CHALLENGES IN FEDERATED CLOUD

Dr. C.S. Rajarajeswari
Department of Computer Science, IGCAS, Puducherry, India

ABSTRACT
Cloud computing refers to the use of resources residing on a remote machine, delivered to end user as a service on demand over Internet and pay per use. There are three models capable of delivering services namely Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). IaaS is the basic cloud service model which provides virtual infrastructure such as servers, data storage space etc. PaaS delivers user development environment services, whereby a user can develop and execute in-house built applications. The service includes Operating System (OS), programming language execution environment, databases and web servers. SaaS refers to the access of applications reside ng in the cloud. Cloud computing can be deployed as a public, private, community and hybrid cloud. Public clouds are available to all users in a pay-as-you-go manner. Private cloud refers to internal datacenters of an organization that are not available to the public. Community clouds are shared by several organizations and are usually built for specific requirements. Hybrid clouds are the mixture of the above three deployment models. In hybrid cloud, applications and data can be independently managed and allowed to move across the environment.
Index Terms: Cloud. Federated cloud, Resource provisioning, Service Level Agreement.

I. INTRODUCTION

Cloud Service Providers (CSP) offers many services that can extend benefit to its customers such as fast access of data from any location, scalability, pay-per-use, data storage, data recovery, protection against hackers, on-demand security controls and infrastructure facilities. CSP should ensure the security and also responsible for service infrastructure.

Single cloud model offers numerous advantages to both end users and business of all sizes. Some of the advantages are cost efficiency, availability, resiliency, redundancy, scalability, quick deployment, ease of integration, increased storage capacity, device diversity, location independence, backup and recovery. Some of the disadvantages in single cloud model are discussed below. (i) Security and Privacy - user gives data and information that might be sensitive and confidential. It is the responsibility of CSP to manage, protect and retain the information. Similarly, privacy in the cloud is another huge issue. Companies and users trust their provider to protect data from unauthorized users. (ii) Dependency and vendor lock-in - “vendor lock-in” is the implicit dependency on the provider. It is difficult to migrate from a provider once customers have committed with that provider. If a user wishes to switch to some other provider, it is really difficult to transfer huge data from the current provider to the new provider. (iii) Limited control and flexibility - Since the applications and services run on third party virtual environments, users have limited control over the function and execution of the hardware and software. (iv) Data Integrity - The data in the cloud may suffer damage during transition operation from or to the cloud provider. The risk of attacks from both inside and outside the cloud provider exists.

Moving from single cloud model to hybrid cloud model is reasonable and important in many aspects. Single cloud providers are predicted to be less popular due to the risks of service availability failure and the possibility of

malicious insiders. The main purpose of moving to hybrid cloud is to improve single cloud by distributing reliability, trust and security among multiple cloud providers. This distribution decreases the cost of switching and offers better fault tolerance. Therefore, the storage load spreads among several providers. Hybrid cloud is the use of two or more clouds to minimize the risk
of service availability failure, corruption of data, loss of privacy and vender lock-in. The service unavailability can occur due to breakdown of hardware, software or system infrastructure. A hybrid cloud strategy improves overall enterprise performance by using different infrastructures to meet the needs of customers. Hybrid cloud environment is capable of processing users demand, among several cloud resource infrastructures and dynamically balances the workload.

Hybrid cloud leads to inefficient performance for many reasons. (i) Management overhead – It requires a higher level of expertise in determining when to move from one cloud to another. This increases the overall management, including investments and monitoring. Dynamic expansion and resizing capability is possible in hybrid cloud model. Even though workload is manageable, the cost calculation leads to a serious problem in hybrid cloud. The cost calculation is based on the geographical location of data centers. This leads to unproductive result in hybrid cloud. (ii) Autonomy – The capability to deploy user's applications on different cloud providers has the clear advantage of reducing dependency on a single vendor. The ability to easily switch vendors means that the user can take advantage of the most attractive offers available at any given time.

A federated cloud is an amalgamation of several cloud providers that are brought together to meet an individual or business's needs [10,11]. The study of federated cloud computing is still in its start. Federated cloud model provides effective resource provisioning and on-time delivery of services. It is critical to predict the demands in the cloud environment. Even if the demand is unpredictable, federated cloud model intelligently undertake the decisions related to dynamic scaling or de-scaling of incoming service requests. Federated cloud not only offers the hardware, software and infrastructure, but can also steer traffic from different customers through the fastest possible parts of the network. Some clouds are best suited than others for a particular task. Different cloud providers support different platforms with constantly changing packages of capabilities. User might prefer to pay more for specific deployments with special capabilities, while continuing to take advantage of lower costs offered by a different provider where those capabilities are not relevant.

Some of the advantages of Federated Cloud model are (i) Cost-effectiveness- Federated Clouds provide a larger amount of resources, which helps to improve cost-effectiveness and quality. This includes improvement for both the user and provider such as reducing completion time, increasing system throughput and optimizing the resource utilization. (ii) Under-utilized- A cloud can decide to provide resources to other clouds when it realizes that its data center is under-utilized at a given time. (iii) Diverse geographical locations- Cloud service providers establish their data centers worldwide. Hence, there is a possibility of load sharing and performance improvement. (iv) Avoidance of vendor lock-in - Federated cloud can freely transit workload among service providers to avoid vendor lock-in. In case a provider changes a policy or pricing that impact negatively, clients can easily migrate to some other providers. (v) Better SLA to customers- Cloud provider can provide better Service Level Agreements (SLA) to its customers, as a result of competition. (vi) Guaranteed availability- During unexpected disasters, the cloud system is able to recover the services by federating with other service providers.

The main aim of federated cloud service brokering has received a lot of attention in academic and industry in recent years. Broker based federated cloud processing is an emerging concept in cloud based services. It connects a set of technologies, protocols and languages to communicate between customers and providers.

II. RELATED WORK

Federated cloud computing environment named Inter-Cloud [12] support the scaling of applications across multiple vendor clouds. The idea behind their introduced federation concept was to enhance cloud providers provisioning capabilities in case of sudden spikes in workload by leasing available computational and storage capabilities from other service providers. The main components of the proposed architecture are a cloud broker, a cloud exchange and a cloud coordinator. A client initiates a cloud broker in order to meet the specified QoS targets, whereas cloud coordinators, acting as gateway between their internal datacenters and external clouds, publish their services to the federation. Cloud exchange acts as a mediator bringing together service providers and customers. It aggregates infrastructure demands from the application brokers and matches them against the available resources published by the cloud coordinators. The framework still a research vision and its development was planned in context of the cloudbus project. However, the simulation results showed that the federation approach brings significant benefits to user's application performance. However, this vendor-oriented endeavor of Inter-Cloud has a specific control plane rather than a setting that it was based on future standards and open interfaces which are available to be shared in the academic community. In addition, knowledge sharing, experimentation and testing within their systems have been limited to the wide range of researchers.

Proposed frameworks [1] compare the performance of different cloud services such as Amazon EC2, Windows Azure and Rackspace. These works again focused on comparing the low level performance of cloud services such as CPU and network throughput. In this work, performance data are used to measure various QoS attributes and evaluate the relative ranking of cloud services.

III. CHALLENGES IN FEDERATED CLOUD

A. Resource Provisioning

The increasing demands in cloud computing has resulted in more heterogeneous infrastructure, making interoperability an area of concern. Due to this, it becomes a challenge for cloud customers to select appropriate cloud service provider (CSP) and hence it ties them to a particular CSP [5]. This is where intercloud computing comes into play. Although intercloud computing is still in its infancy, its purpose is to allow smooth interoperability between clouds, regardless of their underlying infrastructure. This allows users to migrate their workloads across clouds easily. Cloud brokerage is a promising aspect of intercloud computing.

Since there are many providers in the federated cloud management system, users get confusion in choosing the optimal service provider for their requirements [6]. Each provider provides variety of services to user requirements. Some providers offer best performance for certain types of parameters, while others for some other parameters. Each user has different types of input requirements. Hence it is necessary to calculate the rank of CSP in the federated cloud system. Existing ranking model in federated cloud system did not consider various performance parameters which are specific to cloud computing. Ranking model is proposed in the federated cloud architecture in which CSPs are ranked using a mathematical model. User suggested service provider is also taken into consideration in the resource allocation process. The objective is to allocate the incoming request to the available and eligible best CSP.

B. Resource Management

Resource management model, keeps in view different types of services, different customer types, customer characteristic, pricing and refunding. Service Level Agreement (SLA) is defined as a formal agreement between cloud service providers and customers [3]. Flexible and reliable management of SLA agreement is of ultimate importance for both cloud providers and consumers [4,7]. The general function of the SLA management is divided into three parts. They are SLA Negotiation, SLA Monitoring and SLA Termination

A. SLA Negotiation

In this layer a negotiation process is started by the cloud customer with the help of the broker in order to reach an SLA agreement between customers and cloud service providers.

B. SLA Monitoring

Based on the agreement between customer and service provider resources are provided to customers. Resources are provided to customers based upon the resource request given by a customer. During the resource management process a broker verifies whether any SLA violation occurs between customer and service provider. If any SLA violation occurs between them then the broker sends a SLA violation message to the service provider and the transaction dropped immediately between them.

C. SLA Termination

After completing the execution, customer sends a SLA termination request message to the cloud service provider. Based on the termination request given by the customer, cloud service provider terminates the communication process immediately and sends an acknowledgement termination message to the customer. SLA management (SLA negotiation, SLA monitoring and SLA termination) between cloud user and cloud service provider is a significant process.

C. Load Balancing

In federated cloud environment, it is common to have more than one provider for processing the incoming requests. In such cases, there must be a scheme required to distribute the incoming requests equally among CSPs. Load Balancing (LB) strategies are used to share the computational resources by transparently distributing the workload [2].

It optimize the request distribution based on the factors such as capacity, availability, response time, waiting time, execution time and current workload of service providers [8,9]. In the past few years, many studies have been showing the importance of LB to improve the performance of federated cloud. Incoming requests are distributed to CSP under specific load distribution and processed quickly for improving the quality. LB strategies may be either static or dynamic. Static method uses the information about the performance of providers and the transfer decisions are independent of the current state. Dynamic method uses the current state of CSP and makes load distribution decisions. The advantage of static model is its simplicity.

LB on the existing federated cloud is more time consuming. LB segment involves interaction between CSP for gathering load information, negotiation on load reallocation and migration of workload to different CSP. Traditionally LB approaches on federated cloud are implemented based on the historical information of service providers. The QoS of CSP is achieved using the LB algorithm. In this thesis, two dynamic LB algorithms are proposed to handle the workload among CSPs in the federated cloud. The objectives of the LB algorithms are to improve the efficiency in the federated cloud.

Generally LB algorithm has five major components namely Transfer policy, Selection policy, Location policy, Information policy and Load estimation policy. Transfer policy is to determine when a task should be transferred from one provider to the other. Selection policy focus on selecting the processor for load transfer so that the overall response time may be improved. Location policy determines the availability of required resources for providing services and makes a selection based on the location of resources. Information policy acquires workload related information such as nature of workload and average load of each service provider. It is also responsible for exchange of information and the amount of information to be exchanged. Load estimation policy determines the total workload of a service provider in a system.

IV. CONCLUSION

Cloud computing has become an important technology for outsourcing various resource needs of organizations. Single service providers could not offer quality services to user's requirements in dynamic environment. Single service providers are also lacking in service parameters like throughput, response time etc. when the workload becomes very high. Federated cloud mechanism helps to resolve these difficulties. Now-a-days there are many service providers providing services to user. Cloud users have several challenges for executing their task such as choosing the optimal cloud service provider, security, trust worthy of service provider, cost etc. Choosing the best service provider based on user requirement is very difficult task in federated cloud. Since cloud traffic is unpredictable and busty in nature, there is a possibility of large number of incoming service requests for processing. Hence the workload varies dynamically, some service providers are overloaded and others may be under loaded. In order to balance this situation, to improve the performance of federated cloud broker architecture, load balancing techniques are incorporated at the place of Broker Manager and brokers.

REFERENCES

[1] Amit Nathani, Sanjay Chaudharya and Gaurav Somani (2012), "Policy based resource allocation in IaaS cloud", Future Generation Computer Systems, vol. 28, no. 1, pp. 94–103.

[2] Anandharajan T R V and Bhagyaveni M A (2011)," Co-operative Scheduled Energy Aware Load-Balancing technique for an Efficient Computational Cloud", International Journal of Computer Science Issues, vol. 8, no. 2, pp. 571-576.

[3] Ardagna D, Casolari S, Colajanni M and Panicucci B (2012), "Dual time-scale distributed capacity allocation and load redirect algorithms for cloud systems", Journal of Parallel Distributed Computing, vol. 72, no. 6, pp. 796-808.

[4] Baomin Xu, Chunyan Zhao, Enzhao Hua and Bin Hu (2011), "Job scheduling algorithm based on Berger model in cloud environment", Journal of Advances in Engineering Software, vol. 42, no. 7, pp. 419–425.

[5] Bellur U, Rao C S, Madhu Kumar S.D (2010), "Optimal Placement Algorithms for Virtual Machines", Cornell University Library, pp 1-16.

[6] Benny Rochwerger, David Berltgand, Amir Epstein, David Hadas, Irit Loy, Kenneth Nagin, Johan Tordsson,Carmelo Ragusa, Massimo Villari, Stuart Clayman, Elizer Levy, Alessandro Maraschini, Philippe Massonet, Henar Munoz and Giovanni Toffetti (2011), "Reservoir – When One Cloud is not enough. Computer", vol. 44, no. 3, pp. 44-51.

[7] Bernstein D, Ludvigson E, Sankar K, Diamond S and Morrow M (2009), "Blueprint for the Intercloud –Protocols and Formats for Cloud Computing Interoperability", proceedings of the International Conference on Internet and Web Applications and Services. pp. 328-336.

[8] Bessis N, Sotiriadis S, Cristea V and Pop F (2012), "(in press-b) Meta-scheduling issues in interoperable HPCs, grids and clouds", International Journal of Web and Grid Services", vol. 8, no. 2, pp. 153-172.

[9] Bessis N, Sotiriadis S. Cristea V and Pop F (2012), "(in press-a) An architectural strategy for meta-scheduling in InterCloud", proceedings of the International Workshop on InterCloud and Collective Intelligence in conjunction with the IEEE International Conference on Advanced Information Networking and Applications, pp.1184-1189.

[10] Bilenko M, Mooney R, Cohen W, Ravikumar P and Fienberg S (2003), "Adaptive Name Matching in Information Integration," IEEE Intelligent Systems, vol.18, no. 5, pp. 16–23.

[11] Bin Dong, Xiuqiao Li, Qimeng Wu, Limin Xiao and Li Ruan (2012), "A dynamic and adaptive load balancing strategy for parallel file system with large-scale I/O servers", Journal of Parallel Distribution Computing, vol.72, no. 10, pp. 1254–1268. Buyya R., Ranjan R. and Calheiro R.N (2010), "InterCloud: Utility-Oriented Federation of Cloud Computing Environments for Scaling of Application Services", proceedings of the International Conference on Algorithms and Architectures for Parallel Processing, pp. 13–31.

[12] Buyya R., Ranjan R. and Calheiro R.N (2010), "InterCloud: Utility-Oriented Federation of Cloud Computing Environments for Scaling of Application Services", proceedings of the International Conference on Algorithms and Architectures for Parallel Processing, pp. 13–31.