also have offices in five other countries including the UK. BiZZdesign develops consultancy tools for enterprise architectures and business process management, in addition to offering consultancy, training and support services to other IT companies.

Rongen [101] carried out an independent case study with BiZZdesign. It should be emphasised that we were not involved in this case study, and Rongen worked from the information in our publications only [13], [95], [104]. His work aimed to support them in investigating the feasibility of migrating their *InSite* tool[68], an enterprise architecture collaboration and communication software package, to the cloud. InSite is currently installed and maintained by customers on their internal IT infrastructure. However, this is a hassle for customers and can take a long time to install, which results in some customers not using the software. BiZZdesign would like to address this issue by offering InSite as Software-as-a-Service; this also presents the commercial opportunity of expanding their market. The outcome of the case study was a report written for BiZZdesign, which was evaluated by conducting a semi-structured interview with their CTO.

InSite is developed as a Java servlet that runs within Apache Tomcat and can connect to various types of databases including MySQL, SQL Server and Oracle. The tool uses a REST and a SOAP interface to communicate with backend servers, while the web interface is developed using Adobe Flash. Therefore, it is possible to deploy InSite on any public IaaS cloud. However, a separate deployment would be required for each customer as the software is not multi-tenant.

7.4.1 Cloud Suitability Checklist

The cloud suitability checklist and the benefits and risk assessment tool were used during the course of six weeks that involved software developers, sales staff and

[68] http://www.bizzdesign.com/tools/insite

managers at BiZZdesign. The cloud suitability checklist was used during a session with two software developers, where the first four items of the checklist (elasticity, network, processing, and access to hardware) were discussed. Furthermore, a questionnaire was created based on the last four items of the checklist and sent to 12 customers to find out more about their requirements regarding availability and dependability, security, data confidentiality and privacy, in addition to regulations and compliance. Following these discussions, two cloud providers (Rackspace UK and Cloudsigma Switzerland) were selected for further investigation where tests were carried out to confirm the technical feasibility of running InSite on their clouds. These clouds were selected as the data saved in the InSite application is highly confidential, and most of BiZZdesign's customers' IT policies do not allow data to be stored outside of Europe. Overall, Rongen's evaluation [101] of the cloud suitability checklist concluded that:

BiZZdesign indicates the checklist used is useful as a tool to do a quick assessment of the requirements for the application. It provides a starting point from where the most problematic issues can be further investigated. This aids in shortening the possible list of providers and reduces the time and effort required to get the right information from the right providers.

Rongen also mentioned that while the checklist highlights potential issues with deploying application on public clouds, it does not do a good job of highlighting the suitability of a given cloud for a particular application. This requires the specific cloud services of a provider to be examined to assess its suitability for a given application.

7.4.2 Benefits and Risks Assessment

The benefits and risks assessment tool was used by five software developers (senior and junior) and two members of the management team during the case study. Participants were asked to go through the 19 benefits and 39 risks and score each one using a 1 to 10 scale to signify their importance. They knew about the InSite cloud feasibility study but were not informed of the two selected clouds upfront to avoid creating a bias. The results were analysed by averaging the scores across each group of participants (developers and managers) and per risk/benefit category (technical, security, financial, organisational, legal). The top three benefits and risks were identified and included in the final report for BiZZdesign.

The top three perceived benefits were:

1. Anywhere/anytime/any device (desktop, laptop, mobile etc.) access to computational resources and applications can be setup without too much effort. This in turn simplifies collaboration amongst users and simplifies application support and maintenance.
2. Opportunity to offer new products or services or trial products to gauge the level of interest from customers.
3. Reduced risks of technological obsolescence as cloud providers update the infrastructure.

It is interesting to note that the risks and benefits assessment exercise triggered discussions within the developer group about their application architecture, and how the benefits would be far greater if they changed their architecture to use the cloud's characteristics.

The top three perceived risks were:

1. Lack of information on jurisdictions used for data storage and processing. Leads to non-compliance with regulations that require certain types of data to be kept in national boundaries (e.g. the eighth principle in the UK's Data Protection Act 1998 that requires personal data to be kept within the EEA).
2. Non-compliance with data confidentiality regulations. Unauthorised access to data by cloud providers.
3. Performance is worse than expected (e.g. CPU clock rate, I/O and network data transfer and latency rates). It might be difficult to prove to the cloud provider that their system performance is not as good as they promised in their SLA as the workload of the servers and the network can be highly variable in a cloud. This might lead to disputes and litigation.

A major difference was noticed between the scores given to the legal and organisational risks between the developers and managers. For both categories, developers scored the risks higher than the managers. Rongen mentions that "this could be explained by better insight and more perceived control by management in these fields" [101]. The risk assessment exercise also highlighted the need for data encryption and a decision was made to run a test case on a larger scale. The BiZZdesign CTO and Rongen also investigated the roles and responsibilities involved in deploying and maintaining their application on the cloud as part of the risk assessment exercise. Overall, Rongen's evaluation [101] of the benefits and risks assessment tool concluded that:

This tool inspired all participants to consider the risks and benefits in more detail and really understand the way the cloud works. It gives a decent view into the organisations views toward the cloud. Because the participants are not educated or experienced with the cloud it can't provide a reliable view of the true benefits that will be experienced. BiZZdesign and the researcher feel this step [the risks and benefits assessment] should be executed first. It stimulates involvement and indicates what is considered to be important within the organisation.

Rongen also mentioned that although the tool is very effective, the participants did not understand some of the benefits and risks mentioned in the tool. These benefits and risks were identified and needed rephrasing to provide a more comprehensive description.

7.4.3 Cost Modelling

Rongen did not contact us as part of his case study and therefore did not have access to the prototype cost modelling tool (ShopForCloud was still under development during this time). Rongen performed cost modelling manually (presumably using a spreadsheet) and compared the costs of using the Rackspace UK and Cloudsigma Switzerland clouds. Rongen reported a 30% cost difference between the two cloud providers [101]. It is interesting to note that during the cost modelling exercise, BiZZdesign indicated that their customers also need a cost modelling tool and they are investigating the possibility of developing such functionality as part of the enterprise architecture modelling tools.

7.4.4 Overall Evaluation

Rongen and BiZZdesign expected the tools to provide a low-effort method that covered the issues that need to be investigated during cloud adoption. By low-effort, they meant a task that could be done over a couple of days and not use too many resources at the company. Rongen believes that the tools should be used in an iterative process, where the effort in each iteration increases but the results are more concrete and take the form of a go/no go nature. The use of the tools highlighted the need for BiZZdesign to re-architect their InSite application in contrary to their first intention, which was not to take on a re-development exercise. However, the short term benefits of using the cloud for testing and demonstration purposes was also highlighted to them.

Overall, BiZZdesign "was impressed with the results and indicated the tools provided great value […] as they were easy to use and made sure the process was both efficient and effective. It created basic knowledge about the cloud by involving the right people and performing a technical examination" [101]. Rongen commented that organisations that have more experience with cloud computing are likely to use the tools as more of a checklist to ensure all bases are covered during cloud adoption decisions (BiZZdesign did not have experience with cloud computing).

7.5 Discussion

All of the case studies described in this chapter involved organisations that were dealing with existing systems that were possible candidates for cloud migrating for various reasons. The case studies used the tools in different ways:

1. The first case study (a university department's computing services) focused on investigating the costs of using a public IaaS cloud versus buying servers.
2. The second case study (an academic digital library and search engine) focused on investigating the costs of using different cloud providers.
3. The third case study (the R&D department of a large media corporation) focused on investigating the costs of using different deployment options within a cloud.

4. The fourth case study (a Dutch IT consultancy) focused on investigating the feasibility of cloud migration and identifying the major benefits and risks involved from different stakeholder perspectives.

It is usually not feasible for organisations to migrate all of their existing systems to a public cloud. However, organisations with so-called “green field” systems could use the tools presented in this thesis to investigate their system deployment options in public clouds. Organisations that do more “brown field” system development can also use these tools to investigate their system deployment options, however, such systems are often integrated with legacy systems and their migration will not be as straightforward. The case study with the university’s computing services illustrated this point, where migrating all of the systems was not feasible, however, individual systems could have been considered for migration (e.g. the backup and archiving system). Although the primary focus of the first three case studies was on costs, they also used the benefits and risks assessment tool as discussed next.

7.5.1 Benefits and Risks Assessment

Following the first case study, the benefits and risks assessment tool was provided to the head of system administration at the university department involved in the case study and his feedback was as follows:

I think I spent about 30 minutes on the spreadsheet. Overall I think that the spreadsheet is useful. We are currently considering using cloud server instances (probably EC2 micro instances) to provision Tomcat for software engineering and distributed systems teaching. I could build a business case based solely on the risks and benefits in the spreadsheet.

During the second case study, the benefits and risks assessment was performed by the technical director of CiteSeer^{X}. The exercise took around one hour, during which 7 benefits and 13 risks were identified as important. CiteSeer^{X} found the benefits and risks assessment useful as they had not considered the availability and security risks of using public clouds to the extend that were highlighted by the assessment tool.

During the third case study, the assessment was carried out by the IT manager of the media corporation’s R&D department who provided a local view of the benefits and risks. The exercise took around two and a half hours and included discussions with one of their senior software engineers. Furthermore, the benefits and risks assessment tool was used by one of the IT managers who works directly with corporation's CIO; this provided a corporate view of the benefits and risks. The technical analysis revealed 8 important benefits and 25 important risks; the organisational analysis suggested 4 important benefits and 15 important risks.

As discussed in Chapter 4, it can be useful to get a holistic picture of the benefits and risk from an enterprise perspective. One way to do this is to calculate the weighted average of the benefits/risks and chart them on a radar graph, as shown in Figure 40 and Figure 41. This weighted average can be useful as it provides a high-level view of the different types of benefits/risks that an organisation faces. However, generally a small number of benefits/risks dominate because of organisation-specific considerations, thus the weighted average should not be seen as a definitive overview of the benefits/risks.

The weighted average can be calculated by multiplying the number of benefits/risks in each category (organisational, legal, security, technical or financial) by the weight of each benefit/risk (unimportant = 1 … very important = 5), and dividing the result by the total number of benefits/risks in that category.

Figure 40 shows the weighted average of the benefits for the second and third case studies. It shows that in the case of the digital library, the technical benefits of using public IaaS clouds were more important than the organisational and financial benefits. Hence, the technical ability to deal with volatile demand patterns and cater for a growing number of users would be one of the main motivations for using the cloud. In the case of the media corporation, the R&D department also views the technical benefits as more important than the organisational and financial ones. From their perspective, the simplified provisioning of computational resources, and anywhere/anytime access to resources are some of the motivations for using the cloud. Whereas it is clear that their corporate IT department views financial and organisational benefits as more important than technical ones.

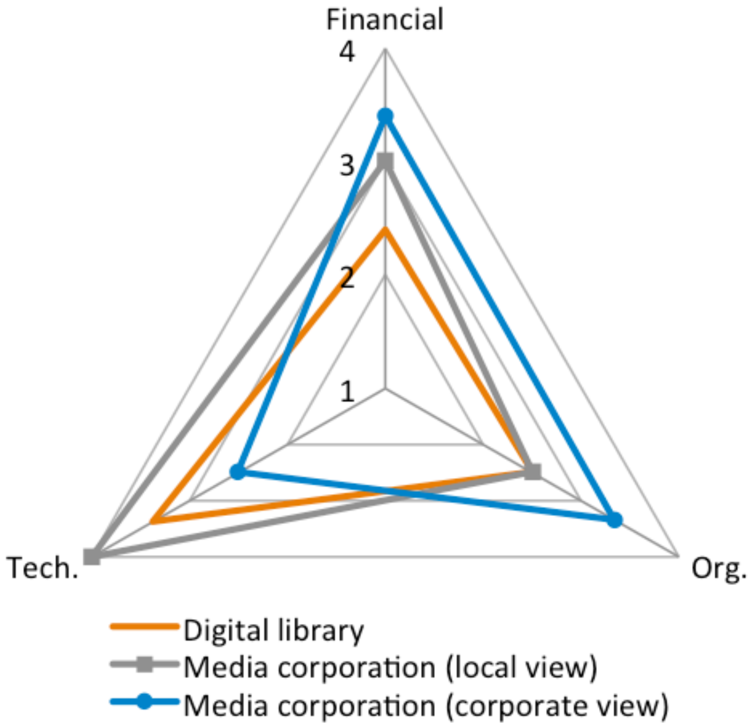


Figure 40: The importance of the different types of benefits of cloud migration in case studies presented in this paper (1=unimportant, 2=little important, 3=moderately important, 4=important, 5=very important)

Figure 41 shows the weighted average of the risks of cloud migration. In the digital library case study, it is clear that the main risks are financial, technical and security related; legal and organisational risks are not important as are not relevant to small enterprises that deal with non-sensitive data. The R&D department of the media corporation has a good understanding of their local systems and appreciates the importance of the different types of risks, whereas their corporate IT department is mostly concerned with the organisational and legal risks. This is understandable as corporate IT are probably best equipped to deal with those risks, however, they are not too concerned with financial risks as budgets are probably managed by local divisions.

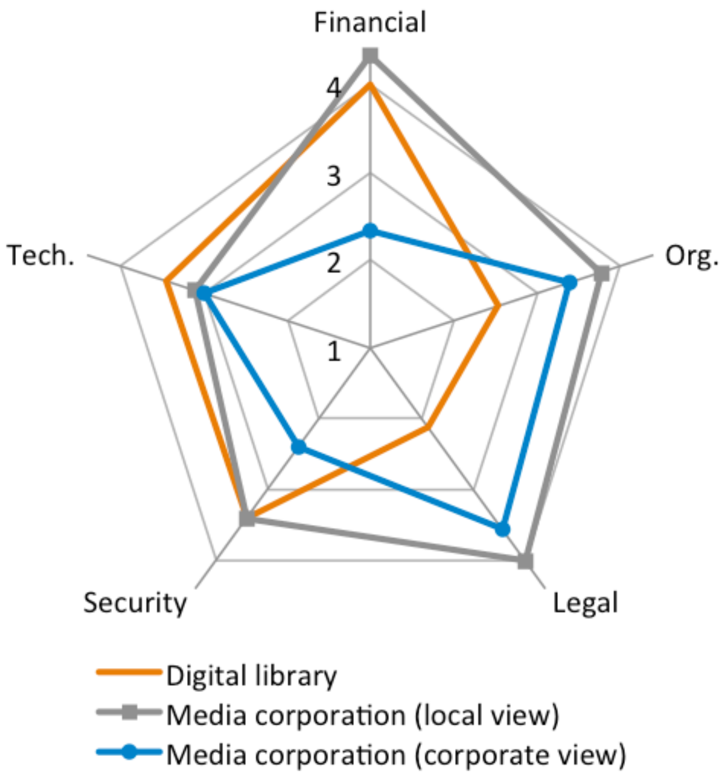


Figure 41: The importance of the different types of risks of cloud migration

The case studies show that there can be multiple perspectives on the benefits/risks of cloud migration, and even within a single enterprise, different divisions have different views. Motivations and concerns of different stakeholders have to be considered during cloud adoption decisions. The digital library might use public IaaS clouds for a subset of their system components. The media corporation's R&D division is unlikely to migrate its systems to public IaaS clouds due to the identified risks; however, it might use the cloud for specific functions such as offsite backup and disaster recovery.

Overall the four case studies have shown that the cloud suitability checklist and the benefits and risks assessment tool collectively provide organisations with a starting point for risk assessment as they inform people about the various benefits and risks. Although the tools do not provide a framework for a yes/no decision to be made, they do trigger discussions within organisations and provide a low-effort means to gather the perspectives of different stakeholders in an organisation. This can be seen as the most beneficial aspect of using these tools.

7.5.2 Cost Modelling

The case studies presented in this chapter demonstrate that the cost modelling tool addresses the challenges of cloud cost estimation (as defined in Section 3.5), and that the output of cost modelling informs cloud adoption decisions by providing important information to decision makers. The case studies also demonstrated that the cost modelling tool is scalable and can model the size of systems that are typically found in enterprises to be suitable candidates for cloud migration.

One of the limitations of the case studies presented in this chapter is that they only discuss infrastructure costs of using public IaaS clouds. Cost modelling has to be used in conjunction with project management and software cost estimation techniques to enable the full costs of cloud migration to be investigated. Furthermore, most large organisations have already invested in existing IT infrastructure, or have signed multi-year lease contracts with IT vendors. In such cases, it can be difficult to justify the use of public IaaS clouds over existing IT infrastructure. The use of private clouds, such as Eucalyptus or CloudStack, might be appropriate in such organisations. The cost of using private clouds is much more predictable as they are usually licenced on a per-CPU-core basis, but would also require the use of additional tools such as RightScale's cloud management platform that is used by many organisations to manage their private clouds.

Cost modelling requires detailed information about the actual resource consumption of applications, such as disk I/O rates and data transfer estimates between different systems. Some organisations do not have this level of information available to them, as traditionally this did not matter since servers were over-provisioned and procured upfront. In such cases, cost modelling can be performed on a range of scenarios to present decision makers with a range of cost estimates rather than a single figure. What-if style cost analysis can also be useful in such cases to examine the costs of different business cases or growth rates.

Overall, the four case studies presented in this chapter showed that:

1. The cost effectiveness of using public clouds is situation dependent rather than universally less expensive than traditional forms of IT provisioning.

2. Using public clouds can be financially attractive if their elasticity is utilised. However, this usually requires enterprise applications to be re-architected as older applications were either not designed as multi-tenant systems, or they cannot scale-out to use multiple servers as they rely on the server to be scaled-up to provide more resources (e.g. CPU and RAM).
3. Performing upfront cost modelling is beneficial as there can be significant differences between different cloud providers, and different deployment options within a single cloud.
4. During such modelling exercises, a system's elasticity patterns must be taken into account to produce more accurate cost estimates, and the notion of elasticity patterns that was presented during this thesis provides one simple way to do this.

8 Evaluation of ShopForCloud

The case studies presented in the previous chapter were one aspect of the cost modelling tool's evaluation. They provided a deep insight into how the tool could be used in practice and showed that there is value in modelling the deployment costs of a system to compare different cloud providers and deployment options. They also demonstrated that the notion of elasticity patterns is a useful method of modelling a system's elasticity to obtain more accurate cost estimates.

This chapter provides a further, quantitative and qualitative evaluation of the cost modelling tool by describing an experiment that was used to measure the accuracy of ShopForCloud's cost estimates (Section 8.1). The usage data that was obtained during the four-month operation of ShopForCloud provides many insights into how the tool was actually used in practice. This data is summarised and presented in Section 8.2, where the lessons learned from the data are also discussed. Finally, Section 8.3 presents the direct feedback that was received from ShopForCloud users.

8.1 Accuracy of Cost Estimates

The test cases developed during the design and implementation of ShopForCloud (Chapter 6) verified the correctness of the calculations being performed during the cost simulation. However, they did not verify the accuracy of the cost estimates compared against actual deployment costs. An experiment was carried out to measure this accuracy for a test system, which was modelled in ShopForCloud and then deployed on the cloud to obtain its actual deployment cost. The aim of the experiment was to measure the accuracy of the cost estimates, which could have been done using one of the case study systems described in the previous chapter. However, this would have cost thousands of dollars and was not practical. Therefore, a fairly small system was designed to reduce the costs of the experiment but at the same time provide enough scope for the accuracy of the cost estimates to be measured.

The accuracy of the cloud provider's cost calculators were also measured during this experiment and compared to ShopForCloud; these calculators included AWS's Simple Monthly Calculator[69] and Rackspace's cost calculator[70]. Cloudorado[71] and

[69] http://calculator.s3.amazonaws.com/calc5.html

RightScale[72] also provide cost calculators as part of their websites, which were also included in the experiment. Finally, a spreadsheet was created to manually calculate the cost of the test deployment; this spreadsheet is included in Appendix B.

It should be noted that the aim of ShopForCloud is not to provide 100% accurate cost estimates, but rather to provide users with a ballpoint figure that is 'good enough'. This point, which is further discussed in Section 8.3, was mentioned by some of ShopForCloud's users. The tool also creates an understanding about the costs of using public IaaS clouds, and how the costs breakdown between different resources.

8.1.1 Experimental Setup

The experiment involved setting up a test deployment for three months on two clouds, namely the AWS US-West Northern California cloud (used as the primary hosting environment) and the Rackspace UK cloud (used as the backup hosting environment). The test deployment, shown in Figure 42, consists of the servers, storage and database required for a typical web application such as a photo sharing website:

- Users were represented by a remote node. They browsed the site uploading and downloading photos, thus creating data transfer in and out of the cloud. Elasticity patterns were created to represent a doubling of users every month, which in turn was assumed to double the data transferred in/out of the AWS cloud.
- The web server was an On-Demand Linux Micro instance that used the database server and the S3 bucket; however, this data transfer did not need to be modelled as it was within the same cloud meaning that it was free. A pattern was created to represent an extra web server being needed every month to handle the growing number of users. Each web server also used a 10GB EBS volume.

---

[70] http://www.rackspace.co.uk/cloud-hosting/learn-more/calculator/
[71] http://www.cloudorado.com/
[72] http://www.rightscale.com/tco-calculator/

- The photos were stored in an AWS S3 bucket. The size of this bucket was also assumed to double every month due to more users uploading photos (represented by another pattern).
- The database server was an On-Demand Linux Micro instance and it was assumed to store user details and site statistics. The database used a 20GB EBS volume, and it was assumed that monthly database backups would be stored on another cloud for disaster recovery purposes.
- The database backups were stored on Rackspace's CloudFiles storage service (similar to AWS S3). A pattern was created to show that this storage would increase by 2GB each month.
- A redundant web and database server were also created on the Rackspace UK cloud for disaster recover purposes. These servers were 512MB RAM Linux instances (Rackspace does not have 'types' of servers like AWS EC2 does, they are named based on their RAM instead).

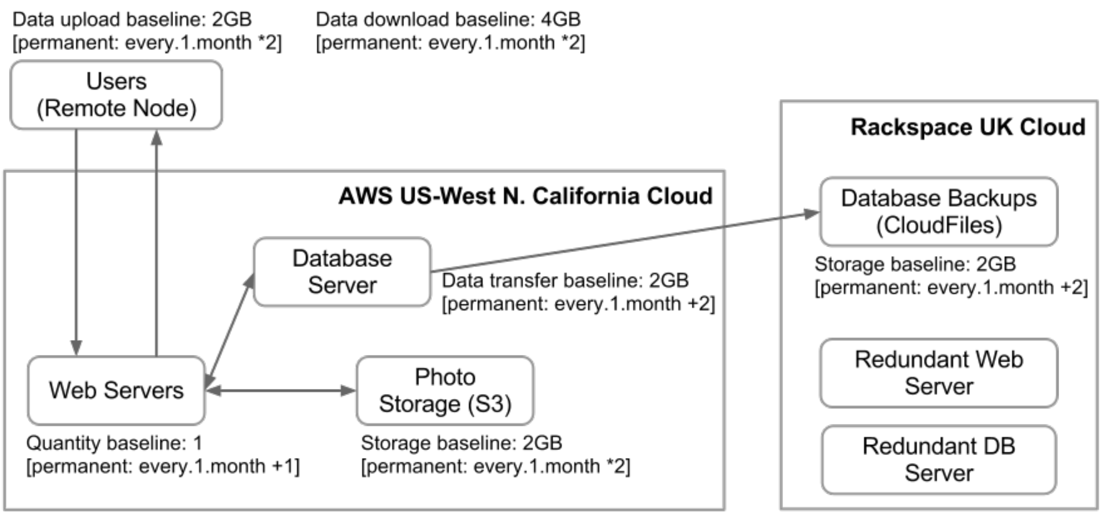


Figure 42: Deployment and patterns created for cost experiment

The deployment was modelled using ShopForCloud before the start of the experiment and the cost estimates were saved. The other tools that were previously mentioned were also used to obtain cost estimates for the deployment. The cloud invoices were collected from AWS and Rackspace after the experiment completion, and used to evaluate the accuracy of the cost estimates produced by ShopForCloud, as described in Section 8.1.2.

The experiment was carried out during February, March and April of 2012 and the following actions were manually performed to mimic the increasing size of the deployment (as modelled by elasticity patterns):

- February
    1. Add 2 servers to AWS
    2. Add 2 servers to Rackspace
    3. Upload 2GB to AWS S3
    4. Download 6GB from AWS S3
    5. Upload 2GB to Rackspace CloudFiles
- March
    1. Add 1 server to AWS
    2. Upload 4GB to AWS S3
    3. Download 12GB from AWS S3
    4. Upload 4GB to Rackspace CloudFiles
- April
    1. Add 1 server to AWS
    2. Upload 8GB to AWS S3
    3. Download 22GB from AWS S3
    4. Upload 6GB to Rackspace CloudFiles

A simple Ruby program[73] was executed to generate 1GB text files that were used during the above actions.

8.1.2 Results

Figure 43 shows a chart of the actual deployment cost after 3 months (dark grey bar) versus the estimated costs produced by various tools (light grey bars). The actual deployment cost was $296.99, while the cost estimate produced by ShopForCloud was $295.26, which represents an accuracy of 99% for the experimental system. The ShopForCloud estimate was similar to the one produced manually using a spreadsheet (described in Appendix B). Users would not necessarily always achieve this level of accuracy for their systems, but the experiment suggests that a high level of accuracy

[73] http://www.skorks.com/2010/03/how-to-quickly-generate-a-large-file-on-the-command-line-with-linux/

can be achieved. The main point being made here is that ShopForCloud can produce estimates that are as accurate as manually calculated estimates, but doing the calculations manually requires expertise of the pricing schemes of different cloud providers, and estimates will be out of date when prices are changed. Using ShopForCloud to forecast the cost of non-trivial deployments with multiple elasticity patterns is also likely to be quicker than manually estimating costs using spreadsheets.

The AWS and Rackspace cost calculators do not support the notion of elasticity. Therefore, they only produce cost estimates for the first month of a deployment, although they could be updated every month and used in a similar manner to the manual spreadsheet. Their cost estimate was $234.48, which represents an accuracy of 79%. This accuracy is very much dependant on the deployment being modelled and is likely to be reduced for more realistic deployments with significant elasticity patterns.

Cloudorado do not support AWS Micro instances and could only provide estimates based on Small instances, which are more expensive than Micro instances. They also do not support the notion of elasticity, and overall they over-estimated by around 50%. The RightScale Total Cost of Ownership calculator is based on RightScale's proprietary data (probably collected from their customers), and also includes their $500/month cloud management fee. This fee was not required for the test deployment and was therefore reduced from their estimate, which resulted in them under estimating the cost by around 235%.

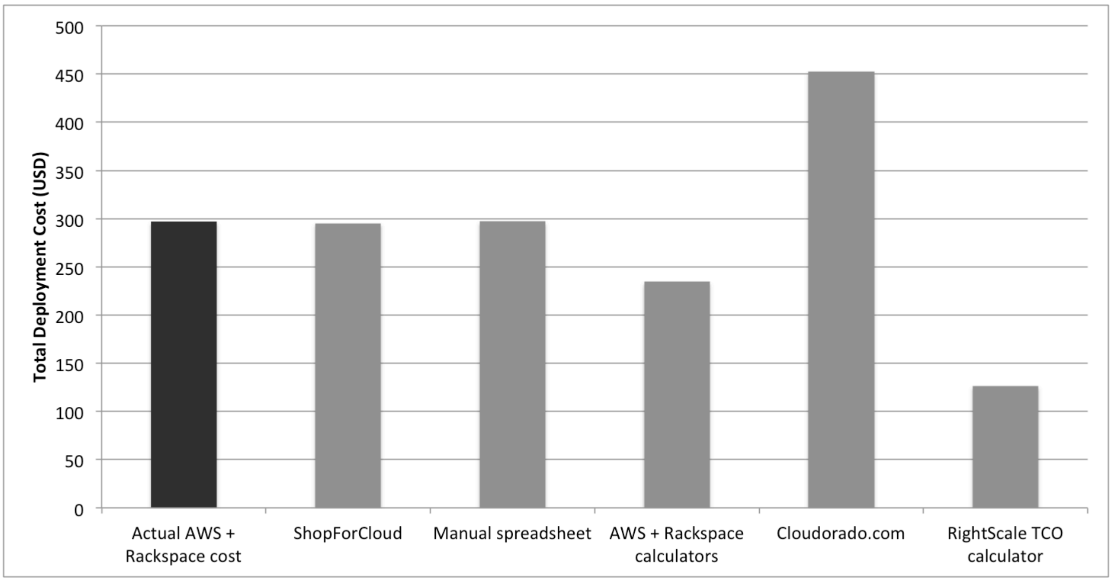


Figure 43: Actual deployment costs versus estimated costs from different tools

Overall, the experiment shows that cost reports produced by ShopForCloud will be accurate assuming that the deployment parameters supplied to the cost simulation represent what actually happens during the system deployment phase. However, as discussed during the case studies, there are scenarios where the error margins in the simulation parameters are high, as some organisations do not know their detailed infrastructure requirements. In such cases, it would be best to create multiple deployments representing extremes in parameter values; this would enable the user to get a range of cost estimates rather than a single estimate.

8.2 User Data

ShopForCloud was run as a free hosted-service between February and May 2012. Users had to create accounts on the system to use it during this period, which enabled us to gather data on how the tool was being used. This section presents some of that data in an anonymised form, and discusses the lessons that were learned from the user data.

8.2.1 Usage Statistics

A total number of 270 users signed-up for ShopForCloud during the four months that it was operational. Users were asked to provide their company name on the sign-up form. These companies included start-ups and SMEs, large corporations, IT consultancies, cloud providers and ISPs as well as universities. There were also a few industry analysts who signed-up. When a new user signs-up, an example deployment is created for them to help them understand the concepts behind ShopForCloud. A Getting Started Guide is also provided (see Appendix A), which guides users through creating the example deployment, creating patterns and additional costs, and getting cost reports.

The following list provides an overview of the user statistics, which do not include the resources created as part of the example deployment (i.e. they reflect what users did on their own). The statistics are discussed in Section 8.2.3 following a look at the various deployment scenarios that were modelled by users.

- 26% of users signed-in more than once, 11% signed-in more than twice and 7% signed-in more than three times.

- 32% of users created one or more deployments.
- 132 deployments were created but only 87 reports were created. This means that some deployments were created with no cost report (i.e. the user did not click on the button to create a corresponding report). In total, around 200 cost simulations were run meaning that some reports were re-generated, which happens when a user changes a deployment and re-generates its report, or the user re-generates the report after cloud prices have changed. The most common cost report duration was 3 years (which is the default), although there were also a few 1 year and 4 year reports.
- 219 servers, 114 storage, 61 databases and 118 data transfer links were created in the above deployments. Section 8.2.2 provides an overview of the scale of deployments that were created by describing several of them.
- 31 additional costs and 104 data transfer links were created. Only 6% of users created new remote nodes (i.e. in addition to the "Users" remote node that is created as part of the example deployment).
- 7% of users changed their default currency from United States Dollars to their local currency, which included British Pounds, Euros, Singapore Dollars and Indian Rupees.

Elasticity patterns were not commonly used; only 27 patterns were created (by 10 users). The patterns included the following names and rules:

- Linear growth: `[permanent: every.1.year on every.2.months +100]`
- Growth: `[temporary: every.1.year on aug +5]`
- Server: `[permanent: every.1.year *2]`
- Scale out (double servers): `[temporary: every.1.year on every.3.months *2]`, `[temporary: every.1.year on every.4.months /2]`
- Pattern for snapshots (10% increase on on-going basis): `[permanent: every.1.year *1.1]`
- 10% more users: `[permanent: every.1.year on every.1.month *1.1]`

This usage of elasticity patterns also reflects the discussions during our case studies, which suggested that the majority of cloud users might not actually need elasticity. Simply knowing that their systems can scale to use more resources *if required* is sufficient for such users. These users are likely to use the cloud not because of elasticity, but because of the ease of deployment and the low up-front costs that are associated with public IaaS clouds.

8.2.2 Usage Scenarios

The case studies presented in Chapter 7 described three scenarios where we used the prototype cost modelling tool to support organisations estimate their costs and compare different deployment options. These case studies involved comparing the costs of using public IaaS clouds versus buying servers, comparing different cloud providers, and comparing different deployment options within a cloud. However, these scenarios were performed by us; whereas the following scenarios were carried out without our involvement or support. As previously mentioned, users only had access to the Getting Started Guide and a small help text on each page of ShopForCloud. The following 7 users were selected as they used ShopForCloud in an interesting way:

User 1: This user's company is a large software development firm in South America. The user has signed-in 18 times and made 8 deployments containing a total of 65 servers, storage and databases with 45 data transfer links between them. The user has used ShopForCloud to obtain cost estimates for their IT systems, and compared the costs of using AWS on-demand versus reserved instances.

User 2: This user's company is an international publishing firm. They have signed-in 4 times and created 4 deployments that also compared the costs of AWS on-demand instances versus reserved instances in addition to creating a deployment to represent a business case (presumably to try and persuade management to use public clouds).

User 3: This user's company is an international telecoms and managed services provider; they have signed-in 3 times and created a deployment to estimate the cost of running basic virtual machines on the cloud. This, as well as the other ISP's who have signed-up, might be an indication that they are considering competing with AWS, and are using the tool to see how their prices compare.

User 4: This user signed-in 4 times, and made 4 deployments comparing the AWS and Rackspace clouds for a deployment containing 200 servers.

User 5: This user's company is a startup that develops iPhone apps. They signed-in once and used the tool to compare the cost of their storage on AWS S3 Reduced Redundancy Storage ($228), Microsoft Azure's Blob Storage ($338) and Rackspace CloudFiles ($360).

User 6: This user's company is a software development company in the Oil & Gas industry. The user signed-in 7 times and made 5 deployments comparing the cost of their deployment on AWS, Microsoft Azure and a mixture of both. They also compared the cost of alternative deployment options such as creating a separate database server for each of their clients.

User 7: This user's company develops industrial control systems. They signed-in once and created two deployments to compare the costs of keeping backups in AWS S3 (4TB). They compared the cost of using the following two types of servers to manage their backups: AWS Medium-Utilization Reserved 1-Year Standard Large and Medium-Utilization Reserved 3-Year Hi-CPU Medium.

8.2.3 Lessons Learned from User Data

The significant number of users that signed-up for ShopForCloud shows that there is a need in industry for cloud cost modelling. The user company data shows that this need might be for different reasons including:

1. Startups, SMEs and IT departments of large corporations using the tool to obtain cost estimates for their IT systems and compare options.
2. Universities using the tool to obtain cost estimates of deploying their research software on public IaaS clouds.
3. Cloud providers and ISPs using the tool to see how they compare against bigger and more popular IaaS cloud providers.

When the prototype cost modelling tool was developed, it was envisaged that "enterprises can assess the feasibility of using cloud computing in their organisations quickly and cheaply without outside consultants [… or] to verify the claims made by

IT consultancies and cloud service providers" [95]. While it is not possible to verify this, the user company data shows that IT consultancies themselves are interested in cost modelling. Thus ShopForCloud could be used as a consultancy tool where consultants become the expert tool users and help organisations investigate their deployment options. This scenario could be useful for organisations that do not have experience or knowledge of cloud computing such as the Dutch IT consultancy that was described in Section 7.4.

The various usage scenarios (described in the previous section) show that organisations are using ShopForCloud to:

1. Obtain cost estimates of using public IaaS clouds.
2. Compare the costs of using different cloud providers.
3. Compare the costs of using different deployment options within a cloud, such as using different types of instances or purchase options (on-demand versus reserved).

These scenarios were also investigated during the case studies described in Chapter 7 using the prototype cost modelling tool. However, whereas the prototype cost modelling tool was probably not going to be used by users due to its usability issues and the difficulties in installing Eclipse (as described in Section 5.3), ShopForCloud made cost modelling simple for users, who were able to conduct similar investigations as the case studies.

The usage statistics show that remote node creation is unnecessary as only 6% of users create them. The most commonly used remote node was "users" (created by default), and the users that did create other remote nodes named them "customers" or "clients". Therefore, this modelling notation can be omitted and the default "users" node could be present during the data transfer link creation. This would simplify the user interface by reducing the number of steps required to model data transfer costs.

The 45 deployments that were created without cost reports show that the report creation process needs to be streamlined. One way to streamline the report creation process would be to have a single web-page, with tables for servers, storage, databases and data transfer links, that can be used to create deployments and see the cost report immediately below the tables. Having a single web-page for deployment creation should be fine as the usage data shows that deployments usually contain less than 10 nodes. This would also make the user interface more responsive as the cost reports can be automatically updated as soon as the deployment is changed.

Cost simulations can take one or two seconds to run, hence, it is feasible to create a single-page deployment with its cost report being updated by AJAX. However, this will increase the number of simulations that are performed, which will place significantly more processing load on the servers. Multiple worker processes can be used to run the simulation jobs if the queue length gets beyond acceptable waiting times.

The use of elasticity patterns was lower than expected. This might have been due to user interface issues as the creation and attachment of patterns is not integrated into the deployment creation process. Pattern creation is done on a separate web-page and users have to jump back and forth between pages to attach a pattern to several deployment attributes. Furthermore, usage statistics showed that the types of patterns that users created were quite simple as they had one or two rules, and either described a peak or gradual growth in usage. Therefore, patterns are probably best attached to the overall deployment rather than its individual attributes.

In this case, users would first select one of two options:

1. *Show my costs if I get popular suddenly* (e.g. a temporary peak), or
2. *Show my costs if I have X% monthly growth*.

If users select the temporary peak option, they would be asked to enter a multiplier of how many times their deployment might grow by during a temporary peak. If they select the monthly growth option, they would be asked to enter the percentage that their deployment is likely to grow by every month. This can be translated to a percentage increase for storage, data transfer etc., or to add up to an increase in the quantity of servers or storage nodes. For example, a 10% monthly increase will increase the quantity of servers by 1 after 8 months as $1.1^8 = 2.14$. Hence, the existing pattern notations and simulation engine would still be used, but a new user interface could be implemented to simplify how elasticity patterns are created and used in ShopForCloud.

8.3 User Feedback

The feedback that was received from the use of the prototype cost modelling tool in the case studies was discussed in Chapter 7. This section focuses on the user feedback that was received from ShopForCloud, either through semi-structured interviews, emails or UserVoice[74]. UserVoice is a web-based feedback system that enables users to provide feedback for a site or ask for support. Ten different ideas were proposed on UserVoice and 36 votes were casted for these ideas during the four months that ShopForCloud was operational.

The top three ideas on UserVoice were:

1. *Make servers/storage/databases cloud agnostic and compare cloud providers for me*. This idea was proposed by a user and received 14 votes. The user wants the tool to automatically compare cloud providers for them based on the specifications of their deployment (e.g. CPU and RAM) and geographical location. This feature could be implemented by making the tool generate multiple deployments from the user's specifications; the cost of the alternative deployments could be charted on the same cost report to help users compare their options.
2. *Which server/storage/database type should I choose?* This idea was proposed by us and received 10 votes. There are currently over 400 types of servers, storage and databases on ShopForCloud and it can be difficult for users to know which type to select if they are not familiar with the various cloud providers. This feature could ask users a few questions to find out about their requirements and suggest a couple of alternatives for them to select.
3. *Add more cloud providers.* This idea was added to UserVoice by us to engage with users and find out what other cloud providers they required. It is interesting to note that several cloud providers, including SoftLayer.com, IBM Cloud, ShiftToTheCloud.com and Zunicore.com, left comments on this thread and asked for their respective clouds to be supported in ShopForCloud.

The other ideas that were suggested on UserVoice included the ability to create overview reports from several cost reports, distinguishing between servers with

[74] https://shopforcloud.uservoice.com

persistent storage (such as Rackspace servers) and ephemeral storage (such as EC2 servers), the ability to share deployments and cost reports with other users (presumably in the same organisation), and having a monthly cost breakdown on the report page that shows how the elasticity patterns are effecting each month's cost. Overall, the feedback provided on UserVoice shows that users are engaged with ShopForCloud and expect it to be further developed to support them in selecting cloud providers and deployment options.

Two of the people involved in the case studies were also asked to use ShopForCloud and provide feedback. The software engineer at the R&D department of the media corporation (Section 7.3) commented on "how much more user friendly" the tool has become and how they could use it in the future. They also mentioned that the server/storage/database type selection needs to be improved (as also mentioned in UserVoice). The person conducting the case study with the Dutch IT consultancy (7.4) mentioned that "most novice users are mostly just looking for the cheapest option to get a server in the cloud. Have them choose a region and minimum specs and suggest the cheapest options" (also mentioned in UserVoice).

A semi-structured interview was conducted with the co-founder of a SaaS company who uses AWS to host their web application. They were interested in using ShopForCloud as they already had spreadsheets that were used to manually estimate their costs and compare different cloud providers (they compared AWS, GoGrid and FlexiScale). Initially it was important for them to understand their deployment costs as they were using a freemium business model, and the paid users had to produce enough profits to cover the costs of the free users. In their case, staff costs far outweighed the deployment costs, and therefore the cost estimates had to be 'good enough' and not 100% accurate.

Going forward, they are interested in comparing their costs between different clouds as more cloud providers become online and competition between them increases. This company found elasticity patterns useful, as cash-flow management is important and they need to know how their monthly costs would change (e.g. ones of their client's was a graduation photography company that had storage peaks during the graduation months in the summer and winter).

A second semi-structured interview was carried out with three technical staff from a software company in the Oil & Gas industry. They have already deployed multiple systems on a cloud; this cloud is not currently supported in ShopForCloud. They estimate their costs manually using spreadsheets, and are interested in using ShopForCloud to compare their current cloud with the others that are supported in ShopForCloud. The automatic comparison between cloud providers that was mentioned in UserVoice also interests them, however, they did mention that it is difficult to move to another cloud once they have deployed their systems on a particular cloud.

Like some of the organisations involved in the case studies, they also do not measure their data transfer between systems, and would find it helpful to model several scenarios to obtain a range of costs for their deployment. Overall, they mentioned that ShopForCloud would "definitely be a useful tool" for them, but they are unsure how often they would use it. They create a number of bid documents for their clients every month, hence they could potentially use the tool to obtain cost estimates for these bid documents.

This chapter has shown that the accuracy of the cost estimates produced by ShopForCloud seems to be good enough for the existing user-base. The ShopForCloud user feedback was positive and was in line with the feedback that was received from the case studies. The usage data provided insights into how the tool is being used and how it could be improved in the future by further simplifying it. The suggestions provided by users showed that they are engaged with this tool and see value in using it.

9 Conclusion

Over the last few years, cloud computing has caused a major change in the way that computing is delivered in enterprises. More and more, computing is shifting away from a product that is owned, to a service that consumed. This shift has caused enterprises to investigate the adoption of cloud computing as public clouds can offer "scalability, reliability, security, ease of deployment, and ease of management for customers, traded off against worries of trust, privacy, availability, performance, ownership, and supplier persistence" [1].

The aim of this project was to support organisations during cloud adoption and assist system deployment decisions in public clouds. Many enterprises are not familiar with the benefits, risks and costs of using public clouds as cloud computing brings a major shift in how IT is provisioned and consumed within organisations. Our work had two broad objectives, which were achieved by this work:

1. Support enterprises in assessing the benefits and risk of using public clouds.

An initial case study was carried out to explore and identify cloud adoption challenges. The following two tools were developed to support decision makers:

- Cloud Suitability Checklist: this tool comprises a simple list of questions to provide a rapid assessment of the suitability of public IaaS clouds for a specific IT system.
- Benefits and Risks Assessment tool: this tool is a spreadsheet that includes the general benefits and risks of using public clouds, which were identified by reviewing over 50 academic papers and industry reports.

Enterprises need to have a discussion about the benefits and risks of using cloud computing to raise awareness, and ensure that decision makers understand the issues involved during cloud adoption from different stakeholder perspectives. The case studies presented in this thesis and the 230 downloads of the Benefits and Risks Assessment tool demonstrate that these tools provide a starting point for those discussions and support enterprises in assessing the benefits and risk of using public clouds.

2. Provide enterprises with a tool to model the costs of deploying systems in public clouds.

As part of their cloud adoption decision making process, enterprises need to be able to estimate the costs of using public clouds. The Elastic Cost Modelling tool that was developed during this project enables enterprises to model the cost of deploying their systems in public clouds including the cost of elasticity, alternative deployment options and cloud providers. The tool also enables enterprises to perform 'what-if' style analysis to understand the effects on their costs if cloud providers change their prices or if they consume more/less resources than their original estimates. The case studies presented in this thesis and the data collected from the 270 users that used the online version of the Elastic Cost Modelling tool, ShopForCloud.com, highlight the simplicity and usefulness of this tool.

Overall, these three tools collectively enable decision makers to investigate the benefits, risks and costs of using public clouds, and effectively support them in making system deployment decisions.

9.1 Lessons Learned from Case Studies

Four case studies were carried out as part of this thesis to explore the problem space and evaluate the tools that are proposed.

The first case study illustrates the potential benefits and risks associated with the migration of an IT system in the oil and gas industry from an in-house datacentre to Amazon EC2 from a broad variety of stakeholder perspectives across the enterprise. Our results show that the system infrastructure in the case study would have cost significantly less on EC2, and using cloud computing could have potentially eliminated many of the support calls for this system. These findings seem significant enough to call for a migration of the system to the cloud. However, the case study revealed that there are significant risks associated with this. Whilst the benefits of using the cloud are attractive, it is argued that decision-makers should consider the overall organisational implications of the changes brought about with cloud computing to avoid implementing local optimisations at the cost of organisation-wide performance.

The second (academic digital library and search engine) and third (media corporation) case studies show that there can be multiple perspectives on the benefits/risks of cloud migration, and even within a single enterprise, different divisions have different views. Motivations and concerns of different stakeholders have to be considered during cloud adoption decisions. Overall these three case studies show that the cloud suitability checklist and the benefits and risks assessment tool collectively provide organisations with a starting point for risk assessment as they inform people about the various benefits and risks. Although the tools do not provide a framework for a yes/no decision to be made, they do trigger discussions within organisations and provide a low-effort means to gather the perspectives of different stakeholders in an organisation. This can be seen as the most beneficial aspect of using these tools.

The fourth case study focused on cost modelling. This case study, as well as the other case studies, investigated the cost of deploying systems on public clouds; the studies show that:

1. The cost effectiveness of using public clouds is situation dependent rather than universally less expensive than traditional forms of IT provisioning. Running systems on the cloud using a traditional 'always on' approach can be less cost effective, and the elastic nature of the cloud has to be used to reduce costs. Decision makers have to model the variations in resource usage and their systems' deployment options to obtain accurate cost estimates.
2. Using public clouds can be financially attractive if their elasticity is utilised. However, this usually requires enterprise applications to be re-architected as older applications were either not designed as multi-tenant systems, or they cannot scale-out to use multiple servers as they rely on the server to be scaled-up to provide more resources (e.g. CPU and RAM).
3. Performing upfront cost modelling is beneficial as there can be significant cost differences between different cloud providers, and different deployment options within a single cloud.

During such modelling exercises, the variations in a system's load (over time) must be taken into account to produce more accurate cost estimates, and the notion of elasticity patterns that is presented in this thesis provides one simple way to do this.

9.2 Critical Evaluation of the Tools

The Cloud Suitability Checklist and the Benefits and Risks Assessment Tool are useful in starting discussions around cloud adoption in enterprises. Although these tools can be used to support decision making, they could have been further developed to include more formalised processes such as issue mapping techniques to guide discussions. This would have enabled various stakeholders to use the tools to map out their arguments for/against cloud adoption. However, the addition of extra processes can also have negative effects as they add complexity to the tool and so require additional learning overhead. Low-effort approaches are more likely to be used by practitioners in industry. The tools could also have been further developed to automatically generate the spider charts mentioned in Chapter 4 based on the data input by the user into the Benefits and Risks Assessment tool.

ShopForCloud, the elastic cost modelling tool, provides useful features to estimate and forecast the costs of using public clouds. Such forecasts could also be made manually, using spreadsheets for example, but only by users who have the expertise of how the various cloud provide pricing schemes work. A very basic version of elasticity patterns could also be included during such forecasts by, say adding a percentage growth every month. Overall this manual approach can be quite time consuming, especially for larger systems, and the estimates become out-dated as cloud providers change their prices. However, the manual approach does enable the user to completely customise the process and include cloud services and prices that are not supported by ShopForCloud. Examples of this limitation are the cost of data transfer between availability zones in an AWS cloud, or the cost of using 'NoSQL' databases such as AWS DynamoDB or Google BigQuery.

There are also other limitations of ShopForCloud, which arise out of the assumptions that were made during its development; for example the price of storage is stored on a per-month basis, meaning that users cannot estimate the cost of using storage for a few hours or days per month. Another limitation is the lack of support for attaching elasticity patterns to AWS reserved instances due to differences between how the cost simulation algorithm calculates upfront reservation costs, and how AWS treats reserved instances as a 'pool of resources' that are used whenever possible (e.g. if a user purchases 10 reserved instance but launches 15 on-demand instance, the reserved prices are applied for 10 and 5 are priced at on-demand rates). These limitations can be addressed in future work by developing customised costing algorithms for cloud providers that have specific requirements.

The usage data collected from ShopForCloud highlighted several areas in the user-interface that need improvement. The process of creating a deployment and generating its cost report requires too many clicks, during which some users drop-off and leave the tool (shown by the 45 deployments that were created without cost reports). These improvements, described in Section 8.2, stress the importance of usability and simplicity of such simulation tools if they are to be used by practitioners in industry. ShopForCloud could be improved by enabling users to estimate public cloud costs using a more 'cause-and-effect' interface, where a user changes the deployment and instantly sees the updated cost report. This would be feasible as the cost simulation takes around a second for a deployment with ten nodes, and some optimisations have already been applied to reduce the number of database queries being made by the simulation engine.

ShopForCloud currently includes prices from three cloud providers, namely AWS, Microsoft Azure and Rackspace. These cloud providers alone have over 2,000 prices, which were collected using scrapers in ShopForCloud. However, this approach is not scalable as there are many cloud providers worldwide. An API that enables cloud providers to push their prices into the ShopForCloud database would be better as they can change their prices at anytime and without the scraper maintenance overheard for us.

On a broader note, organisations also need to consider the indirect costs of using different public cloud services. For example, using Platform-as-a-Service offerings, such as Heroku, save time but usually cost more than do-it-yourself solutions that are based on Infrastructure-as-a-Service offerings such as AWS EC2. ShopForCloud does not currently attempt to take into account these cost trade-offs, nor the cost of software changes that may be required to gain maximum benefit from using cloud services. Therefore, ShopForCloud can provide cost estimates of deploying systems on public clouds but its outputs should be used as part of a broader investigation by organisations that are considering cloud adoption.

9.3 Threats to Validity

The lessons learned from the case studies point to general trends in cloud adoption but are limited by the usual drawbacks of case studies in that it can be difficult to generalise from individual cases. Furthermore, none of the companies involved in the case studies went onto deploy their systems on public clouds, which meant that it was not possible to conduct follow-up studies to see the effects that the use of our tools and techniques had on their actual decision making process.

The experiment that was conducted to measure the accuracy of the cost estimates had a shortcoming, as there were no actual applications running on the servers. This meant that disks and network activity were not exercised in a realistic manner. Since some cloud providers charge for disk read/write requests or database transactions, the actual system costs could have thus been slightly different from what was actually obtained in the experiments. However, as previously mentioned, the aim of our work was to provide ‘good enough’ estimates for decision making purposes and not 100% accurate forecasts.

ShopForCloud had a relatively large number of users but only a few were actually interviewed and valuable insights might have been missed. A random selection of users could have been surveyed to gain further feedback and ask for their insights. However, it is not clear if the survey data would have added anything significant on top of the findings from the case studies, user feedback via the UserVoice portal, user data analysis and the semi-structured interviews.

9.4 Future Work

The feedback gathered from the tools pointed to a number of areas that could be improved in the future, including auto-generating the spider charts from the Benefits and Risks Assessment tool and improving the user-interface in ShopForCloud. In addition to these changes, the following three areas could be further developed in future work:

System deployment models (for infrastructure) could perhaps be automatically created from data collected through system monitoring tools. This might be useful for large-scale systems due to the time saved from creating such models manually.

However, such models would not include details of which clouds or services would be suitable for a system, thus such information would have to be supplied by decision makers. A more suitable scenario for automation might be for existing cloud users to “import” their system deployment details into the cost modelling tool to investigate the cost of other system deployment options.

A recommendation engine could be developed in ShopForCloud to recommend deployment options to users based on constraints set by them. For example, a user could receive recommendations for clouds and server types based on their basic infrastructure requirements such as the hardware specifications, performance requirements and location-based constraints for clouds (e.g. only EU clouds).

An automatic optimisation system could be developed on top of the ShopForCloud simulation engine to run several simulations and find the cheapest deployment option for a given deployment. This system could be linked to the recommendation engine to enable users to describe their requirements and see their deployment options ranked in terms of cost estimates.

Currently users have to create elasticity patterns manually in ShopForCloud. This is often based on previous data or experience, and it might be useful for users to have a tool that analyses their application logs or usage data to extract elasticity patterns. The research group that has already extended elasticity patterns to support probabilistic rules [100], is currently working on this idea. Enterprises often know patterns that occur in their business domain (e.g. number of customers increase in a certain season), and these patterns are likely to influence the load on their IT systems. Therefore it is not clear how useful such an automatic pattern extraction feature would be, given that the aim is to get ‘good enough’ cost estimates.

Looking back on the research carried out as part of this thesis, I should have probably created all of the tools as web applications and launched them publicly from the outset. This would have enabled me to gather feedback and test the usefulness of these tools quicker. It might also have been interesting to investigate different auto-scaling algorithms and tools for web applications, and evaluate their cost effectiveness.

References

[1] H. Erdogmus, “Cloud Computing: Does Nirvana Hide behind the Nebula?,” *IEEE Software*, vol. 26, no. 2, pp. 4–6, 2009.

[2] M. Creeger, “CTO roundtable: cloud computing,” *Communications of the ACM*, vol. 52, no. 8, pp. 50–56, 2009.

[3] J. Bar, “Amazon EC2 Beta,” 2006. [Online]. Available: http://aws.typepad.com/aws/2006/08/amazon_ec2_beta.html. [Accessed: 13-Feb-2012].

[4] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “Above the Clouds: A Berkeley View of Cloud Computing.,” *University of California at Berkeley*, 2009.

[5] L. Vaquero, L. Merino, and J. Caceres, “A break in the clouds: towards a cloud definition,” *SIGCOMM Computer Communications Review*, vol. 39, no. 1, pp. 50–55, 2009.

[6] L. Youseff, M. Butrico, and D. Da Silva, “Toward a Unified Ontology of Cloud Computing,” in *Grid Computing Environments Workshop (GCE ’08)*, 2008, pp. 1–10.

[7] P. Mell and T. Grance, “The NIST Definition of Cloud Computing,” *National Institute of Standards and Technology*, 2009.

[8] R. Yanosky, “From Users to Choosers: The Cloud and the Changing Shape of Enterprise Authority,” in *The Tower and the Cloud*, R. Katz, Ed. EDUCAUSE, 2008, pp. 126–136.

[9] J. Erbes, H. R. Motahari-nezhad, and S. Graupner, “The Future of Enterprise IT in the Cloud,” *Computer*, vol. 45, no. 5, pp. 66–72, May 2012.

[10] W. Fellows, “Partly Cloudy, Blue-Sky Thinking About Cloud Computing,” *451 Group*, 2008.

[11] H. R. Motahari Nezhad, B. Stephenson, and S. Singhal, “Outsourcing Business to Cloud Computing Services: Opportunities and Challenges,” *HP Laboratories*, 2009.

[12] J. Elson and J. Howell, “Refactoring Human Roles Solves Systems Problems,” in *Workshop on Hot Topics in Cloud Computing (HotCloud)*, 2009.

[13] A. Khajeh-hosseini, D. Greenwood, and I. Sommerville, “Cloud Migration: A Case Study of Migrating an Enterprise IT System to IaaS,” in *IEEE 3rd Int. Conf. on Cloud Computing*, 2010, pp. 450–457.

[14] B. Grobauer, T. Walloschek, and E. Stocker, “Understanding Cloud-Computing Vulnerabilities,” *IEEE Security & Privacy Magazine*, 2010.

[15] M. Jensen, J. O. Schwenk, N. Gruschka, and L. L. Iacono, “On Technical Security Issues in Cloud Computing,” in *IEEE International Conference on Cloud Computing (CLOUD-II 2009)*, 2009, pp. 109–116.

[16] N. Jesper, “On our extended downtime, Amazon and what’s coming,” 2009. [Online]. Available: http://blog.bitbucket.org/2009/10/04/on-our-extended-downtime-amazon-and-whats-coming/. [Accessed: 01-Dec-2009].

[17] Cloud Security Alliance, “Security Guidance for Critical Areas of Focus in Cloud Computing,” *Cloud Security Alliance*, 2009.

[18] D. Catteddu and G. Hogben, “Cloud Computing: benefits, risks and recommendations for information security,” *European Network and Information Security Agency (ENISA)*, 2009.

[19] J. Viega, “Cloud Computing and the Common Man,” *Computer*, vol. 42, no. 8, pp. 106–108, 2009.

[20] L. M. Kaufman, “Data Security in the World of Cloud Computing,” *IEEE Security & Privacy*, vol. 7, no. 4, pp. 61–64, 2009.

[21] S. Mansfield-devine, “Danger in the clouds,” *Network Security*, vol. 12, 2008.

[22] A. Joint, E. Baker, and E. Eccles, “Hey, you, get off of that cloud?,” *Computer Law & Security Review*, vol. 25, no. 3, 2009.

[23] P. Jaeger, J. Lin, J. Grimes, and S. Simmons, “Where is the cloud? Geography, economics, environment, and jurisdiction in cloud computing,” *First Monday*, vol. 14, no. 5, 2009.

[24] M. Nelson, “The Cloud, the Crowd, and Public Policy,” *Issues In Science And Technology*, vol. 25, no. 4, 2009.

[25] GSA OCIO (Office of The Chief Information Officer), “Infrastructure as a Service (IaaS) offerings,” 2009. [Online]. Available: https://www.fbo.gov/index?s=opportunity&mode=form&id=d208ac8b8687dd9c6921d2633603aedb&tab=core&_cview=0&cck=1&au=&ck. [Accessed: 01-Dec-2009].

[26] P. Saripalli and B. Walters, “QUIRC: A Quantitative Impact and Risk Assessment Framework for Cloud Security,” in *2010 IEEE 3rd International Conference on Cloud Computing*, 2010, pp. 280–288.

[27] B. Grobauer and T. Schreck, “Towards incident handling in the cloud,” in *Proceedings of the 2010 ACM workshop on Cloud computing security workshop (CCSW ’10)*, 2010, pp. 77–85.

[28] P. Saripalli and G. Pingali, “MADMAC: Multiple Attribute Decision Methodology for Adoption of Clouds,” in *2011 IEEE 4th International Conference on Cloud Computing*, 2011, pp. 316–323.

[29] K. P. Yoon and C. Hwang, *Multiple Attribute Decision Making: An Introduction*. London: Sage Publications, 1995.

[30] S. C. Misra and A. Mondal, “Identification of a company’s suitability for the adoption of cloud computing and modelling its corresponding Return on Investment,” *Mathematical and Computer Modelling*, Mar. 2010.

[31] M. Menzel and R. Ranjan, “CloudGenius: decision support for web server cloud migration,” in *Proceedings of the 21st international conference on World Wide Web - WWW ’12*, 2012.

[32] M. Menzel, M. Schönherr, and S. Tai, “(MC2)2: criteria, requirements and a software prototype for Cloud infrastructure decisions,” *Software: Practice and Experience*, Aug. 2011.

[33] Accenture Technology Labs, “Accenture Cloud Computing Accelerator,” *Accenture*, 2009.

[34] Computer Sciences Corporation, “Cloud Adoption Assessment,” *Computer Sciences Corporation*, 2010.

[35] R. Buyya, H. Stockinger, J. Giddy, and D. Abramson, “Economic Models for Management of Resources in Grid Computing,” in *SPIE International Symposium on The Convergence of Information Technologies and Communications (ITCom 2001)*, 2001.

[36] S. Clearwater and B. A. Huberman, “Swing Options: a Mechanism for Pricing IT Peak Demand,” in *11th International Conference on Computing in Economics*, 2005.

[37] G. A. Thanos, C. Courcoubetis, and G. D. Stamoulis, “Adopting the Grid for Business Purposes: The Main Objectives and the Associated Economic Issues,” in *Grid Economics and Business Models: 4th International Workshop (GECON)*, 2007.

[38] F. Wu, L. Zhang, and B. A. Huberman, “Truth-telling Reservations,” *Algorithmica*, vol. 52, no. 1, pp. 65–79, 2008.

[39] J. O. Fito, M. Macias, and J. Guitart, “Toward business-driven risk management for Cloud computing,” in *2010 International Conference on Network and Service Management*, 2010, pp. 238–241.

[40] M. Klems, J. Nimis, and S. Tai, “Do Clouds Compute? A Framework for Estimating the Value of Cloud Computing,” *Markets, Services, and Networks, Lecture Notes in Business Information Processing*, vol. 22, 2009.

[41] E. Walker, “The Real Cost of a CPU Hour,” *Computer*, vol. 42, no. 4, pp. 35–41, Apr. 2009.

[42] E. Walker, W. Brisken, and J. Romney, “To Lease or Not to Lease from Storage Clouds,” *Computer*, vol. 43, no. 4, pp. 44–50, Apr. 2010.

[43] M. Assuncao, A. Costanzo, and R. Buyya, “Evaluating the cost-benefit of using cloud computing to extend the capacity of clusters.,” in *18th ACM Int. Symp. on High Performance Distributed Computing (HPDC ’09)*, 2009, pp. 141–150.

[44] E. Deelman, G. Singh, M. Livny, B. Berriman, and J. Good, “The cost of doing science on the cloud: the Montage example,” in *2008 ACM/IEEE conference on Supercomputing (SC’08)*, 2008, pp. 1–12.

[45] D. Kondo, B. Javadi, P. Malecot, F. Cappello, and D. P. Anderson, “Cost-benefit analysis of Cloud Computing versus desktop grids,” in *2009 IEEE international Symposium on Parallel & Distributed Processing*, 2009, pp. 1–12.

[46] G. Singer, I. Livenson, M. Dumas, and S. N. Srirama, “Towards a Model for Cloud Computing Cost Estimation with Reserved Instances,” in *Proc. of 2nd Int. ICST Conf. on Cloud Computing, CloudComp 2010*, 2010.

[47] B. Suleiman, S. Sakr, R. Jeffery, and A. Liu, “On understanding the economics and elasticity challenges of deploying business applications on public cloud infrastructure,” *Journal of Internet Services and Applications*, vol. 2, no. 3, pp. 1–21, Dec. 2011.

[48] V. T. K. Tran, K. Lee, A. Fekete, A. Liu, and J. Keung, “Size Estimation of Cloud Migration Projects with Cloud Migration Point (CMP),” in *2011 International Symposium on Empirical Software Engineering and Measurement*, 2011, pp. 265–274.

[49] A. J. Albrecht and J. E. Gaffney, “Software Function, Source Lines of Code, and Development Effort Prediction: A Software Science Validation,” *IEEE Transactions on Software Engineering*, vol. SE-9, no. 6, pp. 639–648, Nov. 1983.

[50] L. Gillam, B. Li, J. O’Loughlin, and A. P. S. Tomar, “Fair Benchmarking for Cloud Computing Systems,” *University of Surrey*, 2012.

[51] Z. Ou, H. Zhuang, J. K. Nurminen, A. Ylä-Jääski, and P. Hui, "Exploiting Hardware Heterogeneity within the Same Instance Type of Amazon EC2," in *HotCloud'12 Proceedings of the 4th USENIX conference on Hot Topics in Cloud Ccomputing*, 2012.

[52] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. D. Rose, and R. Buyya, "CloudSim: A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms," *To appear in Software: Practice and Experience*, 2010.

[53] A. Li, X. Yang, S. Kandula, and M. Zhang, "CloudCmp: shopping for a cloud made easy," in *Proceedings of the 2nd USENIX conference on Hot topics in cloud computing*, 2010.

[54] G. Booch, J. Rumbaugh, and I. Jacobson, *Unified Modeling Language User Guide*, 2nd Editio. Addison-Wesley Professional, 2005.

[55] SysML Partners, "OMG SysML 1.2 Specifications," 2010. [Online]. Available: http://www.sysml.org/docs/specs/OMGSysML-v1.2-10-06-02.pdf. [Accessed: 05-Jun-2012].

[56] P. Goldsack, J. Guijarro, S. Loughran, A. Coles, A. Farrell, A. Lain, P. Murray, and P. Toft, "The SmartFrog configuration management framework," *ACM SIGOPS Operating Systems Review*, vol. 43, no. 1, pp. 16–25, Jan. 2009.

[57] C. Cassell and G. Symon, *Essential Guide To Qualitative Methods In Organizational Research*. Sage Publications Ltd, 2004, pp. 323–324.

[58] T. Wengraf, *Qualitative Research Interviewing: Biographic Narrative and Semi-Structured Methods*. Sage Publications Ltd, 2001, pp. 5–6.

[59] R. K. Yin, *Case Study Research. Design and Methods*. Sage Publications Ltd, 1994.

[60] M. A. Vouk, "Cloud computing issues, research and implementations," in *30th International Conference on Information Technology Interfaces (ITI 2008)*, 2008, pp. 31–40.

[61] T. Sullivan, "Nick Carr: The ways cloud computing will disrupt IT," 2009. [Online]. Available: http://www.cio.com.au/article/296892/nick_carr_ways_cloud_computing_will_disrupt_it. [Accessed: 10-Jan-2010].

[62] T. Ristenpart, E. Tromer, H. Shacham, and S. Savage, "Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds," in *16th ACM Conference on Computer and Communications Security*, 2009, pp. 199–212.

[63] S. Pearson, "Taking account of privacy when designing cloud computing services," in *ICSE Workshop on Software Engineering Challenges of Cloud Computing*, 2009, pp. 44–52.

[64] M. Mowbray, "The Fog over the Grimpen Mire: Cloud Computing and the Law," *SCRIPTed Journal of Law, Technology and Society*, vol. 6, no. 1, pp. 132–146, Apr. 2009.

[65] R. Whiteside, "Taylor Woodrow Migrates 1,800 Users to Google Apps," 2008. [Online]. Available: http://googleenterprise.blogspot.com/2008/07/taylor-woodrow-migrates-1800-users-to.html. [Accessed: 10-Jan-2010].

[66] M. L. Markus, "Power, Politics, and MIS Implementation," *Communications of the ACM*, vol. 26, no. 6, pp. 430–444, 1983.

[67] P. G. W. Keen, "Information systems and organizational change," *Communications of the ACM*, vol. 24, no. 1, 1981.

[68] O. Hanseth and K. Braa, "Hunting for the Treasure at the End of the Rainbow: Standardizing corporate IT Infrastructure," *Computer Supported Cooperative Work (CSCW)*, vol. 10, no. 3, pp. 261–292, Sep. 2001.

[69] L. Lapointe and S. Rivard, "A multilevel model of resistance to information technology implementation," *MIS Quarterly*, vol. 29, no. 3, pp. 461–491, 2005.

[70] J. Rooksby and A. Khajeh-Hosseini, "Diagnostic work in cloud computing," in *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12*, 2012, pp. 335–338.

[71] I. Sommerville, *Software Engineering*, 9th Editio. Addison Wiley, 2010.

[72] L. Loh and N. Venkatraman, "An Empirical Study of Information Technology Outsourcing: Benefits, Risks, and Performance Implications," in *Proc. ICIS 1995, Paper 25*, 1995.

[73] M. C. Lacity and R. Hirschheim, "The Information Systems Outsourcing Bandwagon," *Sloan Management Review*, vol. 35, no. 1, pp. 73–86, 1993.

[74] J. Dibbern, T. Goles, R. Hirschheim, and B. Jayatilaka, "Information systems outsourcing: a survey and analysis of the literature," *ACM SIGMIS Database*, vol. 35, no. 4, pp. 6–102, Nov. 2004.

[75] S. Overby, "Outsourcing: Bringing IT Back Home," *CIO*, 2003.

[76] J. Varia, "Architecting for the Cloud: Best Practices," *Amazon Web Services*, 2010.

[77] L. Herbert and J. Erickson, "The ROI Of Software-As-A-Service," *Forrester Research*, 2009.

[78] J. Jurison, "The role of risk and return in information technology outsourcing decisions," *Journal of Information Technology*, vol. 10, no. 4, pp. 239–247, Dec. 1995.

[79] P. Palvia, "A dialectic view of information systems outsourcing: Pros and cons," *Information & Management*, vol. 29, no. 5, pp. 265–275, Nov. 1995.

[80] A. Khajeh-hosseini, I. Sommerville, and I. Sriram, "Research Challenges for Enterprise Cloud Computing," *LSCITS Technical Report*, 2010.

[81] Sun Microsystems, "Introduction to Cloud Computing architecture," *Sun Microsystems*, 2009.

[82] K. Jeffery and B. Neidecker-Lutz, "The Future of Cloud Computing: Opportunities for European Cloud Computing beyond 2010," *European Commision*, 2010.

[83] J. F. Rayport and A. Heyward, "Envisioning the Cloud: The Next Computing Paradigm," *Marketspace*, 2009.

[84] D. Chappell, "Windows Azure and ISVs: A Guide for Decision Makers," *David Chappell & Associates*, 2009.

[85] P. Mell and T. Grance, "Effectively and Securely Using the Cloud Computing Paradigm," *National Institute of Standards and Technology*, 2009.

[86] D. McDonald, C. Breslin, and A. MacDonald, "Final report from the JISC Review of the Environmental and Organisational Implications of Cloud Computing in Higher and Further Education," *University of Strathclyde and JISC*, 2010.

[87] T. Wood, E. Cecchet, K. K. Ramakrishnan, P. Shenoy, J. V. D. Merwe, and A. Venkataramani, "Disaster Recovery as a Cloud Service: Economic Benefits & Deployment Challenges," in *Proceedings of the 2nd USENIX conference on Hot topics in cloud computing (HotCloud'10)*, 2010.

[88] K. Sripanidkulchai, S. Sahu, Y. Ruan, A. Shaikh, and C. Dorai, "Are clouds ready for large distributed applications?," in *ACM SIGOPS Operating Systems Review*, 2010, vol. 44, no. 2, pp. 18–23.

[89] R. Clarke, "Computing Clouds on the Horizon? Benefits and Risks from the User's Perspective," in *23rd Bled eConference*, 2010.

[90] D. Durkee, "Why cloud computing will never be free," *Communications of the ACM*, vol. 53, no. 5, p. 62, May 2010.

[91] M. C. Lacity, L. P. Willcocks, and D. F. Feeny, "The Value of Selective IT Sourcing," *Sloan Management Review*, pp. 13–25, 1996.

[92] B. A. Aubert, M. Patry, and S. Rivard, "A framework for information technology outsourcing risk management," *ACM SIGMIS Database*, vol. 36, no. 4, pp. 9–28, Oct. 2005.

[93] M. Earl, "The risks of outsourcing IT," *Sloan Management Review*, pp. 26–32, 1996.

[94] J. Rhoton, *Cloud Computing Explained: Implementation Handbook for Enterprises*, 2nd Editio. Milton Keynes, UK: Recursive Press, 2009.

[95] A. Khajeh-Hosseini, D. Greenwood, J. W. Smith, and I. Sommerville, "The Cloud Adoption Toolkit: supporting cloud adoption decisions in the enterprise," *Software: Practice and Experience*, vol. 42, no. 4, pp. 447–465, Apr. 2012.

[96] Amazon Web Services, "Security Best Practices," *Amazon Web Services*, 2010.

[97] M. Dalheimer and F. Pfreundt, "GenLM: License Management for Grid and Cloud Computing Environments," in *2009 9th IEEE/ACM international Symposium on Cluster Computing and the Grid (CCGRID)*, 2009, pp. 132–139.

[98] S. Bradshaw, C. Millard, and I. Walden, "Contracts for Clouds: Comparison and Analysis of the Terms and Conditions of Cloud Computing Services," *Queen Mary School of Law Legal Studies*, 2010.

[99] R. Harms and M. Yamartino, "The Economics of the Cloud," *Microsoft*, 2010.

[100] K. Johnson, S. Reed, and R. Calinescu, "Specification and Quantitative Analysis of Probabilistic Cloud Deployment Patterns," in *7th Haifa Verification Conference (HVC 2011)*, 2011.

[101] B. Rongen, "Making the case for migration of information systems to the cloud," in *16th Twente Student Conference on IT*, 2012.

[102] M. Glautier and B. Underdown, *Accounting: Theory and Practice*, 7th Editio. Financial Times/Prentice Hall, 2001.

[103] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," *Communications of the ACM*, vol. 53, no. 4, 2010.

[104] A. Khajeh-Hosseini, I. Sommerville, J. Bogaerts, and P. Teregowda, "Decision Support Tools for Cloud Migration in the Enterprise," in *IEEE 4th Int. Conf. on Cloud Computing*, 2011, pp. 541–548.

Appendix A – ShopForCloud Getting Started Guide

The cloud cost modelling tool has a 'Getting Started Guide' that is available online at: https://sites.google.com/site/alikhajeh1/sfc_guide.pdf. This guide introduces the basic concepts of the tool, describes how a deployment can be created and how its cost report can be produced.

Appendix B – Cost experiment manual calculations

The following spreadsheet was created to manually calculate the costs of the test deployment used during the cost experiment, which was described in Section 8.1. The spreadsheet shows the monthly cost of servers, storage and data transfer, which were adjusted accordingly each month based on the predicated usages (i.e. the elasticity patterns were manually processed). Manual calculations can be time consuming to do for large deployments or for deployments with many patterns.

Using AWS US-West Northern California and Rackspace UK clouds						
Month 1	Servers	Storage	Data Transfer	Total		
AWS (USD)	24	3.36	0.72	28.08		
RAX (GBP)	19.2	0.22	0	19.42	30.76	USD, Used Google Currency Converter
			Monthly total in USD	58.84		1 GBP = 1.5840 USD
Month 2	Servers	Storage	Data Transfer	Total		
AWS (USD)	55.8	4.52	1.44	61.76		
RAX (GBP)	29.76	0.44	0	30.2	47.84	USD
			Monthly total in USD	109.6		
Month 3	Servers	Storage	Data Transfer	Total		
AWS (USD)	73	5.74	2.64	81.38		
RAX (GBP)	29.2	0.66	0	29.86	47.3	USD
			Monthly total in USD	128.68		
			Total Deployment Cost (sum of 3 months)	**297.12**	**USD**	

Appendix C - Publications

This thesis includes material from the following publications.

Khajeh-Hosseini A, Greenwood D, Smith J W, Sommerville I. 2012. *The Cloud Adoption Toolkit: Supporting Cloud Adoption Decisions in the Enterprise*. Software: Practice and Experience. DOI: 10.1002/spe.1072

Khajeh-Hosseini A, Sommerville I, Bogaerts J, Teregowda P. 2011. *Decision Support Tools for Cloud Migration in the Enterprise*. IEEE 4th Int. Conf. on Cloud Computing (CLOUD 2011), Washington DC, USA. DOI: 10.1109/CLOUD.2011.59

Khajeh-Hosseini A, Greenwood D, Sommerville I. 2010. *Cloud Migration: A Case Study of Migrating an Enterprise IT System to IaaS*. IEEE 3rd Int. Conf. on Cloud Computing (CLOUD 2010). Miami, USA. DOI: 10.1109/CLOUD.2010.37

Khajeh-Hosseini A, Sommerville I, Sriram I. 2010. *Research Challenges for Enterprise Cloud Computing*. LSCITS Technical Report.

Sriram I, Khajeh-Hosseini A. 2010. *Research Agenda in Cloud Technologies*. LSCITS Technical Report.

Smith J W, Khajeh-Hosseini A, Ward J S, Sommerville I. 2012. *CloudMonitor: Profiling Power Usage*. IEEE 5th Int. Conf. on Cloud Computing (CLOUD 2012), Hawaii, USA.

Rooksby J, Khajeh-Hosseini A. 2012. *Diagnostic Work in Cloud Computing*. ACM Conf. on Computer Supported Cooperative Work (CSCW 2012). Seattle, USA. DOI: 10.1145/2145204.2145257