Proceedings of the 11th International Conference on Manufacturing Research (ICMR2013), Cranfield University, UK, 19th – 20th September 2013, pp 195-200

COST MODELLING FOR CLOUD COMPUTING UTILISATION IN LONG-TERM DIGITAL PRESERVATION

Essam Shehab
Isaac Sanya
Mohamed Badawy
Fatma Ocal

Manufacturing and Materials Department
Cranfield University
Cranfield
Bedford, MK43 0AL
{e.shehab, i.o.sanya, m.badawy, f.ocal}@cranfield.ac.uk

Bartosz Namiesnik
Julien Morineau
Stephen Odika
Zaira Fernandez Ortiz

Manufacturing and Materials Department
Cranfield University
Cranfield
Bedford, MK43 0AL
{b.namiesnik, j.morineau, z.fernandezortiz, s.o.odika}@cranfield.ac.uk

ABSTRACT

The rapid increase in volume of digital information can cause concern among organisations regarding manageability, costs and security of their information in the long-term. As cloud computing technology is often used for digital preservation purposes and is still evolving, there is difficulty in determining its long-term costs. This paper presents the development of a generic cost model for public and private clouds utilisation in long-term digital preservation (LTDP), considering the impact of uncertainties and obsolescence issues. The cost model consists of rules and assumptions and was built using a combination of activity-based and parametric cost estimation techniques. After generation of cost breakdown structures for both clouds, uncertainties and obsolescence were categorised. To quantify impacts of uncertainties on cost, three-point estimate technique was employed and Monte Carlo simulation was applied to generate the probability distribution on each cost driver. A decision-support cost estimation tool with dashboard representation of results was developed.

Keywords: Cloud Computing, Cost Modelling, Uncertainties, Long-Term Digital Preservation.

1 INTRODUCTION

In current information society, the volume of digital information produced everyday has been increasing rapidly. As information is often required to be kept over a long period, there is a growing need of managing it to ensure its accessibility, sustainability and security over time. Cloud computing has emerged to organisations with a number of business benefits, e.g. cost saving, higher performance and increased availability. As the concept of cloud is relatively new, it is difficult to determine long-term costs of using this technology.

This paper presents the development of a cost model for cloud computing in utilisation of LTDP by including the impact of uncertainties and obsolescence issues on cost. The research study is part of the ENSURE project, a Seventh Framework Programme (FP7) project focused on providing a total solution for LTDP. The next section of the paper contains the literature review on areas related to the scope of the research. Section 3 describes an overview of the research methodology employed. Section 4 describes the main results of the research study: cost breakdown structure, uncertainties categorisation and costing and interactive cost modelling tool. In Section 5, the techniques that have been used to validate the research results are discussed.

2 RELATED RESEARCH

LTDP brings attention to storage and maintenance of digital information with the intention of accessing the content in the future regardless of any obsolescence in format, software and hardware (Factor, 2008). As digital information is fragile, vulnerable and susceptible to damage as it is created, used and stored (Waugh et al., 2000), it can be lost due to inability to access stored data and loss in quality or technical failures. The most common digital preservation techniques are migration, emulation, encapsulation and museum (K. Ho lee et al., 2002).

Cloud computing has been defined by the National Institute of Standard and Technology (NIST) as a *"model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction"* (Mell and Grance, 2011).

Uncertainty described as a state of inability to specify an outcome or event with certainty affects costs, technical performance and operation processes in current environment of rapid changing technologies in aspect of cloud computing (Erkoyuncu, 2011). The changes occurred in technology, economy, needs, demands lead to better understanding and detailed study of these uncertainties (MacManus and Hastings, 2006). Obsolescence is the process of becoming out of support or end of life, e.g. when product or service is no longer supported by original manufacturer (Dowling, 2004). Refreshing is one of the digital preservation techniques to keep the infrastructure up-to-date with most recent technology. This can be used to prevent obsolescence and failure of infrastructure components (Barateiro et al., 2010).

It is worth to mention that none of current cost models is focused on LTDP activities in cloud computing technology, which is essential since as a long term process, LTDP is especially subject to a number of uncertainties and obsolescence issues that can occur over time and have an impact on cost. Therefore, there is an apparent lack of a reliable tool capable of estimating and comparing costs for many deployment models, which would be useful for manufacturing industry planning to move into the cloud. This research study has made crucial findings on the costing of uncertainties and obsolescence issues on cloud computing for LTDP activities and development of cost model using combination of activity and parametric cost estimation technique.

3 RESEARCH METHODOLOGY

The challenge of developing a cost model for cloud computing utilisation in LTDP with consideration the impact of uncertainties and obsolescence issues, lies in the need of a consistent and rigorous methodology. The structured sequence of procedures and practices has the purpose of filling the previously described knowledge gap in related research. The adopted research methodology aims at successful and efficient development of the cost model for public and private clouds. Figure 1 illustrates three phases of this methodology: generation of cost breakdown structures, categorisation of uncertainties and obsolescence, development of cost model and Monte Carlo simulation.

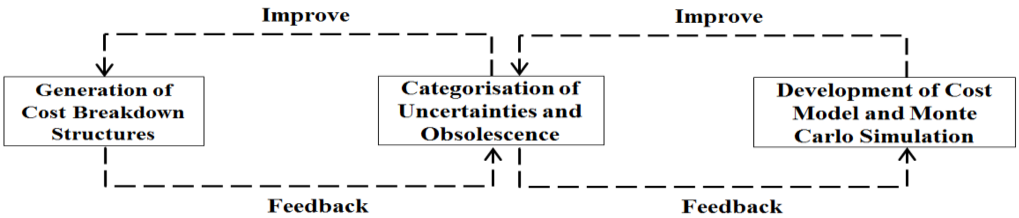


Figure 1: Research Methodology

To start with Phase 1, a number of interviews were conducted with wide range of industries, and then cost drivers were identified as consequences of the analysis of knowledge captured. The cost breakdown structures for public and private clouds were generated and validated by industry experts. Uncertainties and obsolescence issues were explored and categorised into high level classification from both cloud service provider and end-user perspectives. Phase 3 was carried out for the development of cost model. As a result of knowledge gained from phases 1 and 2, the cost model consisting of parameters, equations, rules and assumptions was built using a combination of activity-based and parametric cost estimation techniques; therefore it can be called a hybrid cost model. The impact of categorised uncertainties and obsolescence was assessed, and three-point estimation technique was employed. Then, Monte Carlo simulation was constructed to illustrate probability distribution of future costs. Finally, the cost estimation tool with front-end graphical user interface and a dashboard representation was developed to support decision making on preserving digital information for public and private cloud.

4 RESULTS

4.1 Cost Breakdown Structures

The cost breakdown structure (CBS) for public cloud is presented in Figure 2. It consists mostly of service charges paid to CSPs and other operational costs.

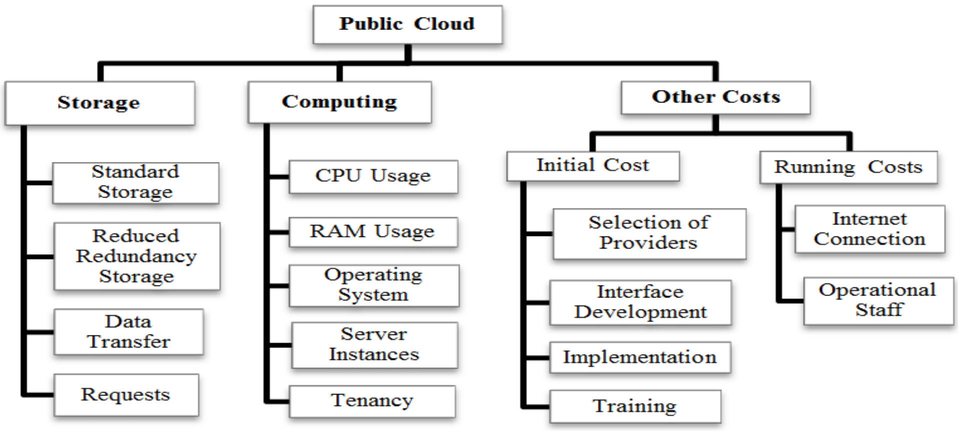


Figure 2: Cost Breakdown Structure for Public Cloud

The two main costs of public cloud utilisation have been identified as storage cost and computing cost. The standard storage cost depends on the exact volume of monthly used storage (in gigabytes). Other elements of the storage cost are size of monthly data transfer and number of requests sent to the cloud. Total computing cost is determined by a number of utilised server instances and their monthly utilisation in hours. The other running costs of public cloud utilisation are broadband connection cost and operational staff cost. As for the initial investment cost, there can be certain initial costs related to selection of providers and services that will fit the organisational requirements, the development and implementation of interfaces and ensure that employees are properly trained to work with cloud technology.

The cost breakdown structure for private cloud is presented Figure 3. The high level breakdown of cost drivers is categorised into acquisition, implementation, running, and redundancy costs.

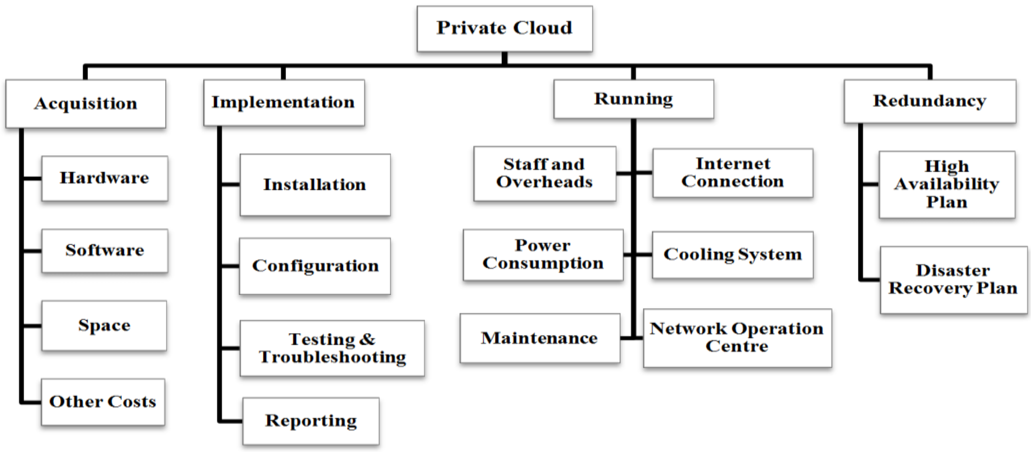


Figure 3: Cost Breakdown Structure for Private Cloud

The acquisition costs are composed of tangible and intangible assets that organisations purchase. These assets are composed of infrastructure such as hardware, software, space, and other costs e.g. power distribution system, cooling system, generator, fire protection, security facilities etc. Second category is implementation costs such as installation, configuration, testing, troubleshooting and reporting costs which depends on the size of data centre and number of labour required. Running (recurring) costs include a number of fixed costs, i.e. monthly operational staff costs and broadband connection cost, as well as variable costs, e.g. cost of power consumed by devices, cooling infrastructure, Network Operation Centre (NOC) and maintenance. Finally, redundancy cost is divided into two parts: high availability plan and disaster recovery plan. Both require acquisition and implementation of backups which are used to reduce downtime of system or to ensure that the system can be restored with best effort without any loss of data in case of any disaster. The amount of redundant infrastructure depends on company's policies. Redundancy plan involves also an additional cost of regular data transfer between primary and redundant systems.

4.2 Categorisation of Uncertainties in Cloud Computing

During deployment of both public and private clouds, there are uncertainties that have minor and major impact on cost, technical performance and operation processes. According to conducted interviews carried out along this research study, the identified uncertainties are classified into five categories: business, economic, regulatory, technological, and physical uncertainties.

Business uncertainties discuss the relationship between organisations and their behaviours. These behaviours relate to unknown situations that both cloud service provider and end user could experience. Business uncertainties can be broken down into four areas such as cloud service provider existence in the future, organisational culture, market trends, and business growth. Economic uncertainties highlights the importance of the economic state of a nation, country, or organisation and their effects on costs. Economic uncertainties can be broken down into inflation, energy prices (e.g. electricity and currency fluctuation if organisation operates in different regions or countries). Regulatory uncertainties interact with governmental and organisation policies, legislation, and standards. Additionally, any changes within the industry standards on digital preservation in cloud and international trade tariffs will have impact on cost. Technological uncertainties relate to the availability, elasticity of infrastructure, data representation (image, text or numeric), and the possibility of new technology. These can be divided into technology evolution, data format, data security, and obsolescence of hardware, software, people's skills and preservation plans. Physical uncertainties relate to tangible and intangible geographical elements, processes that have impact on costs e.g. natural disasters, climate changes, and solar occurrence. These uncertainties have lower chance of occurrence, however potentially high impact on cost if they do.

4.3 Costing Uncertainties and Monte Carlo Simulation

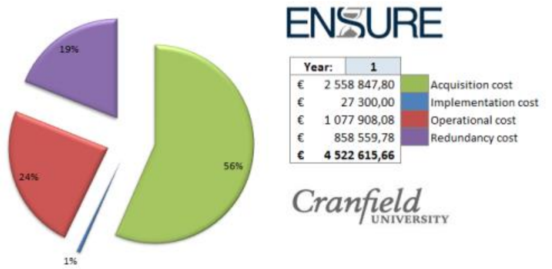
After categorisation of uncertainties, three-point estimation technique was employed for embedding impact of uncertainties into the cost model. Firstly, the maturity level of project definition was analysed based on the framework developed by Association of Advancement for Cost Engineering International (AACEI). Based on the user's input into the cost estimation tool, expected accuracy range is applied for best-case estimate and worst-case estimate. When the best-case, worst-case and most likely estimate values have been identified for each cost driver, Monte Carlo simulation for probability distribution is applied on these three value ranges. The results is displayed in a dashboard representation in order to view cost trends and select specific costs on specific years to support decision making.

4.4 Cost Modelling Tool

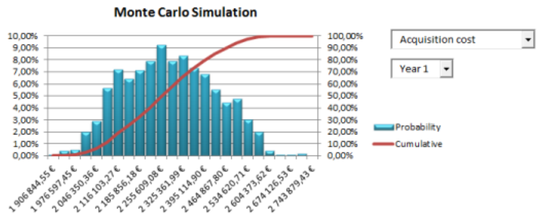
The cost modelling tool, which has been developed using Microsoft Excel and Visual Basic for Applications, is a software implementation of public and private cloud cost model. The tool consists of a number of modules. First, the user has to enter all their preservation requirements, which are needed to run the cost modelling calculations, using an interactive input interface. The calculated costs are then presented in two different worksheets, which are dashboard representation and detailed text representation. The elements of the Dashboard interface are presented in Figure 4.

INPUTS	RESET ALL	
Preserved data per month	600	Terabyte
Number of years to preserve information	15	
Number of servers in the Data Centre	25	
Number of generators for the Data Centre	3	
Amount of Diesel fuel to store	100	litre
Option for space for the Data Centre	Lease	
Number of Computer Room Air Conditioning (CRAC) units	4	
Server operating system	LINUX	
Cloud computing operating system	OpenStack	
Type of terminals	Hardware	
Number of terminals	20	
Redundancy ratios for High Availability and Disaster Recovery	15,00%	10,00%
Maturity level of project definition deliverables	Medium	

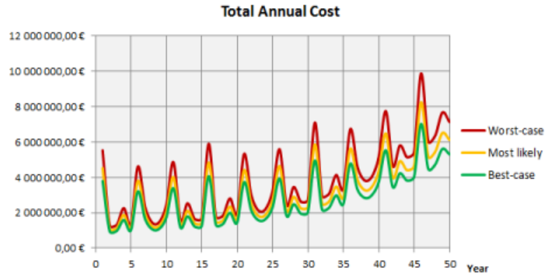
List of all user inputs; the user can alter any of the inputs to do a quick simulation of how this will affect different costs.



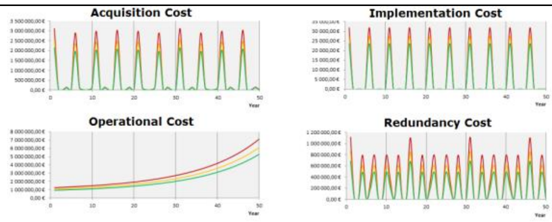
Pie chart showing percentage structure of main cost elements in a year selected by user.



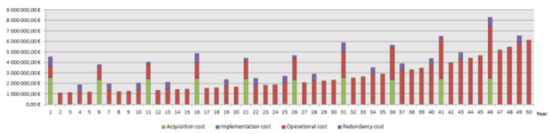
Bar chart showing Monte Carlo simulation results (probability distribution of cost) for a cost driver and year selected by user.



Line chart showing total annual cost (most likely, worst-case and best-case) over up to 50 years.



Set of line charts showing main cost elements (most likely, worst-case and best-case) over up to 50 years.



Bar chart showing total annual cost structure over up to 50 years.

Figure 4: Overview of the Dashboard Elements

Apart from the charts, users can also see a detailed table of all costs with sub-totals and the total annual cost for every year in the detailed annual cost worksheet. In addition to the most likely cost values, the best-case and worst-case values are also provided for each high-level cost.

All parameters for cost modelling calculations (e.g. unit costs) are located in the public cloud and private cloud parameter worksheets, which are hidden for the users. There is also a separate Monte Carlo hidden worksheet, in which the random cost samples needed for Monte Carlo simulation are generated and the Monte Carlo charts are constructed.

5 VALIDATION

All results, methods and techniques employed throughout the research were validated in order to make the outcomes of the research credible. Validation process was continuous and lasted through all three phases of the research. The validation activities were based on a number of interviews and meetings with digital preservation and cloud computing experts. Every interview provided expert feedback on previously obtained research results and improved knowledge. The results of the interviews were systematically analysed and the outcomes of this analysis were used to revise and improve the deliverables. At the end of the last phase of the research study, the developed cost modelling tool for public and private cloud was validated by the ENSURE consortium members in a series of presentation sessions.

6 CONCLUSIONS

The research described within this paper focused on development of a generic cost model for cloud computing utilisation in LTDP. Its main tangible result is an innovative cost estimation tool, which is easy to use and does not require users to have a wide technical knowledge about cloud implementation. Therefore, it may serve as a decision-support tool, not just a cost calculation tool.

Long-term cost modelling for cloud computing is challenging, since it is very difficult to predict many long-term trends, e.g. prices of hardware, as not enough historical data is available. However, by using three-point estimation and Monte Carlo simulation, the developed tool returns a range of possible costs instead of a single point value, which provides the users with a realistic cost estimation.

The conducted research resulted in crucial addition to knowledge for cloud deployment costs, especially identification of detailed cost breakdown structure for private cloud. Private cloud requires high initial investment cost for purchase of infrastructure. However, the return on investment can be economically viable. Public cloud is a cost effective solution, not requiring organisations to make any considerable investments in advance. However, security and control of data becomes an issue.

There are diverse categories of uncertainties and obsolescence issues that should be identified and carefully considered while estimating costs of cloud deployment. The research provides classification of these issues, as well as an attempt to quantify their impact on cost and include it in the cost model.

In manufacturing industry, there are numerous uncertainties and obsolescence issues in long-term that have major and minor impact on cost. This generic cost model developed within this research can be employed for manufacturing, service delivery etc. industries as a decision-supporting cost estimation tool.

ACKNOWLEDGEMENTS

The authors would like to thank European Commission for funding this research, which is a part of ENSURE project (www.ensure-fp7.eu) under Seventh Framework Programme (FP7/2007-2013). They would also like to thank all people who have contributed directly or indirectly to the success of the research project, especially those who shared their knowledge and experience during meetings and interviews.

REFERENCES

Xu, X. (2012), "From cloud computing to cloud manufacturing", *Robotics and Computer-Integrated Manufacturing*, vol. 28, no. 1, pp. 75-86.

Barateiro, J., Antunes, G., Freitas, F. and Borbinha, J. (2010), *Designing Digital Preservation Solutions: A Risk Management-Based Approach*, The International Journal of Digital Curation [Online], Vol. 5, No. 1, pp. 5-16, available at: http: www.ijdc.net/index.php/ijdc/article/view/143/205 (accessed 10/04/2012).

Dowling, T. (2004), *Technology insertion and obsolescence*, Journal of Defence Science, vol.9, no.3, pp.151-155.

Erkoyuncu, J. A. (2011). *Cost Uncertainty Management and Modelling for Industrial Product Service Systems* (PhD thesis). Cranfield University, Cranfield, Bedford, UK

Factor, M. (2008). *Long Term Digital Preservation*. IBM Haifa, available at: http://www.ndpp.in/presentation/National_Workshop2008/Mr._Vijay_K_Garg.pdf (accessed 11/04/2013)

K. Ho lee et al. (2002), *The state of the art and practice in digital preservation*, Journal of Research of the National Institute of Standards and Technology vol. 107, no.1, January-February2002.

McManus, H. L. and Hastings, D. E. (2006). *A Framework for Understanding Uncertainty and its Mitigation and Exploitation in Complex Systems*. Fifteenth Annual International Symposium of the INCOSE, Rochester, NY, pp. 1-20.

Mell, P., Grance, T. (2011). *The NIST Definition of Cloud Computing*. NIST Special Publication 800-145, National Institute of Standards and Technology, U.S. Department of Commerce.

Waugh, A., Wilkinson, R., Hills, B., Dell'oro, J. (2000). *Preserving digital information forever*. In: Proceedings of the fifth ACM conference on Digital libraries, 2-7 June 2000, San Antonio, TX, pp. 175-184