REVIEW OF PRACTICAL ISSUES OF RESOURCE & RISK MANAGEMENT IN CLOUD COMPUTING

Manoj Kumar*

Abstract: *Computers have become an indispensable part of life. We need computers everywhere, be it for work, research or in any such field. As the use of computers in our day-to-day life increases, the computing resources that we need also go up. For companies like Google and Microsoft, harnessing the resources as and when they need it is not a problem. But when it comes to smaller enterprises, affordability becomes a huge factor. With the huge infrastructure come problems like machines failure, hard drive crashes, software bugs, etc. This might be a big headache for such a community. Cloud Computing offers a solution to this situation.*

In this paper, we get rid of several foundation challenges in building a cloud-scale resource management system based on past research and shipping cluster resource management products. Additionally, we converse various techniques to grip these challenges, along with the pros and cons of each technique. We expect to stimulate future research in this area to extend practical solutions to these issues.

Keywords – *Clouding Computing, Resource Management, Practical issues, security issue*

*Associate Professor, Department of Computer Sc. & Engg. Green Hills Engineering College, Kumarhatti, Solan

I. INTRODUCTION

The Greek myths tell of creatures plucked from the surface of the Earth and enshrined as constellations in the night sky. Something similar is happening today in the world of computing. Data and programs are being swept up from desktop PCs and corporate server rooms and installed in "the compute cloud". In general, there is a shift in the geography of computation.

What is cloud computing exactly? As a beginning here is a definition "An emerging computer paradigm where data and services reside in massively scalable data centers in the cloud and can be accessed from any connected devices over the internet"

Like other definitions of topics like these, an understanding of the term cloud computing requires an understanding of various other terms which are closely related to this. While there is a lack of precise scientific definitions for many of these terms, general definitions can be given.

Cloud computing is an emerging paradigm in the computer industry where the computing is moved to a cloud of computers. It has become one of the buzz words of the industry. The core concept of cloud computing is, quite simply, that the vast computing resources that we need will reside somewhere out there in the cloud of computers and we'll connect to them and use them as and when needed.

Computing can be described as any activity of using and/or developing computer hardware and software. It includes everything that sits in the bottom layer, i.e. everything from raw compute power to storage capabilities. Cloud computing ties together all these entities and delivers them as a single integrated entity under its own sophisticated management.

Cloud is a term used as a metaphor for the wide area networks (like internet)or any such large networked environment. It came partly from the cloud-like symbol used to represent the complexities of the networks in the schematic diagrams. It represents all the complexities of the network which may include everything from cables, routers, servers, data centers and all such other devices.

Computing started off with the mainframe era. There were big mainframes and everyone connected to them via "dumb" terminals. This old model of business computing was frustrating for the people sitting at the dumb terminals because they could do only what they were "authorized" to do. They were dependent on the computer administrators to give them permission or to fix their problems. They had no way of staying up to the latest innovations

Resource management is a foundation job required of any man-made system. It affects the three fundamental criteria for system evaluation as given below:

1. Performance
2. Functionality
3. Cost

Inefficient resource management has a unswerving unconstructive consequence on performance and cost. It can also indirectly influence system functionality. Some functions the system provides might become too costly or unproductive due to pitiable performance. A cloud computing infrastructure is a multifaceted system with a huge number of collective resources. These are subject to unpredictable requests and can be affected by exterior events beyond user control. The cloud resource management requires multifarious policies and decisions for multi-objective optimization. It is enormously challenging because of the complication of the system, which makes it impracticable to have precise comprehensive state information.

II. POLICIES AND MECHANISMS

A policy typically refers to the principal guiding decisions, whereas mechanisms represent the means to implement policies. Separating policies from mechanisms is a guiding principle in computer science. Butler Lampson and Per Brinch Hansen offer solid arguments for this separation in the context of OS design. User can loosely group cloud resource management policies into five classes:

The explicit goal of an admission control policy is to prevent the system from accepting workloads in violation of high-level system policies. For example, a system may not accept an additional workload that would prevent it from completing work already in progress or contracted. Limiting the workload requires some knowledge of the global system state. In a dynamic system, this information is often obsolete at best. The capacity allocation means allocating resources for individual instances. An instance is service activation. Locating resources that are subject to multiple global optimization constraints requires user to a search a large space when the state of individual systems is changing so rapidly. User can perform load balancing and energy optimization locally, but global load-balancing and energy-optimization policies encounter the same difficulties as the ones already discussed. Load balancing and energy optimization are correlated and affect the cost of providing the services.

The common meaning of the term load balancing is that of evenly distributing the load to a set of servers. In cloud computing, a critical goal is minimizing the cost of providing the service. In particular, this also means minimizing energy consumption. This leads to a different meaning of the term load balancing. Instead of having the load evenly distributed among all servers, we want to concentrate it and use the smallest number of servers while switching the others to standby mode, a state in which a server uses less energy. In our example, the load from D will migrate to A and the load from C will migrate to B. Thus, A and B will be loaded at full capacity, whereas C and D will be switched to standby mode. Quality of service is that aspect of resource management that's probably the most difficult to address and, at the same time, possibly the most critical to the future of cloud computing. Resource management strategies often jointly target performance and power consumption.

Dynamic voltage and frequency scaling (DVFS) techniques such as Intel SpeedStep and AMD PowerNow lower the voltage and the frequency to decrease power consumption. Motivated initially by the need to save power for mobile devices, these techniques have migrated to virtually all processors, including those used in high-performance servers. As a result of lower voltages and frequencies, the processor performance decreases. However, it does so at a substantially slower rate than the energy consumption. Virtually all optimal or near-optimal mechanisms to address the five policy classes don't scale up. They typically target a single aspect of resource management, such as admission control, but ignore energy conservation. Many require complex computations that can't be done effectively in the time available to respond. Performance models are complex, analytical solutions are intractable, and the monitoring systems used to gather state information for these models can be too intrusive and unable to provide accurate data.

Therefore, many techniques are concentrated on system performance in terms of throughput and time in system. They rarely include energy tradeoffs or QoS guarantees. Some techniques are based on unrealistic assumptions. For example, capacity allocation is viewed as an optimization problem, but under the assumption that servers are protected from overload.

III. RELATED WORK

Kandalintsev et al. (2012) stated that software methods did not have control over low-level hardware circuit modules. Built-in hardware methods had very fine-grained control, but their impact was limited to a specific microchip unit. In this study they seemed to address this problem by developing algorithms that improve interoperability and combine the benefits of both software and hardware approaches delivering significant energy savings.

Rathore et al. (2011) stated that In case of the High Performance Computing (HPC), providing adequate resources for user applications was crucial. For instance, a computing center that a user has access to cannot handle the user applications with short deadlines due to limited computing infrastructure in the center. Therefore, to get the application completed by the deadline, users usually tried to get access to several computing centers (resources). However, managing several resources, potentially with different architectures, was difficult for users. Another difficulty was optimally scheduling applications in such environment. In this paper we were giving the strategy how the resource managed in cloud environment.

Irwin et al. (2010) argued that the cloud paradigm was also well suited for handling data-intensive applications, characterized by the processing and storage of data produced by high-bandwidth sensors or streaming applications. The data rates and the processing demands varied over time for many such applications, making the on-demand cloud paradigm a good match for their needs. However, today's cloud platforms needed to evolve to meet the storage, communication, and processing demands of data-intensive applications. We presented an ongoing GENI project to connect high-bandwidth radar sensor networks with computational and storage resources in the cloud and used this example to highlight the opportunities and challenges in designing end-to-end data-intensive cloud systems.

Gulati et al. (2011) shed light on some of the key issues in building cloud-scale resource management systems, based on five years of research and shipping cluster resource management products. Furthermore, they discussed various techniques to provide large scale resource management, along with the pros and cons of each technique. they hoped to motivate future research in this area to develop practical solutions to these issues.

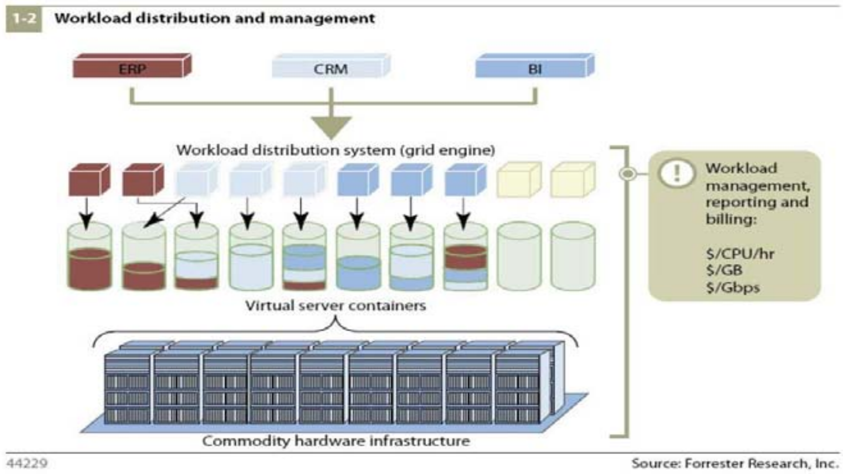
Hu et al. (2010) argued that the resource provisioning for cloud computing, an important issue was how resources may be allocated to an application mix such that the service level agreements (SLAs) of all applications are met. A performance model with two interactive job classes was used to determine the smallest number of servers required to meet the SLAs of both classes. For each class, the SLA is specified by the relationship: Prob [response time ≤ x] ≥ y. Two server allocation strategies are considered: shared allocation (SA) and dedicated allocation (DA). For the case of FCFS scheduling, analytic results for response time distribution were used to develop a heuristic algorithm that determined an allocation strategy (SA or DA) that required the smallest number of servers. The effectiveness of this algorithm was evaluated over a range of operating conditions. The performance of SA with non-FCFS scheduling was also investigated. Among the scheduling disciplines considered, a new discipline called probability dependent priority was found to have the best performance in terms of requiring the smallest number of servers.

Sasitharagai et al. (2013) stated that the problem of dynamic resource management for a large-scale cloud environment was mitigated with optimized high throughput performance. The resource management framework consisted of, Gossip protocol that ensured fair resource allocation among sites by calculating Memory Load Factor and CPU Load Factor and routing table for dynamically managing the tasks. A request partitioning approach based on gossip protocol was proposed that facilitates the cost-efficient and online splitting of user requests among eligible Cloud Service Providers within a networked cloud environment. Following the outcome of the request partitioning phase, the embedding phase - where the actual mapping of requested virtual to physical resources was performed that allows for efficient and balanced allocation of cloud resources. Finally, a thorough evaluation of the overall framework on a simulated cloud environment was made, which offers reliable and dynamic resource management.

Buyya et al. (2010) promised to offer subscription-oriented, enterprise-quality computing services to users worldwide. With the increased demand for delivering services to a large number of users, they needed to offer differentiated services to users and meet their quality expectations. Existing resource management systems in data centers are yet to support

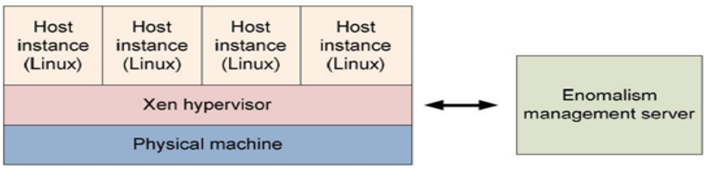
Service Level Agreement (SLA)-oriented resource allocation, and thus needed to be enhanced to realize cloud computing and utility computing. In addition, no work had been done to collectively incorporate customer-driven service management, computational risk management, and autonomic resource management into a market-based resource management system to target the rapidly changing enterprise requirements of Cloud computing. This paper presented vision, challenges, and architectural elements of SLA-oriented resource management. The proposed architecture supports integration of marketbased provisioning policies and virtualisation technologies for flexible allocation of resources to applications. The performance results obtained from our working prototype system shows the feasibility and effectiveness of SLA-based resource provisioning in Clouds.

IV. SYSTEM MODEL



This gives the basic architecture of a cloud computing application. We know that cloud computing is the shift of computing to a host of hardware infrastructure that is distributed in the cloud. The commodity hardware infrastructure consists of the various low cost data servers that are connected to the system and provide their storage and processing and other computing resources to the application. Cloud computing involves running applications on virtual servers that are allocated on this distributed hardware infrastructure available in the cloud. These virtual servers are made in such a way that the different service level agreements and reliability issues are met. There may be multiple instances of the same virtual server accessing the different parts of the hardware infrastructure available. This is to make sure that thereare multiple copies of the applications which are ready to take over on another one's failure. The virtual server distributes the processing between the infrastructure and the computing is done and the result returned. There will be a workload distribution management system, also known as the grid engine, for managing the different requests coming to the virtual servers. This engine will take care of the creation of multiple copies and also the preservation of integrity of the data that is stored in the infrastructure. This will also adjust itself such that even on heavier load, the processing is completed as per the requirements. The different workload management systems are hidden from the users. For the user, the processing is done and the result is obtained. There is no question of where it was done and how it was done. The users are billed based on the usage of the system - as said before - the commodity is now cycles and bytes. The billing is usually on the basis of usage per CPU per hour or GB data transfer per hour.

Server Architecture



Cloud computing makes use of a large physical resource pool in the cloud. As said above, cloud computing services and applications make use of virtual server instances built upon this resource pool. There are two applications which help in managing the server instances, the resources and also the management of the resources by these virtual server instances. One of these is the Xen hypervisor which provides an abstraction layer between the hardware and the virtual OS so that the distribution of the resources and the processing is well managed. Another application that is widely used is the Enomalism server management system which is used for management of the infrastructure platform.

When Xen is used for virtualization of the servers over the infrastructure, a thin software layer known as the Xen hypervisor is inserted between the server's hardware and the operating system. This provides an abstraction layer that allows each physical server to run one or more "virtual servers," effectively decoupling the operating system and its applications from the underlying physical server. The Xenhypervisor is a unique open source technology, developed collaboratively by the Xencommunity and engineers at over 20 of the most innovative data center solutionvendors, including AMD, Cisco, Dell, HP, IBM, Intel, Mellanox, Network Appliance, Novell, Red Hat, SGI, Sun, Unisys, Veritas, Voltaire, and Citrix. Xen is licensed under the GNU General Public License (GPL2) and is available at no charge in both source and object format. The Xen hypervisor is also exceptionally lean-- less than50,000 lines of code. That translates to extremely low overhead and near-native performance for guests. Xen re-uses existing device drivers (both closed and open source) from Linux, making device management easy. Moreover Xen is robust to device driver failure and protects both guests and the hypervisor from faulty or malicious drivers.

The Enomalism virtualized server management system is a complete virtual server infrastructure platform. Enomalism helps in an effective management of the resources. Enomalism can be used to tap into the cloud just as you would into a remote server. It brings together all the features such as deployment planning, load balancing, resource monitoring, etc. Enomalism is an open source application. It has a very simple and easy to use web based user interface. It has a module architecture which allows for the creation of additional system add-ons and plugins. It supports one click deployment of distributed or replicated applications on a global basis. It supports the management of various virtual environments including KVM/Qemu, Amazon EC2 and Xen, OpenVZ, Linux Containers, VirtualBox. It has fine grained user permissions and access privileges.

V. RISKS IN CLOUD COMPUTING

The security risks in cloud computing must be identified by the company in order to get a clear picture about the proper internal controls and related responses that a company should take to ensure the continued smooth operation of the company without fear of data disruption or compromise. Cloud computing is now an accepted part of the array of technology available to accountants. Cloud computing can offer efficiency and cost cutting benefits. Before using cloud technology, however companies should understand the risks and security issues inherent in this new technology. By taking a systematic approach to risk assessment, including creating effective policies for cloud usage and a risk response plan, companies can take advantage of this new technology to increase operational efficiency. All organizations should have policies to establish controls to prevent and detect the unauthorized procurement and use of cloud services, regardless of management's position on venturing into cloud computing. Due to the low cost of initiating cloud services relative to traditional technology purchases, current controls such as expenditure limits may not trigger appropriate attention from management.

Deciding whether to adopt cloud computing requires management to evaluate the internal environment – including the state of business operations, process standardization, IT costs, and the backlog of IT projects – along with the external environment – which includes laws and regulations and the competition's adoption of cloud computing. As management contemplates its cloud computing position and strategies, it should address some key questions, including:

- What is management's stance on outsourcing functions?
- Does the organization anticipate rapid growth that might require using cloud solutions?
- Is the organization in a mature market that might require using cloud computing to save costs to remain competitive?
- Are the organization's operational functions and processes mature and formalized enough to allow for a change in the underlying technology platform?
- What is the capability and maturity of the organization's current IT function?
- How should the organization prepare for cloud computing?
- Should cloud computing be embraced, to capitalize on its benefits, or rejected, to avoid risks such as data breaches or noncompliance with complex e-discovery requirements?
- Who should be involved in the evaluation process, and who makes the decisions?
- How can the organization manage its risks adequately while operating in a business environment with cloud computing?

The variables to be considered when making decisions about cloud computing solutions include business processes to be supported, specific deployment models, specific service delivery models, and the specific vendors that could become service providers.

VI. CONCLUSION

Efficient management of resources at cloud scale while providing proper performance isolation, higher consolidation and elastic use of underlying hardware resources is key to a successful cloud deployment. Existing approaches either provide poor management controls, or low consolidation ratios, or do not scale well. Based on years of experience shipping the VMware DRS resource management solution and prototypes to increase its scale, we have presented some use cases for powerful controls, key challenges in providing those controls at large scale, and an initial taxonomy of techniques available to do so.

REFERENCES

[1]. B. Alexander, "Web 2.0: A New Wave of Innovation for Teaching and Learning?" Learning, vol. 41, no. 2, pp. 32–44, 2006.

[2]. R. Buyya, C. Yeo, and S. Venugopal, "Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities," in Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications (HPCC-08, IEEE CS Press, Los Alamitos, CA, USA), 2008.

[3]. I. Foster, Y. Zhao, "Cloud Computing and Grid Computing 360-Degree Compared," in Grid Computing Environments Workshop, 2008. GCE'08, 2008, pp. 1–10.

[4]. W. Forrest, "How to cut data centre carbon emissions?" Website, December2008. [Online]. Zvailable:http://www.computerweekly.com/Articles/008/12/05/233748/how-tocut-data-centre-carbon-emissions.htm

[5]. J. Koomey, "Estimating total power consumption by servers in the US and the world," Final report. February, vol. 15, 2007.

[6]. L. Wang, G. von Laszewski, "Cloud computing: a perspective study," New Generation Computing, vol. WangLYAH, to appear in 2010.

[7]. G. von Laszewski, A. Younge, X., "Experiment and Workflow Management Using Cyberaide Shell," in 4th International Workshop on Workflow Systems in e-Science (WSES 09) in conjunction with 9th IEEE International Symposium on Cluster Computing and the Grid. IEEE, 2009. [Online].

[8]. G. von Laszewski, F. Wang, "Cyberaide JavaScript: A JavaScript Commodity Grid Kit," in GCE08 at SC'08. Austin, TX: IEEE, Nov. 16 2008

[9]. P. Barham, B. Dragovic, "Xen and the art of virtualization,"in Proceedings of the 19th ACM Symposium on OperatingSystems Principles, New York, U. S. A., Oct. 2003, pp. 164–177.

[10]. VMware, "Understanding Full Virtualization, Paravirtualization, andHardware Assis," VMware, Tech. Rep., 2007. [Online]. Available: http://www.vmware.com/files/pdf/paravirtualization.pdf

[11]. Amazon, "Elastic Compute Cloud." [Online]. Available: http://aws.amazon.com/ec2/

[12]. D. Nurmi, R. Wolski, C. Grzegorczyk, G. Obertelli, S. Soman, L.Youseff, and D. Zagorodnov, "The Eucalyptus Open-source Cloudcomputing System," Proceedings of Cloud Computing and Its Applications, 2008.

[13]. J. Fontan, T. Vazquez, "OpenNEbula: The Open Source Virtual Machine Manager for Cluster Computing," in Open Source Grid and Cluster Software Conference, San Francisco, CA, USA, May 2008.

[14]. S. Adabala, V. Chadha, "From virtualized resources to virtual computing Grids: the In VIGOsystem," Future Generation Comp. Syst., vol. 21, no. 6, pp.896–909, 2005