Generating a Performance Test-bed for Cloud Computing Systems

Feifei Chen
Faculty of Information and Communication Technologies
Swinburne University of Technology
Melbourne, Australia 3122
feifeichen@swin.edu.au

***Abstract*—Cloud computing delivers IT solutions as a utility to users. A common economic objective for both cloud consumers and providers is to minimise their total deployment and operational costs while achieving satisfactory system performance to meet Service Level Agreements (SLAs). Therefore, the trade-off between cost and system performance is needed to be managed to achieve the best cost effectiveness. The PhD project presented in this paper provides an efficient and accurate method to evaluate the trade-off between cost and system performance by proposing a performance testing tool for Cloud systems. This tool can accommodate different Cloud system architectures and adopt different workload and cost models of Cloud systems during the trade-off evaluation process.**

Keywords—Cloud computing, performance engineering, cost, trade-off analysis, energy

I. INTRODUCTION

Cloud computing is a new and promising paradigm which delivers computing as a utility [1]. It provides computation, software, data access and storage services through the Internet. Key advantages of Cloud computing include that users can scale on demand their computing and data storage services without the traditional large upfront investment in computing infrastructure.

Building workable Cloud systems requires deep insight into the architectural and performance characteristics of the Cloud, as well as the ability to find alternatives of system designs and deployments. For business purposes, it is imperative to achieve high performance with low cost of software system deployment in the Cloud. For instance, the response time of data centres may be prolonged because of queuing and high CPU load when large amount of data and user requests are being processed simultaneously. One effective solution to this issue is to purchase more computing resources from the Cloud provider to achieve a better system performance. However, it may result in the unexpectedly high extra cost. In addition, energy consumption has become a critical concern in design modern Cloud systems as high energy consumption directly contributes to operational costs, especially when energy unit costs continue to rise significantly. A common economic objective of Cloud providers is to minimise their total deployment and operational costs while achieving satisfactory system performance to meet Service Level Agreements (SLAs).Therefore, we need to manage the trade-off between cost and system performance to achieve the best cost effectiveness before the deployment of Cloud systems. However, this is not a trivial task, as it requires:

- Running extensive experiments with different parameters/metrics and workloads.
- Collecting appropriate Cloud and application cost/performance measurement.
- Performing cost/performance trade-off analysis.

To this end, our research is to design a performance testing tool that supports both Cloud consumers and providers to meet the above three requirements. In addition, predictive cost and performance models will be developed and integrated into the performance testing tool to improve the efficiency of the trade-off analysis. This performance testing tool can accomodate different system architecutres and adopt different workload and cost models during the trade-off evaluation process.

II. MOTIVATION AND RELATED WORK

Much research has been devoted to building performance testing tools for Cloud systems [2-5]. Most of the performance testing tools are built based on simulation technologies, which is a popular methodology to conduct performance evaluation of new software systems [6, 7]. By modelling the interactions and behaviours of each component of the proposed system based on its architecture, the whole system can be simulated. Simulation modelling is steadily becoming more practical with the availability of more powerful and inexpensive computing resources. The key limitation of simulation modelling is that test results may be inaccurate because of the imperfection of environmental configuration and input data in the simulation [8]. In contrast to simulation, model-based performance test-bed generation [8] provides more accurate test results as a test-bed is close to the real software environments. Compared to simulation, less work has been done in generating a performance test-bed for Cloud systems.

In addition, the core functionalities of performance testing tools for Cloud systems are modelling Cloud user behaviour and developing various load testing plans. Most existing tools provide only a fairly basic model of user behaviour: a sequence of user requests on Cloud servers arranged into repeating groups with multiple threads (to mimic large numbers of Cloud users) [5, 9]. Moreover, most existing tools do not support parameterisation of loading tests to allow configuration of different performance test cases and test data. A rich and realistic Cloud system workload model is required before reliable performance evaluation can be conducted.

Appropriate cost measurements for Cloud systems need to be collected before trade-off analysis. Over the last few years, many large-scale data centres have been built due to the massive growth in demand for high performance Cloud data and computational services. As Cloud computing becomes more widespread, increasing data storage and computation needs increase the energy consumed by large Cloud infrastructures. High energy consumption directly contributes to both deployment and operational costs. Much research effort has been devoted to analyse the cost of Cloud systems. Most existing approaches to cost analysis are based on total cost of ownership and utilisation cost from a Cloud provider's and Cloud consumer's perspectives respectively [10, 11]. Few approaches have taken into account energy consumption into the cost model by setting the energy cost rate to a fixed value [12]. However, energy charging policies vary in different regions because of the variety of the "energy dirtiness rate" which reflects the level of electrical pollution. To our best knowledge, there is little work done in checking the "energy dirtiness rate" while analysing energy cost in Cloud systems.

III. APPROACH

Our aim is to support the analysis of the trade-off between system performance and cost of Cloud systems. We aim to reduce the time and effort of conducting performance testing and trade-off analysis while guaranteeing the effectiveness and accuracy of the evaluation process. Our Cloud performance test-bed is expected to assist Cloud system architects to find alternatives of Cloud systems designs and deployments. This tool is also expected to assist performance engineers derive both static and dynamic system-level performance optimisation strategies for Cloud systems. We will implement the tool using Eclipse GMF (Graphical Modelling Framework) [13], which provides a framework for creating visual editors while being model agnostic.

In an attempt to hide the complexity of the performance testing implementation from Cloud systems architects and performance engineers so that they can focus on analysing the performance results, the performance testing tool will be designed to perform load testing automatically based on architecture and workload models specified by end users. This will significantly reduce the complexity of the load testing suites and architecture-specific performance data collection, which in many cases require more time to understand and develop than the testing itself.

The tool will be developed to automatically generate the deployment scripts for Cloud systems based on high-level architectural descriptions. The tool is expected to support alternative architecture choices for performance evaluation. In order to achieve those goals, we need general abstract Cloud system architecture model so that the attributes of Cloud elements and relationships between them can be configured and modified.

Various Cloud system workloads need to be applied during the performance evaluation process based on different workload models. In order to conduct a reliable performance evaluation, a realistic and comprehensive Cloud system workload model is required. Thus, we need a workload modelling technique so that both Cloud user behaviours and application data can be modelled. A form chart model is a technology independent bipartite state diagram used to simulate user behaviours of submit/response systems [14]. It describes what the user sees as system output, and what the user provides as input to the system at a high level. It captures the structure of the target system from the user's perspectives and can be augmented with probabilities to capture user interactions with the target system. A stochastic form chart model has been extended from basic form chart model with stochastic functions to generate performance testing workloads of Web applications [15, 16]. In the stochastic form chart model, the pages of a web site are represented as bubbles, the actions as boxes and the transitions between them as arrows. In addition to the form chart, a form-oriented model specifies message types and user data for all the pages and actions. The stochastic form chart models can be extended to model the Cloud workloads. Cloud users send service requests to the Cloud and get responses from the Cloud with service results. Therefore, a Cloud system can be considered as a submit/response system. The Cloud user behaviours of requesting Cloud services can be modelled by stochastic form charts.

As energy consumption has become a critical concern in designing modern Cloud systems, we will incorporate energy cost in our cost model. An energy consumption model is required to describe the energy consumption patterns of Cloud systems. An "energy dirtiness rate" will be integrated into our energy consumption model to factor in different environment impacts of different Cloud energy sources. In addition, different Cloud charging and billing models allow different choices of hosting options of applications between and even within Cloud platforms. The total cost in the cost model is composed of costs introduced by energy consumption and Cloud resources consumption.

Based on the Cloud system architecture model, workload model and cost model, we will systematically investigate and analyse trade-off between cost and system performance of Cloud systems. Complementary work will be done in real Cloud environments to evaluate the trade-offs from different cases studies and scenarios. This will provide a set of essential strategies to achieve the best cost-effectiveness of Cloud systems. In addition, a user evaluation will be conducted to evaluate the performance test-bed itself.

Figure 1 shows the overview of the proposed approach. Green areas demonstrate the elements of the work done to date. The Cloud architecture models (1) are defined by the end user. The deployment scripts are automatically generated by the transformation engine (2). The structure of deployed application and corresponding workload models can be abstracted and generated automatically based on both application structure and application data (3) specified by the user. The generated workload model will be translated into load testing scripts (4). Based on the system deployment and workload testing scripts, automatic load testing will be performed (5). Performance measurements (5a) and energy usage data (5b) will be collected during the load testing process. The end user will apply the energy usage data to the

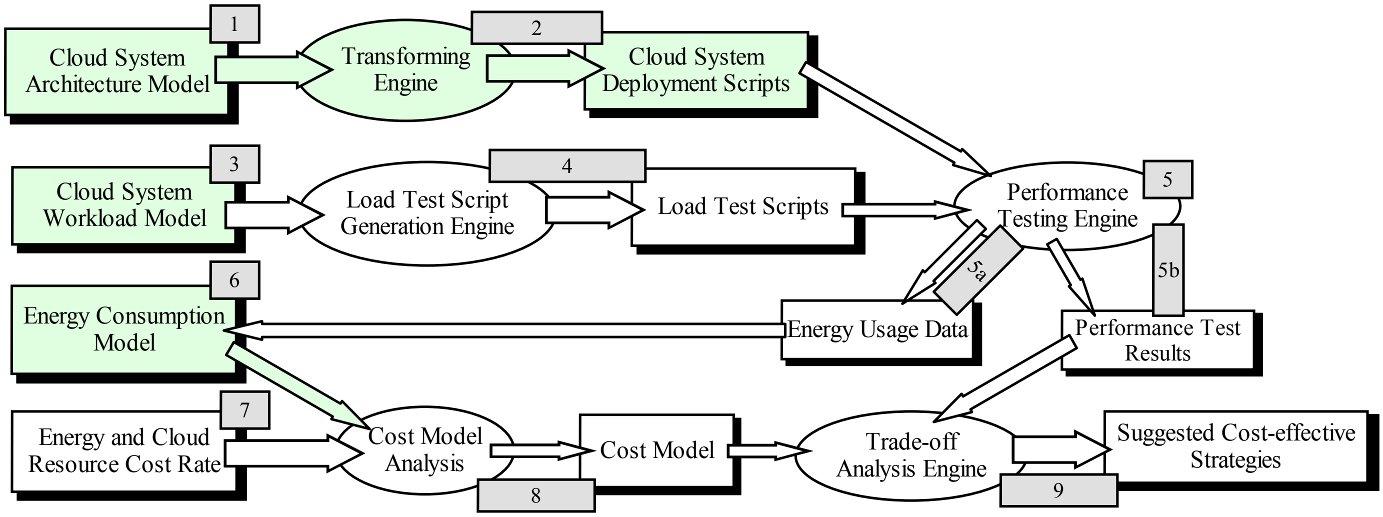

Figure 1. Overview of our cost and performance trade-off analysis approach.

energy consumption model (6) along with the energy and Cloud resource pricing strategy (7) to derive the cost model (8). The trade-off between cost and system performance will be analysed based on the performance measurements and cost model (9). Suggested cost-effective strategies for Cloud system deployment will be presented to end users (9).

IV. METHODOLOY

We are using an incremental and iterative approach to our work with the purpose of reducing risks and uncertainties. All proposed models and implemented tools will be refined after the evaluation of each stage of the research. For each increment, scientific papers will be reported to the academic community through publications in workshops and conferences if significant results have been achieved. An outline of our key research steps is shown below, including:

- Identifying essential elements required in a Cloud system and investigating both interconnection models and behavioural specifications of Cloud system elements.
- Designing and implementing a Cloud system architecture modelling tool.
- Identifying appropriate workload modelling technologies and investigating patterns of Cloud system workload.
- Designing and implementing a Cloud system workload modelling tool.
- Designing and implementing a Cloud load testing engine.
- Integrating the architecture modelling tool, workload modelling tool and load testing engine into our Cloud performance test-bed.
- Evaluating the completeness and correctness of the architecture model and workload model by comparing the performance measurements we already manually collected to the performance data collected by our tool.
- Collecting empirical energy consumption data and performance measurements in Cloud systems and analysing the correlation coefficients of energy consumption and system performance.
- Formulising the energy consumption model based on the collected energy consumption data.
- Designing a energy-aware cost model and integrating it into the performance test-bed.
- Analysing the trade-off between cost and Cloud system performance by performing systematic load testing.

To date we have completed steps (1), (2), (3) and (6) presented in Figure 1.

V. RESEARCH RESULTS AND PROGRESS

Based on our initial design, a Cloud system architecture modelling tool has been developed using GMF. Users can define their Cloud architecture models graphically by specifying the attributes and relationships of required Cloud elements. The completeness and correctness of the architecture model has been validated. The graphical Cloud architecture model is transformed to an XML file for Cloud system deployment.

In order to investigate the energy cost of Cloud systems, we have proposed a general energy consumption model for Cloud systems [17]. In this model, the total energy consumption of a Cloud system is divided into a fixed part and a dynamic part. The fixed part of energy consumption includes the energy consumption during idle time and the energy consumption of the cooling system. The dynamic energy consumption is the energy consumed by running tasks in the Cloud system. Based on the major resource consumed by the task, runtime tasks are categorised into three types: computation-intensive, data-intensive and communication-intensive. We have also identified the parameters of each type of task, i.e. the number of processes, the size of data to be processed and the size of data to be transmitted. These task parameters determine the task workload and impact energy consumption of runtime tasks. To verify and formulise this model, we have conducted extensive experiments to profile the energy consumption in Cloud computing systems based on the abovementioned three types of tasks [18]. We have collected fine-grained energy consumption and performance data with varying system configurations and workloads. Our experimental results show that system configuration and workload can significantly impact the energy consumption and system performance in Cloud systems.

The next stage of our research is ongoing: modelling the workloads of Cloud systems. As we have observed in the experimental analysis of energy consumption, the workloads of different runtime tasks in Cloud systems can significantly impact energy consumption and system performance. Therefore, for different type of runtime tasks, different load test scripts are expected. The workload model will be evaluated by comparing the performance measurements we already manually collected to the performance data collected by our tool.

VI. CONCLUSION

Our initial work has demonstrated that it is necessary to include energy cost into the Cloud system cost model. In addition, the system configurations impact energy consumption and system performance significantly. This provides evidence for identified Cloud elements in our proposed Cloud architecture model. The energy consumption profiling results have shown that the types of applications and user data in the Cloud systems can greatly affect the energy consumption and system performance. This confirms the completeness and correctness of the design and implementation of our workload model.

The key contribution of our research is minimising the time Cloud system architects and performance engineers spend evaluating and analysing the trade-off between system performance and cost before deployment. It helps increase the productivity of load testing of Cloud systems. Furthermore, our research also assists in discovering cost-effective system deployment strategies which can be adopted to support static or dynamic system-level optimisation of Cloud systems.

ACKNOWLEDGMENTS

This PhD research is funded by Australian Research Council under Discovery Project DP110101340. A special acknowledgement is given to the author's supervisors, Prof John Grundy, Prof Yun Yang and A/Prof Jean-Guy Schneider for their guidance and support throughout this research.

REFERENCES

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "Above the clouds: a Berkeley view of cloud computing," UC Berkeley Reliable Adaptive Distributed Systems Laboratory, USA, Technical Report, UCB/EECS-2009-28, Feb 10, 2009.

[2] 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," *Software - Practice and Experience,* vol. 41, no. 1, pp. 23-50, 2011.

[3] I. Sriram, "SPECI, a Simulation Tool Exploring Cloud-Scale Data Centres," in *the 1st International Conference on Cloud Computing(CloudCom 2009)*, Beijing, China, 2009, pp. 381-392.

[4] A. Nunez, J.L.Vazquez-Poletti, A. C. Caminero, J. Carretero, and I. M. Llorente, "Design of a New Cloud Computing Simulation Platform," in *Computational Science and Its Applications*, Santander, Spain, 2011, pp. 582-593.

[5] F. Abbors, T. Ahmad, D. Truscan, and I. Porres, "MBPeT: A Model-Based Performance Testing Tool," in *the 4th International Conference on Advances in System Testing and Validation Lifecycle*, Lisbon, Portugal, 2012, pp. 1-8.

[6] C. M. Woodside, "Software Resource Architecture and Performance Evaluation of Software Architectures," in *the 34th International Conference on System Sciences*, Hawaii, USA, 2001, pp. 1-10.

[7] A. Liu, "Dynamic Distributed Software Architecture Design with PARSE-DAT," in *International Conference on Software Methods and Tools (SMT'00)*, Wollongong, Australia, 2000, pp. 199-207.

[8] J. Grundy, Y. Cai, and A. Liu, "Generation of Distributed System Test-beds from High-level Software Architecture Descriptions," in *the 16th Annual International Conference on Automated Software Engineering (ASE 2001)*, San Diego, CA, USA, 2001, pp. 193-200.

[9] M. Kamra and R. Manna, "Performance of Cloud-Based Scalability and Load with an Automation Testing Tool in Virtual World," in *IEEE the 8th World Congress on Services*, Honolulu, HI, USA, 2012, pp. 57-64.

[10] P. Brebner and A. Liu, "Performance and Cost Assessment of Cloud Services," in *the 8th International Conference on Service Oriented Computing(ICSOC 2010)*, San Francisco, CA, USA, 2011, pp. 39-50.

[11] A. Khajeh-Hosseini, I. Sommerville, J. Bogaerts, and P. B. Teregowda, "Decision Support Tools for Cloud Migration in the Enterprise," in *the IEEE International Conference on Cloud Computing(CLOUD 2011)*, Washington, DC, USA, 2011, pp. 541-548.

[12] X. Li, Y. Li, T. Liu, J. Qiu, and F. Wang, "The method and tool of cost analysis for cloud computing," in *the IEEE International Conference on Cloud Computing (CLOUD 2009)*, Bangalore, India, 2009, pp. 93-100.

[13] (2013, Mar 7). *GMP*. Available: http://www.eclipse.org/modeling/gmp/

[14] D. Draheim and G. Weber, *Form-Oriented Analysis - A New Methodology to Model Form-Based Applications*. New York, USA: Springer, 2005.

[15] D. Draheim, J. Grundy, J. Hosking, C. Lutteroth, and G. Weber, "Realistic Load Testing of Web Applications," in *the 10th European Conference on Software Maintenance and Reengineering (CSMR'06)*, Bari, Italy, 2006, pp. 70-81.

[16] C. Lutteroth and G. Weber, "Modeling a Realistic Workload for Performance Testing," in *the 12th International IEEE Enterprise Distributed Object Computing Conference(ECOC 2008)*, Munich, Germany, 2008, pp. 149-158.

[17] F. Chen, J.-G. Schneider, Y. Yang, J. Grundy, and Q. He, "An energy consumption model and analysis tool for Cloud computing environments," in *the 1st International Workshop on Green and Sustainable Software(GREENS2012)*, Zurich, Switzerland, 2012, pp. 45-50.

[18] F. Chen, J. Grundy, Y. Yang, J.-G. Schneider, and Q. He, "Experimental Analysis of Task-based Energy Consumption in Cloud Computing Systems," in *the 4th ACM/SPEC International Conference on Performance Engineering(ICPE2013)*, 2013, To appear.