Internships as an Application of Cloud Computing Solutions for Education at Universities

Jerzy Kotowski

Institute of Computer Engineering, Control and Robotics,
Wrocław University of Technology,
11/17 Janiszewskiego Street, 50-372 Wrocław, Poland
jerzy.kotowski@pwr.wroc.pl

1 Introduction

Cloud computing is a Web-based processing, that allow to share resources, software, and information over the Internet. Cloud computing helps enterprises and other institutions like schools, universities, etc. transform business and technology. Most cloud computing infrastructures consist of services delivered through common centers and built on servers.

During the past 40 years, information technology has undergone many revolutions in how applications and data have been delivered to users. Mainframes provided a centralized computation facility where end users consumed resources on a shared basis. Client-server architectures offered flexibility and lowered computing costs while bringing more power to the desktop. Mobile computing introduced the notion of anytime, anywhere application access from a laptop or handheld device. Now cloud computing offers a new approach that will enable you to deliver IT services on demand.

Typical benefits of managing in the cloud are: reduced cost, increased storage and flexibility. Experts assume that the challenges of the cloud computing are increasing of data protection, growing data recovery and availability or the growing management capabilities. Clouds are essentially data centers that require high energy usage to maintain operation. High energy usage is undesirable since it results in high energy cost. This problem of minimizing energy consumption maps to the 2-dimensional bin-packing problem. In this paper the detailed model of this problem will be presented.

Cloud computing is a category of computing solutions in which a technology and/or service lets users access computing resources on demand, as needed, whether the resources are physical or virtual, dedicated, or shared, and no matter how they are accessed (via a direct connection, LAN, WAN, or the Internet). This approach is a very useful tool that allow to simplify and upgrade educational process at universities. In this paper we present some particular applications in this area. Namely we present how cloud approach helped to organize students internships at the Wroclaw University of Technology.

R. Moreno-Díaz et al. (Eds.): EUROCAST 2013, Part II, LNCS 8112, pp. 156–165, 2013.

2 A Virtual Machine

System virtualization provides low-cost, flexible and powerful executing environment. It plays an important role in the infrastructure of Cloud computing.

A major problem in the distributed systems is how to ensure that jobs finish their execution within the estimated completion times in the real time computing. Networking research funding agencies in many countries are encouraging research on revolutionary networking architectures that may or may not be bound by the restrictions of the current TCP/IP based Internet. In this area scientists considers new architectures, security mechanisms, content delivery mechanisms, management and control frameworks, service architectures, as like as the routing mechanisms.

Business stakeholder demands on IT are increasing. Every business decision impacts IT, and accelerating market forces reward first movers. Yet most enterprise applications and services are built on top of tightly coupled technology stacks that are challenging to change and costly to manage. Provisioning a new email server or business intelligence engine, for example, can require months of waiting just for hardware purchases and system image configurations.

Caught between shrinking resources and growing business needs, organizations are looking to cloud computing to provide a more efficient, flexible and cost-effective model for computing—one that allows IT to operate much more efficiently and respond faster to business opportunities. The goal is to enable IT as a Service, and cloud computing provides the technical architecture to deliver it.

Cloud computing systems provide a new paradigm to the distributed processing of digital data. As mentioned above Cloud computing's importance rests in the cloud's potential to save investment costs in infrastructure, to save time in application development and deployment, and to save resource allocation overhead. Typical benefits of managing in the cloud are: reduced cost, increased storage and flexibility. Due to these benefits, the use of High Performance Computing infrastructure to run business and consumer based IT applications has increased rapidly during the last few years. Cloud platform facilitates deployment of applications without the cost and complexity of buying and managing the underlying hardware and software layers. For this moment at least six deployments models of the cloud computing are available. The most popular and useful deployment models presently are as follows: a public cloud, a community cloud, a hybrid cloud and a private cloud. In any case a Virtual Machine (VM) is putted into the motion.

A virtual machine is a software implementation of a machine (i.e. a computer) that executes programs like a physical machine. An essential characteristic of a virtual machine is that the software running inside is limited to the resources and abstractions provided by the virtual machine - it cannot break out of its virtual environment. At the moment, there exists hundreds of application that support cooperation between server and user PC, which are being installed by user. Among them, we can distinguish such as: VMware, VM from IBM, etc.

3 Deployment Models

Clouds are essentially data centers that require high energy usage to maintain operation [5]. Today, a typical data center with 1000 racks need 10MW of power to operate. High energy usage is undesirable since it results in high energy cost. For a data center, the energy cost is a significant component of its operating and up-front costs. Therefore, Cloud providers want to increase their profit or Return on Investment (ROI) by reducing their energy cost. Many Cloud providers are thus building different data centers and deploying them in many geographical locations so as not only to expose their Cloud services to business and consumer applications but also to reduce energy cost.

Cloud platform services computing platform and solution stack as a service, often consuming cloud infrastructure and sustaining cloud applications. It facilitates deployment of applications without the cost and complexity of buying and managing the underlying hardware and software layers. Cloud computing is a computing model that delivers information technology (IT) as a service over the Internet. Today, cloud computing has matured to a point where it's considered a mainstream technology service. People touch the cloud every day without knowing it by sending instant messages and sharing les easily between companies over the Internet, staying connected on projects with colleagues and other businesses through new social networking tools that take advantage of the ability to "come together in the cloud" (from CNN).

For this moment at least for deployments models of the cloud computing are available and useful. They are as follows: a public cloud, a community cloud, a hybrid cloud and a private cloud. Public cloud or external cloud describes cloud computing in the traditional mainstream sense, whereby resources are dynamically provisioned on a ne-grained, self-service basis over the Internet. A community cloud may be established where several organizations have similar requirements and seek to share infrastructure. The costs are spread over fewer users than a public cloud (but more than a single tenant). This option may oer a higher level of privacy, security and/or policy compliance.

Hybrid cloud is also called hybrid delivery by the major vendors including HP, IBM, Oracle and VMware who oer technology to manage the complexity in managing the performance, security and privacy concerns that results from the mixed delivery methods of IT services. Private cloud has been described as neologism, but its concept pre-dates the term cloud by 40 years. Enterprise IT organizations use their own private cloud (s) for mission critical and other operational systems to protect critical infrastructures. So, private clouds are not an implementation of cloud computing at all, but are in fact an implementation of a technology subset: the basic concept of virtualized computing.

Cloud computing could also prove to be attractive to academic institutions. The global bailouts of the financial systems with hundreds of billions of tax payers' money are likely to impact on public spending in many countries [16]. Colleges and universities are always on the lookout to upgrade their software and IT hardware in order to attract students and keep pace with the rapid developments in digital technologies.

Cloud computing could provide those institutions with the means to achieve those ambitions at prices they can afford. Furthermore, shifting responsibility to external providers for managing some aspects of their software and hardware infrastructures could also result in cost savings with relation to labor, as fewer IT services staff will be needed than before.

4 Optimization at the Deployment Process

The main area of our interest is to build at the Wroclaw University of Technology the public cloud that will provide educational software to K-12 schools on the Low Silesia district. For this moment we assume that necessary back end elements of the cloud will the computers working at our University. We formulated and solved the deployment procees as a special kind of the discrete optimization problem.

Existing meta-heuristic algorithms are based on ideas found in the paradigm of natural or artificial phenomena. For our purposes we reworked an optimization procedures based on the idea of Genetic Algorithm and Harmony Search Algorithm [11,8]. The basic concept of Genetic Algorithm is designed to simulate processes in natural system necessary for evolution. The newest method in the area of the meta-heuristic algorithms is the Harmony Search Algorithm (HSA). It was conceptualized from the musical process of searching for a perfect state of harmony, such as jazz improvisation. The details of the mathematical model will be described in the next section [1].

5 Problem Description

Given:

- The number of virtual machines. Each of them is described by needed/required resources (CPU performance, memory). Required working hours are also known. It is assumed that these requirements may be different for each machine.
- There is a cloud computing system consisting of multiprocessor computers (host machines) with defined parameters.

Assumptions/restrictions:

- Each virtual machine must be deployed on a single host. It is not allow for a virtual machine to use resources from several hosts.
- The hardware resources of at least one host machines are larger than those required by the virtual machine.
- Any VM has a given priority. It is a real positive number.
- The most important parameter for the VM is its performance. The minimal feasible performance is known as like as its advisable value.
- Working hours for each VM are known.

Task:

Deploy each virtual machine on the particular host.

Mathematical Model

To achieve an appropriate mathematical model of the problem described about first, we have to improve an appropriate notation:

n - the number of virtual machines;
m - the number of hosts;
T - time limit: $t\epsilon\,[0,T]$;
p_{i} - advisable performance of i-th VM, i=1,2,. . . ,n;
q_{i}- minimal necessary performance of i-th VM, i=1,2,. . . ,n;
c_{i}- the priority of i-th VM, i=1,2,. . . ,n;
t_{i}- the starting time of i-th VM, i=1,2,. . . ,n;
v_{i}- the ending time of i-th VM, i=1,2,. . . ,n;
f_j- feasible performance of the j-th host, j=1,2,. . . ,m.

Our aim is to find the optimal value of the two sets of decision variables $x, y\epsilon R^n$, for which

$$F\left(x,y\right)=\sum_{i=1}^{n}c_i 1\left(x_i\right)\frac{y_i}{p_i}\rightarrow max \tag{1}$$

Let $x = (x_1, x_2, ..., x_n)$ and let $y = (y_1, y_2, ..., y_n)$. For any i=1,2,...,n, x_i denotes the number of the host to which i-th virtual machine should be deployed: $x_i\epsilon\left\{1,2,...,m\right\}$. Similarly, for any i=1,2,...,n, y_i denotes the optimal performance of the i-th virtual machine. Of course, for each i=1,2,...,n the following local restriction: $q_i \leq y_i \leq p_i$ should be fulfilled. The feasible solution should also fulfill the set of m global restrictions:

$$\sum_{x_i=j} p_i \leq f_j,\ j = 1, 2, ..., m \tag{2}$$

Problem Analysis

We may study two versions of the problem presented above.

- *Version I:* all parameters of the virtual machines are random numbers. Virtual machines (VM) form the stochastic process. This assumption leads to the management problem in the real time.
- *Version II:* all parameters of the virtual machines are known previously. This assumption leads to the cutting/packing problem (this paper).

One may show that this problem may be easy transformed to the Strip Packing Problem [19]. The Strip Packing Problem or Strip Cutting Problem (SCP) is formulated as follows: to pack (or cut) a set of small rectangles (pieces) into a bin (strip) of fixed width but unlimited length. The aim is to minimize the length to which the bin is filled.

Problem (1)-(2) is classified as 2/V/O/R. It was shown that SCP is NP-complete since it is a generalization of the cutting stock problem, which in turn is a generalization of the famous knapsack problem.

In classic SCP we have a single, initial strip $\boldsymbol{R}$ of fixed width W and unlimited length and a finite set $\boldsymbol{S}$ of n small rectangles (pieces) with dimensions (l_i, w_i), for i=1, . . ., n. The problem is to cut off from the strip all pieces such that: all pieces have fixed orientation, all applied cuts are of guillotine type, no overlapping between pieces and between pieces and the edges of the strip. the aim is to minimize the length to which the initial strip $\boldsymbol{R}$ is filled.

To transform problem (1)-(2) to SCP is enough to assume that we have m strips (hosts). For each strip j its width equals feasible performance of the j-th host, j=1,2,. . . ,m and its length is related to the computing time when the host is in used. Similarly, y_i is the width of virtual machine number i=1, . . ., n.

The Method and the Algorithm

The problem is NP-complete. More, in practice the total number of VM goes to hundreds. Decisions variables are of integer and real type. In the problem occur many additional restrictions. We must also remember about the time consuming procedures (fitness function, decision variables coding and decoding). The final conclusion may be only one: a meta-heuristic approach should be used. Namely we decided to put into the motion the newer version of Harmony Search Algorithm (HSA) named Improved HSA (IHSA).

Classic HSA was recently developed in an analogy with music improvisation process where music players improvise the pitches of their instruments to obtain better harmony. Algorithm scheme contains the following steps:

1. Initialize the problem and algorithm parameters.
2. Initialize the harmony memory.
3. Improvise a new harmony.
4. Update the harmony memory.
5. Check the stopping criterion.

HSA uses the following parameters: *HMS* - The harmony memory HM size, *HMCR* - harmony memory considering rate, *PAR* - pitch adjusting rate, *bw* - an arbitrary distance (bandwidth) and *NI* - the number of improvisations.

The traditional HSA algorithm uses fixed value for both *PAR* and *bw*. In the HSA method *PAR* and *bw* values were adjusted in initialization step (Step 1) and cannot be changed during new generations. The key difference between IHSA and traditional HSA method is in the way of adjusting *PAR* and *bw*. Small *bw* values in final generations increase the fine-tuning of solution vectors, but in early generations *bw* must take a bigger value to enforce the algorithm to increase the diversity of solution vectors. Large *PAR* values with small *bw* values usually cause the improvement of best solutions in final generations which algorithm converged to optimal solution vector. Usually, this parameters changes during the calculation process due to the rules presented in (3)-(4).

$$PAR\,(gn) = PAR_{min} + \tfrac{PAR_{max} - PAR_{min}}{NI} gn \tag{3}$$

$$bw\,(gn) = bw_{max} e^{c \cdot gn} \tag{4}$$

where c in (4) is given as:

$$c = \frac{ln\left(\frac{bw_{min}}{bw_{max}}\right)}{NI} \tag{5}$$

In (3)-(5) gn stands for the generation number.

A numerical application was written in C++. For the proper work an algorithm was equipped in three necessary procedures: coding and decoding procedures used after primary transformation of (3)-(4) to the permutation problem and procedure for designing the fitness value of the considered solution based on the greedy approach.

6 Multipurpose Cloud Center (mc²)

In May 2010, the Faculty of Electronics and IBM Corporation prepared and interesting offer of internship for students of PWr.
Multipurpose Cloud Center (mc^2) is the internship based on the newest IT solutions, which, apart from its educational character in the informatics area, gives the possibility to work from home in the flexible working hours (task work) and what is the most important, they gives the opportunity to prove oneself and demonstrate one's creativity.

To attending the internship, there are invited students especially from the 6th semester of Bachelor degree from any faculties and departments. Making such internship is necessary, according to requirements of the particular act, to graduation of Bachelor studies in all Polish technical Universities.
Internship lasts 4 weeks and they were conducted during three holiday's months: from July to September. It had a form compatible with IBM initiative, called Multipurpose Cloud Center (mc^2), similar to ESI (Educational Students Internship). In every period, almost 300 hundred students could have taken part in the event and every ten of them had been looked after by one mentor from IBM. Apart from that, the University of Technology in Wroclaw selected several tutors for every 30 of trainees.

27 July 2010 IBM Poland and the Wroclaw University of Technology announced signing an agreement concerning the first Multipurpose Cloud Computing Center in Poland.

Presently, in the Cloud, there are 10 modules divided by different scientific areas, from which trainees may choose those that are the most interesting for them. The possibilities of this technology will be consecutively increased in the next years.
The environment aimed to help in organization of remote internship currently enables an access to various numbers of practical workshops for students. The students might be divided into groups, every of which may have opened different combination of workshops, adjusted to the particular group's profile or interests of its members.

Students receive all essential information about internship through the special Web Site. With its help, it is possible for them to: reserve application, check the schedule, notes, and attached materials as well as to obtain current information concerning internship and tasks that should be executed. There is also planned to create a chat module, so that instead communicating via e-mail, students could also discuss with their mentors in the real time and keep conversation history in the same place. Such system facilitates internship management from the level of usual web browser.

Creating user account, rating their work, checking the schedule and statistics will be possible from an administration panel. The employees of the University might have the access to the various numbers of workshops using voucher system. For popularizing virtual machines on the server, there will be responsible Cloud Management System.

The newest environment for internship management will enable using composite systems consisting of several virtual machines with the requirement of constant network communication. Such solution will give the students possibility to implement multiple tasks basing on network services. In the consequence, there will be possible to prepare advanced topologies and popularize them via Cloud Technology.

7 Future Plans and Conclusions

Cloud computing is a useful tool in the modern Information Technology. In this paper we consider a new model for tasks scheduling of VM deployment for business and educational purposes.

In addition to the Multipurpose Cloud Computing Centre, IBM has invited Wroclaw University of Technology to join its two centers of Cloud excellence. The first is IBM's Academic Skills Cloud in which twenty (Q4 2010) academic units throughout the world collaborate to form a large educational Cloud project. The second is an association of elite higher education institutions working in the Cloud Computing technologies called the IBM Cloud Academy.

Wroclaw University of Technology and IBM Poland plan to continue collaboration in the field of education, PHD programs and research. IBM plans to initiate similar programs with other institutes, faculties and data centers in Poland, as well as to integrate with local business communities in the process.

World of Innovation – this is the title of the conference, concerning the latest achievements in the IT field, which took place 3rd April, 2012, in the Wroclaw University of Technology (PWr) and was organized in cooperation with IBM Global Services Delivery Centre in Wroclaw.

During the conference, Wroclaw University of Technology and IBM Haifa Research Laboratory signed the Declaration of Intent, concerning future cooperation. Collaboration will concern creating several laboratory groups for research and development of the latest IT technologies, in the field of Cloud Computing, IT modeling, as well as optimization problems.

References

1. Brzozowska, A., Greblicki, J., Kotowski, J.: Cloud Computing in Educational Applications Methods of Virtual Desktops Deployment. In: Moreno-Díaz, R., Pichler, F., Quesada-Arencibia, A. (eds.) EUROCAST 2011, Part I. LNCS, vol. 6927, pp. 568–575. Springer, Heidelberg (2012)
2. Cagan, J., Shimada, K., Yin, S.: A survey of computational approaches to three-dimensional layout problems. Computer Aided Design 34(8), 597–611 (2002)
3. Cui, Y., Yang, Y., Cheng, X., Song, P.: A recursive branch-and-bound algorithm for the rectangular guillotine strip packing problem. Computers & Operations Research 35, 1281–1291 (2008)
4. Folino, G., Mastroianni, C.: Special section: Bio-inspired algorithms for distributed systems. Future Generation Computer Systems 26, 835–837 (2010)
5. Garg, S.K., Yeo, C.S., Anandasivam, A., Buyya, R.: Environment-conscious scheduling of HPC applications on distributed Cloud-oriented data centers. J. Parallel Distrib. Comput. (2010), doi:10.1016/j.jpdc.2010.04.004
6. Georgis, N., Petrou, M., Kittler, J.: On the generalized stock–cutting problem. Machine Vision and Applications 11, 231–241 (2000)
7. Greblicki, J., Kotowski, J., Szlachcic, E.: The Cutting Stock Problem Recent Application in Information Technology. In: Moreno-Díaz, R., Pichler, F., Quesada-Arencibia, A. (eds.) EUROCAST 2011, Part II. LNCS, vol. 6928, pp. 577–584. Springer, Heidelberg (2012)
8. Greblicki, J., Kotowski, J.: Analysis of the properties of the Harmony Search Algorithm carried out on the one dimensional binary knapsack problem. In: Moreno-Díaz, R., Pichler, F., Quesada-Arencibia, A. (eds.) EUROCAST 2009. LNCS, vol. 5717, pp. 697–704. Springer, Heidelberg (2009)
9. Greblicki, J., Kotowski, J.: Automated design of totally self-checking sequential circuits. In: Moreno-Díaz, R., Pichler, F., Quesada-Arencibia, A. (eds.) EUROCAST 2009. LNCS, vol. 5717, pp. 98–105. Springer, Heidelberg (2009)
10. Kong, X., Lin, C., Jiang, Y., Yan, W., Chu, X.: Efficient dynamic task scheduling in virtualized data centers with fuzzy prediction. Journal of Network and Computer Applications (2010)
11. Kotowski, J.: The use of the method of illusion to optimizing the simple cutting stock problem. In: Proc. of 7th IEEE Conference on Methods and Models in Automation and Robotics, MMAR 2001, vol. 1, pp. 149–154 (2001)
12. Lee, Y.C., Zomaya, A.Y.: Rescheduling for reliable job completion with the support of clouds. Future Generation Computer Systems 26, 1192–1199 (2010)
13. Mateescu, G., Gentzsch, W., Ribbens, C.J.: Hybrid Computing - Where HPC meets grid and Cloud Computing. Future Generation Computer Systems (2010)
14. Ortmann, F.G., Ntene, N., van Vuuren, J.H.: New and improved level heuristics for the rectangular strip packing and variable-sized bin packing problems. European Journal of Operational Research 203, 306–315 (2010)
15. Paul, S., Pan, J., Jain, R.: Architectures for the future networks and the next generation Internet: A survey. Computer Communications 34, 2–42 (2011)
16. Sultan, N.: Cloud computing for education: A new dawn? International Journal of Information Management 30, 109–116 (2010)
17. Taylor, M., Haggerty, J., Gresty, D., Hegarty, R.: Digital evidence in cloud computing systems. Computer Law & Security Review 26, 304–308 (2010)
18. Trkman, P., Gradisar, M.: One-dimensional cutting stock optimization in consecutive time periods. European Journal of Operational Research 179, 291–301 (2007)

19. Wäscher, G., Haußner, H., Schumann, H.: An improved typology of cutting and packing problems. European Journal of Operational Research 183, 1109–1130 (2007)
20. Zhao, C., Zhang, S., Liu, Q., Xie, J., Hu, J.: Independent Tasks Scheduling Based on Genetic Algorithm in Cloud Computing. In: 5th International Conference on Wireless Communications, Networking and Mobile Computing, WiCom. IEEE (2009)