To Cloudify or Not to Cloudify that's the Question for a Scientific Data Center

Prof. Prashant J. Gadakh[1], Prof. Ajitkumar Shitole[2], Prof. Vishal Chaudhari[3]

[1] *Assistant Professor, Computer Engineering Department, I2IT, Hinjawadi, Pune, Maharashtra, India*
[2] *Assistant Professor, Computer Engineering Department, I2IT, Hinjawadi, Pune, Maharashtra, India*
[3] *Assistant Professor, Computer Engineering Department, I2IT, Hinjawadi, Pune, Maharashtra, India*

ABSTRACT

In The age of information technology the idea of turning data centers executing scientific batch jobs into private clouds is as attractive as troubling. Cloud platforms may help both in limiting power consumption and in implementing fault tolerance strategies. However, there is also the fear that performance may worsen, and that the electricity required for longer job duration and fault tolerance implementation may overcome the saved one. In this paper, we present the consumability analysis for assessing the impact of cloud and fault tolerance tunings on scientific processing systems. The analysis considers performance, consumption, and dependability aspects, jointly. The aim is to pinpoint if, for a given system, there is a setting where consumption and job failure rate decrease, while performance is not affected. Applied to the scientific data center at our University, the analysis allowed us to find the proper selection of virtual machines' configuration, consolidation strategy, and fault tolerance tuning.

Keyword: - *Cloud computing, scientific computing, performance, consumption, dependability, modeling*

---

1. INTRODUCTION

Many institutions own scientific data centers to perform heavy computations, typically as long-running, non interactive (batch) jobs. While such computing systems are conceived for high performance, they are rarely used at their maximum capacity [1], [2], wasting energy due to the power consumption of machines uselessly active. Moreover, they are affected by a non negligible number of failures, which cause further energy waste (of unended jobs), and impact system dependability1 [4]. Cloud computing platforms are emerging as a promising means to address the above issues, suggesting cloudification of scientific data centers. Jobs can be executed in virtual machines (VMs) according to an Infrastructure- or Platform- as-a-Service model [5]. Cloudification may allow energy saving by consolidating VMs in a reduced number of physical machines (PMs), and by putting in standby or switching off idle ones [6], [7]. Virtualization is appealing also as it eases the implementation of fault tolerance mechanisms; for instance, the checkpointing of jobs is simplified by snapshotting whole VMs, whereas job replication is turned into VM replication [8], [9]. Unfortunately, virtualization and cloudification may also worsen performance and consumption, frustrating or even outweighing the expected benefits [6], [10]. In [11], we presented preliminary results indicating that performance, consumption, and dependability aspects of a scientific data center mutually affect each other, when using virtualization and fault tolerance. In this paper, we propose a consumability analysis technique for quantitative assessment of such mutual relations when cloudifying a scientific data center. The analysis serves for estimating the impact of cloud solutions, and for proper management and tuning. The technique is based on stochastic models, conceived to be fed with field data from the system under analysis. We apply the analysis to the S.Co.P.E. system at our University, demonstrating that a cloudified scientific data center can be tuned so as to reduce power consumption and job failure probability without affecting performance. After discussing related research (Section 2), the paper proceeds as follows:

1) The consumability models of a physical batch system (PS) (Section 3) and of the corresponding cloudified system (CS) (Section 4) are proposed, encompassing performance, consumption and availability.
2) Real data gathered by monitoring a scientific data center are used to estimate input parameters of PS (Section 5) and CS models (Section 6). This case study serves both for validating the models and as guideline to practitioners;
3) Models are solved (Section 7), demonstrating that it is possible to find a proper trade off among performance, consumption and dependability by: (i) instantiating a number of VMs per PM which reduces the VMs management overhead while providing flexibility and execution isolation, (ii) consolidating VMs with different load types, thus parallelizing operations with different resource requirements, and (iii) selecting and tuning the fault tolerance strategy so as its cost is balanced by the saving of energy due to job failures.

2. RELATED WORK

The common deployment models for cloud computing are public cloud and private cloud [5]. In the former case, a service provider supplies a cloud infrastructure for open use by customers. In the latter, the infrastructure is meant for exclusive use by an organization. Many studies discuss how scientific computing can be migrated to a public cloud to avoid the building and ownership costs of large data centers [6], [10], [12], [13], [14]. The focus of this paper is on private cloud: institutions that already own a data center may adopt cloud and virtualization as management means. For the S.Co.P.E. scientific data center, discussed in this paper, a grid-on-cloud prototype is presented and validated in [15]. Other platforms for virtualizing scientific processing systems by integrating virtual cluster provisioning have also been proposed [16]. While the objectives of these works partially overlap with the ones in this paper, they are still in their infancy and experimental results do not prove the reaching of the goals. More studies try to find the configuration that improves performance, consumption or dependability. They often focus on either (i) the placement of virtual machines on physical machines (VM consolidation), or (ii) the placement of tasks on VMs (similar to the common scheduling problem).

Performance issues are usually faced by searching for consolidation strategies mixing VMs with different load in a same server. Experimental results show that for compute intensive tasks, performance gradually degrades when the number of co-hosted VMs grows, while in the case of I/O operations, performance degradation occurs due to additional delay for data processing [17]. Similarly, CPU-bound and memory-bound VMs can be consolidated for efficiently exploiting resources [7]. Apart from reducing the performance decay introduced by the virtualization layer, how the power consumption may benefit from the proper scheduling of VMs is studied in the Magellan Project [18]. It investigates the role of cloud computing in addressing energy related issues of mid-range data-intensive computing. Main findings revealed that cloud can be used even for scientific computations, but their specific requirements are to be taken into account, above all in the case of I/O bound operations (e.g., communication operations, disk read/write operations). With respect to these studies, we consider not only the impact of consolidation on performance and consumption, but also on dependability. Dependability issues are often considered as “SLA violations” [6], [7], [10], [19]. Using this count as dependability metric is not always correct, since a violation may simply represent the performance going below a certain threshold, while the actual faulty behavior is neglected. The search for energy-aware trade-offs between reliability and performance is discussed in [19], which compares several consolidations. Advantages of implementing checkpointing and replication fault tolerance strategies in virtualized environments are discussed in [8], [9], [22], [23]. Hypervisor-based fault tolerance (HBFT) implements the checkpoint-recovery protocol, but it causes a large overhead to VMs, then several optimizations should be considered for its adoption [9]. The use of VMs’ snapshots for implementing checkpointing appears as a feasible solution, instead [8], [22]. The overhead of the mechanism, which should suspend, snapshot, and resume VMs, can be reduced by using incremental snapshotting [8]. Replication-based fault tolerance is described in [23]. A program is executed in a VM for which an active spare is created; that is, another VM acts as a replica and it is updated up to forty times a second, thus its state is just slightly delayed with respect to the main VM.

The importance of considering the mutual relations among performance, energy efficiency, and dependability attributes in virtualized environments was discussed in our previous paper [11]. Results paved the way towards the consumability analysis for the joint assessment of the three aspects, which in this paper is discussed in detail and applied to a real system for tuning several factors, demonstrating the value added of cloud as a means for scientific data center management.

3. SYSTEM MODELING

We adopt a model-based approach, since it offers an abstraction of the real system leaving out unnecessary details and permits to avoid system implementation. The modeling is from the job execution perspective. In a system running batch jobs, both performance degradation and failures directly reflect on the jobs submitted by the users, and a job, in turn, affects the consumption of the system. The problem is to identify the relations among performance levels, energy consumption dynamics, and failures. We use a hierarchical approach; that is, performance, consumption, and failures are first modeled separately and then composed in an overall model taking into account their mutual relationships.

3.1 Performance Model

The performance model describes the correct execution flow of a batch job according to popular resource managers/job schedulers for batch systems, such as Globus Toolkit,3 Torque, 4 Univa Grid Engine.5 A job submitted to the system is queued; then it becomes running when resources for its execution are identified and its execution time arrives; when the execution terminates, the job is completed and results are returned to the user during the exiting stage. The SRN in Fig. 1 models such a job life cycle. Transition are represents the job arrival; then, a job is queued and a token is added to place Q. Scheduling and starting of a job execution are modeled by transition sch. Upon its firing, a token is moved from Q to R. Transition *cpl* depicts the completion of the job by removing a token from R and adding a token to C. Upon firing of transition *exg*, a token is removed from C, mimicking the notification of the final result to the user. gs is a guard (or enabling) function used for inhibiting the scheduling capacity when no enough free resources are available for executing a job. This depends on the number of already running jobs and on the total available resources. Probability distributions of inter-arrival, scheduling, completion units.

4. CLOUDIFIED SYSTEM MODELING

In the case of a cloud-based batch system, we consider (i) jobs executed in VMs, (ii) each PM hosting up to a certain number of VMs, (iii) switched on only PMs running at least one VM, and in standby the others (i.e., with a reduced power consumption, but requiring a certain time before VMs can be hosted). The performance and failure models have to consider additional events specific to a CS (e.g., VM creation, failure of a VM creation). The consumption model does not change with respect to the physical system (it differs for the weights, which are inputs of the model. The system basically a cloud based batch system, hence performance is the major challenging factor in that model.

4.1 Performance Model

Before executing a job in a VM, that machine is to be instantiated on a PM. This operation is called provisioning of a VM. After the execution, the machine is to be deleted and resources of the PM freed. This is called deprovisioning of a VM [35]. Hence, with respect to the performance model discussed in Section 3.1, we add two transitions modeling the provisioning phase and the deprovisioning phase.

5. CONCLUSIONS

Cloudification of scientific data centers is attractive for improving their management and efficiency. Nevertheless, all that glitters is not gold. Cloud is based on virtualization, which significantly impacts performance. Also, the hypervisor causes a consumption increase in the hosting nodes. Finally, failures happen, and fault tolerance has a cost. Cost benefit analysis of management strategies for a scientific processing system can neither overlook any of performance, consumption and dependability aspects, nor their mutual effects. This paper presented the consumability analysis to estimate the impact of cloudification on such attributes, and the case study of a real scientific data center, which demonstrated the effectiveness of this analysis as a means for administrators to properly tune their systems. Results showed that Cloudification can definitely

6. REFERENCES

[1] A. Verma, P. Ahuja, and A. Neogi, "Power-aware dynamic placement of HPC applications," in Proc. 22nd Annu. Int. Conf. Supercomput., 2008, pp. 175–184.
[2] T. K. Samuel, T. Baer, R. G. Brook, M. Ezell, and P. Kovatch, "Scheduling diverse high performance computing systems with the goal of maximizing utilization," in Proc. Int. Conf.High Perform. Comput., 2011, pp. 1–6.
[3] A. Avizienis, J. C. Laprie, B. Randell, and C. Landwehr, "Basic concepts and taxonomy of dependable and secure computing,"
IEEE Trans. Dependable Secure Comput., vol. 1, no. 1, pp. 11–33, Jan.–Mar. 2004.
[4] A. Moody, G. Bronevetsky, K. Mohror, and B. R. De Supinski,
"Design, modeling, and evaluation of a scalable multi-level checkpointing system," in Proc. ACM/IEEE Int. Conf. High Perform.
Compu., Netw., Storage Anal., 2010, pp. 1–11.
[5] NIST - National Institute of Standards and Technology, P. Mell, and T. Grance, "The NIST definition of Cloud Computing," Special
A. Beloglazov, J. Abawajy, and R. Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for
Cloud computing," Future Generation Comput. Syst., vol. 28, no. 5, pp. 755–768, 2012.
[7] C. Mastroianni, M. Meo, and G. Papuzzo, "Probabilistic consolidation of virtual machines in self-organizing cloud data centers," IEEE Trans. Cloud Comput., vol. 1, no. 2, pp. 215–228, Jul. 2013.
[8] B. Nicolae and F. Cappello, "BlobCR: Efficient checkpoint-restart for HPC applications on IaaS clouds using virtual disk image snapshots," in Proc. ACM Int. Conf. High Perform. Comput., Netw.,
Storage Anal., 2011, pp. 1–11.
[9] J. Zhu, Z. Jiang, Z. Xiao, and X. Li, "Optimizing the performance of virtual machine synchronization for fault tolerance," IEEE Trans. Comput., vol. 60, no. 12, pp. 1718–1729, Dec. 2011.
[10] K. Le, J. Zhang, J. Meng, R. Bianchini, Y. Jaluria, and T. D. Nguyen, "Reducing electricity cost through virtual machine placement in high performance computing clouds," in Proc. Int. Conf. High Perform. Comput., Netw., Storage Anal., 2011, pp. 1–12.
[11] M. Cinque, D. Cotroneo, F. Frattini, and S. Russo, "Cost-benefit analysis of virtualizing batch systems: Performance energydependability trade-offs," in Proc. 6th IEEE/ACM Int. Conf. Utility Cloud Comput., 2013, pp. 264–268.
[12] A. Iosup, S. Ostermann, N. Yigitbasi, R. Prodan, T. Fahringer, and D. Epema, "Performance analysis of cloud computing services for many-tasks scientific computing," IEEE Trans. Parallel Distrib.
Syst., vol. 22, no. 6, pp. 931–945, Jun. 2011.
[13] E. Roloff, M. Diener, A. Carissimi, and P. O. A. Navaux, "High performance computing in the cloud: Deployment, performance and cost efficiency," in Proc. Int. Conf. Cloud Comput. Technol. Sci.,
2012, pp. 371–378.
[14] X. Zhu, L. T. Yang, H. Chen, J. Wang, S. Yin, and X. Liu, "Realtime tasks oriented energy-aware scheduling in virtualized clouds," IEEE Trans. Cloud Comput., vol. 2, no. 2, pp. 168–180,
Apr. 2014.
[15] G. B. Barone, R. Bifulco, V. Boccia, D. Bottalico, R. Canonico, and L. Carracciuolo, "GaaS: Customized grids in the clouds," in Proc. 18th Int. Conf. Parallel Process. Workshops, 2013, pp. 577–586.

[16] Y. Zhao, Y. Zhang, W. Tian, R. Xue, and C. Lin, "Designing and deploying a scientific computing cloud platform," in Proc. ACM/
IEEE 13th Int. Conf. Grid Comput., 2012, pp. 104–113.
[17] S. Lee, R. Panigrahy, V. Prabhakaran, V. Ramasubramanian, K. Talwar, L. Uyeda, and U. Wieder, "Validating heuristics for virtual machines consolidation," Microsoft Res., Redmond, WA, USA, Tech. Rep. MSR-TR-2011-9, 2011.
[18] K. Yelick, S. Coghlan, B. Draney, and R. S. Canon, U.S. Dept.
Energy, "The Magellan report on cloud computing for science," Magellan Project Final Report, Dec. 2011.
[19] W. Deng, F. Liu, H. Jin, X. Liao, H. Liu, and L. Chen, "Lifetime or energy: Consolidating servers with reliability control in virtualized cloud datacenters," in Proc. IEEE 4th Int. Conf. Cloud Comput.
Technol. Sci., 2012, pp. 18–25.
[20] R. Ghosh, F. Longo, F. Frattini, S. Russo, and K. S. Trivedi,
"Scalable analytics for IaaS cloud availability," IEEE Trans. Cloud Comput., vol. 2, no. 1, pp. 57–70, Jan. 2014.
[21] R. Ghosh, F. Longo, R. Xia, V. K. Naik, and K. S. Trivedi, "Stochastic model driven capacity planning for an Infrastructureas- a-Service cloud," IEEE Trans. Serv. Comput., vol. 7, no. 4, pp. 667–680, Sep. 2013.
[22] I. Krsul, A. Ganguly, J. Zhang, J. A. B. Fortes, and R. J. Figueiredo, "VMPlants: Providing and managing virtual machine execution environments for grid computing," in Proc. ACM/IEEE Conf. Supercomput., 2004, p. 7.
[23] B. Cully, G. Lefebvre, D. Meyer, M. Feeley, N. Hutchinson, and A. Warfield, "Remus: High availability via asynchronous virtual machine replication," in Proc. 5th USENIX Symp. Netw. Syst. Design Implementation, 2008, pp. 161–174.
[24] G. Ciardo, A. Blakemore, P. F. Chimento Jr., J. K. Muppala, and K. S. Trivedi, "Automated generation and analysis of Markov reward models using stochastic reward nets," in Linear Algebra, Markov Chains, and Queueing Models, IMA Volumes Math. Appl., vol. 48, C. Meyer and R. J. Plemmons. Heidelberg, Germany: Springer, 1993, pp. 145–191.
[25] B. Subramaniam and W. C. Feng, "Statistical power and performance modeling for optimizing the energy efficiency of scientific computing," in Proc. IEEE/ACM Int. Conf. Green Comput. Commun., 2010, pp. 139–146.
[26] R. Bertran,M. Gonzelez, X. Martorell, N. Navarro, and E. Ayguade, "A systematic methodology to generate decomposable and responsive power models for CMPs," IEEE Trans. Comput., vol. 62, no. 7, pp. 1289–1302, Jul. 2013.
[27] E. Heien, D. Kondo, A. Gainaru, D. LaPine, B. Kramer, and F. Cappello, "Modeling and tolerating heterogeneous failures in large parallel systems," in Proc. Int. Conf. High Perform. Comput., Netw., Storage Anal., 2011, pp. 1–11.
[28] C. Di Martino, Z. Kalbarczyk, R. K. Iyer, F. Baccanico, J. Fullop, and W. Kramer, "Lessons learned from the analysis of system failures at petascale: The case of blue waters," in Proc. Int. Conf. Dependable Syst. Netw., 2014, pp. 610–621.
[29] Rice University—Division of Information Technology. (2013, Apr.). Why Are my jobs not running? [Online]. Available: http:// rcsg.rice.edu/rcsg/shared/scheduling.html
[30] IGI—Italian Grid Infrastructure. (2014, Jul.). Troubleshooting guide for CREAM [Online]. Available: https://wiki.italiangrid. it/twiki/bin/ view/CREAM/TroubleshootingGuide.
[31] D. Cotroneo, F. Frattini, R. Natella, and R. Pietrantuono, "Performance degradation analysis of a supercomputer," in Proc. IEEE 23rd Int. Symp. Softw. Rel. Eng. Workshops, 2013.
[32] K. S. Trivedi, Probability and Statistics with Reliability, Queuing and Computer Science Applications, 2nd ed. New York, NY, USA: Wiley, 2001.
[33] C. Hirel, B. Tuffin, and K. S. Trivedi, "SPNP: Stochastic Petri Nets. Version 6," in Proc. 11th Int. Conf. Comput. Perform. Eval.: Model. Techn. Tools, 2000, pp. 354–357.
[34] A. J. Oliner, R. K. Sahoo, J. E. Moreira, and M. Gupta, "Performance implications of periodic checkpointing on large-scale cluster systems," in Proc. 19th IEEE Int. Parallel Distrib. Process. Symp., 2005,p. 299.
[35] F. Frattini, R. Ghosh, M. Cinque, A. Rindos, and K. S. Trivedi,
"Analysis of bugs in Apache Virtual Computing Lab," in Proc.
IEEE/IFIP Int. Conf. Dependable Syst. Netw., 2013, pp. 1–6.
[36] F. Frattini. (2014, May). Consumability analysis of batch processing systems, Ph.D. dissertation [Online]. http://143.225.8
1.109/ www.mobilab.unina.it/tesi/PhDThesis_Flavio_Frattini_2014.pdf
[37] A. Rezaei, H. Salimi, and M. Sharifi, "Improving software dependability using system-level virtualization: A survey," in Proc. IEEE 24th Int. Conf. Adv. Inf. Netw. Appl. Workshops, 2010, pp. 195–199.

[38] M. Chtepen, F. H. A. Claeys, B. Dhoedt, F. De Turck, P. Demeester, and P. A. Vanrolleghem, “Adaptive task checkpointing and replication: Toward efficient fault-tolerant grids,” IEEE Trans. Parallel Distrib. Syst., vol. 20, no. 2, pp. 180–190, Feb. 2009.
[39] Universit_a degli Studi di Napoli Federico II. (2014, Jul.). S.Co.P.E. [Online]. Available: http://www.scope.unina.it.
[40] CERN. (2014, Jul.). ATLAS experiment [Online]. Available:
http://atlas.ch