EXPERIMENTALLY PROVE THE WORKLOAD COMPLEXITIES OF A LARGE SCALE UTILITY CLOUD

S.Muthukaruppan[1], R.Ebenezer Princy [2]

[1,2] Department of Comp. Science and Engg, Mount Zion College of Engg and Tech
Anna University, Chennai, (India)

ABSTRACT

In a cloud computing environment the analysis of characteristics and workload of data is purely difficult as well as it is really complicated to improve the resource management and operational conditions at the time of guaranteeing the Quality of Service (QOS) architecture. The workload models are analyzed based on the performance of cloud at the service end. This system basically operates with three different categories of actors, the Client, the data owner and the administrator. All the three actors are properly register their identity to the system and get the proper authenticated mean to prove their identity. Once the owner registers their identity, the system allows the owner to upload the required resource into the cloud server and all the data into the server are maintained in an encrypted strategy. The workload calculation begins at the client end, while the client start searching process the resource which is uploaded by the server getting ranked into the server, based on the ranking only the data will be organized and showed into the client port. Each and every time the workload of the uploaded resource can be viewed by the owner as well as the administrator. The Administrator has the full access rights of the entire system; the admin can easily activate and deactivate the owner and user based on their behavior as well as transactions. For all the entire system fully supports the efficient maintenance of resource in a cloud medium as well as calculating the workloads and provides the energetic support to the client for searching the data.

Keywords: Quality of Service (QOS), Cloud Management System (CMS), Key Performance Indicator (KPI), Cloud Providers.

1. INTRODUCTION

Cloud computing environments are large-scale heterogeneous systems that are required to meet Quality of Service requirements demanded by consumers in order to fulfil diverse business objectives. Such system characteristics result in a diversity of Cloud workload in terms of user behavior, task execution length and resource utilization patterns. In this context, Workload is defined as: "The amount of work assigned to, or done by, a client, workgroup, server, or system in a given time period" and consists of two components: tasks and users. Tasks are defined as the basic unit of computation assigned or performed in the Cloud, and a user is defined as the actor responsible for creating and configuring the volume of tasks to be computed. In order to further enhance the effectiveness of managing Cloud computing environments there are two critical requirements. The first is that such environments require extensive and continuous analyses in order to understand and quantify the characteristics of system components. The second is the exploitation of the parameters derived from such analyses in order to develop simulation models which accurately reflect the operational conditions.

Analysis and simulation of Cloud tasks and users significantly benefits both providers and researchers, as it enables a more in-depth understanding of the entire system as well as offering a practical way to improve data center functionality. For providers, it enables a method to enhance resource management mechanisms to effectively leverage the diversity of users and tasks to increase the productivity and Quality of Service of their systems. For example, exploiting task heterogeneity to reduce performance interference of physical servers or analyzing the correlation of failures to resource consumption. For researchers, simulation of Cloud workload enables evaluation of theoretical mechanisms supported by the characteristics of Cloud data centers. Ideally such simulation parameters are derived from the empirical analysis of large-scale production Cloud data centers. Failure to do so results in misleading assumptions about the degree of workload diversity that exists within the Cloud and the creation of unrealistic simulation parameters. This consequently results in limitations to the usefulness and accuracy of simulation parameters. However, deriving such analyses is challenging in two specific areas. The first and most critical problem is that there are few available data sources pertaining to large-scale production utility Clouds, due to business and confidentiality concerns. This is a particular challenge in academia, which relies on the very few publicly available Cloud trace logs. The second problem is analysis and simulation of realistic workloads; this is due to the massive size and complexity of data that a typical production Cloud can generate in terms of sheer volume of users and server events as well as recording resource utilization of tasks.

1.1 Empirical Analysis of Cloud Workflow

Recently, there has been initial work from the analysis of limited Cloud traces from Google and Yahoo! in an effort to provide mechanisms to analyze and characterize workload patterns. However, such efforts are predominately constrained to traces of short observational periods and coarse-grain statistics which are not sufficient to characterize the workload diversity of Cloud environments. In addition, there have been a number of approaches that analyze the diversity of workload by classifying tasks according to critical characteristics. However, none of these provide a comprehensive study of the diversity of users and tasks, or provide a model containing sufficient details about the model parameters obtained from the analyses in order to be of practical use to researchers.

The objective of this paper is to present an in-depth empirical analysis of workload and its diversity in a large scale production Cloud computing data center. Additionally, this work aims to provide a validated simulation model that includes parameters of tasks and users to be made available for other researchers to use. The analysis is conducted using the data from the second version of the Google Cloud trace log, which contains over 25 million tasks, submitted by 930 users over the observational period of a month. There are three core contributions within this work:

- An in-depth statistical analysis of the characteristics of workload diversity within a large-scale production Cloud. The analysis was performed over the entire trace log time span as well as a number of observational periods to investigate patterns of diversity for both users and tasks within the system.
- An extensive analysis of distribution parameters derived from the workload analysis that can be applied to simulation tools by other researchers.
- A comprehensive validation of the simulation model based on empirical and statistical methods. A significant contribution of the simulation model provided is that it does not just replay the data within the trace log. Instead, it creates patterns that randomly fluctuate based on realistic parameters. This is important in order to emulate dynamic environments and to avoid just statically reproducing the behavior from a specific period of time.

A secondary contribution of this paper is presenting practical applications of the model obtained to identify sources of inefficiencies and enhance resource-management and energy usage in virtualized Cloud environments. This approach applies the methodology of analysis introduced in our previous approach, but is substantially different in a number of ways. First, this paper focuses specifically on a substantial analysis of Cloud diversity for tasks and users. Additionally, we analyze the entire trace log time span and three additional observational periods, instead of just two days-which limited the original approach's applicability, as it could potentially omit crucial behavior within the overall Cloud environment. Furthermore, extensive analysis and parameter details are provided for user and task distributions.

1.2 Diversity Patterns in Cloud

According to the NIST, the Cloud computing model has the following five essential characteristics: on-demand self service, resource pooling, broad network access, rapid elasticity and measured service. These characteristics create highly dynamic environments where customers from different contexts coexist submitting workloads with diverse resource requirements at anytime. Workloads by themselves have properties or attributes that describe their behavior. These attributes are normally expressed by the type and amount of resources consumed and other attributes that could dictate where a specific workload can or cannot be executed. For example, security requirements, geographical location, or specific hardware constraints such as processor architecture, number of cores or Ethernet speed among others described. As more and more customers adopt Cloud platforms to fulfill their IT requirements, Cloud providers need to be prepared to manage highly heterogeneous workloads that are served on the top of shared infrastructure. Workloads can be broadly classified according to the fundamental resources that they consume in terms of CPU, memory and storage-bound workloads. Moreover, depending on the interaction with the end-users, they can also be classified as latency sensitive and batch workloads. Common examples of workloads running in multi-tenant Cloud data centers according to include Business Intelligence, scientific high-performance computing, gaming and simulation.

1.3 Importance of Workload Models in Cloud

Models abstract reality to aid researchers and providers in understanding system environments in order to develop or enhance such systems. Workload models enable a way to actually study Cloud environments and the effect of workload variability on the performance and productivity of the overall system. Specifically, they support researchers and providers in further understanding the actual status and conditions of the Cloud system and identify Key Performance Indicators (KPI) necessary to improve operational parameters. Such models can be used in a number of research domains including resource optimization, security, dependability and energy-efficiency. In order to produce realistic models, it is critical to derive their components and parameters from real-world production trace logs. This leads to capturing the intrinsic diversity and dynamism of all co-existing components within the system as well as their interactions. Moreover, realistic workload models enable the simulation of Cloud environments whilst being able to control selected variables to study emergent system-wide behavior, as well as support the estimation of accurate forecasting under dynamic system conditions to improve QoS offered to users. This supports the enhancement of Cloud Management Systems (CMSs) as it allows providers to experiment with hypothetical scenarios and assess their decisions as a result of changes within the Cloud environment (i.e., Capacity planning for increased system size, alteration of the workload scheduling algorithm, performance tradeoffs, and service pricing models.

II. RELATED WORK

Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system. The major part of the project development sector considers and fully survey all the required needs for developing the project. For every project Literature survey is the most important sector in software development process. Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system such as what type of operating system the project would require, and what are all the necessary software are needed to proceed with the next step such as developing the tools, and the associated operations.

2.1 Intercloud: Utility-Oriented Federation of Cloud Computing Environments for Scaling of Application Services

Cloud computing providers have setup several data centers at different geographical locations over the Internet in order to optimally serve needs of their customers around the world. However, existing systems do not support mechanisms and policies for dynamically coordinating load distribution among different Cloud-based data centers in order to determine optimal location for hosting application services to achieve reasonable QoS levels. Further, the Cloud computing providers are unable to predict geographic distribution of users consuming their services, hence the load coordination must happen automatically, and distribution of services must change in response to changes in the load. To counter this problem, we advocate creation of federated Cloud computing environment (InterCloud) that facilitates just-in-time, opportunistic, and scalable provisioning of application services, consistently achieving QoS targets under variable workload, resource and network conditions. The overall goal is to create a computing environment that supports dynamic expansion or contraction of capabilities

(VMs, services, storage, and database) for handling sudden variations in service demands. This paper presents vision, challenges, and architectural elements of Inter-Cloud for utility-oriented federation of Cloud computing environments. The proposed InterCloud environment supports scaling of applications across multiple vendor clouds. We have validated our approach by conducting a set of rigorous performance evaluation study using the Cloud Sim toolkit. The results demonstrate that federated Cloud computing model has immense potential as it offers significant performance gains as regards to response time and cost saving under dynamic workload scenarios.

2.2 Characterizing Task Usage Shapes in Google Compute Clusters

The increase in scale and complexity of large compute clusters motivates a need for representative workload benchmarks to evaluate the performance impact of system changes, so as to assist in designing better scheduling algorithms and in carrying out management activities. To achieve this goal, it is necessary to construct workload characterizations from which realistic performance benchmarks can be created. In this approach, we focus on characterizing run-time task resource usage for CPU, memory and disk. The goal is to find an accurate characterization that can faithfully reproduce the performance of historical workload traces in terms of key performance metrics, such as task wait time and machine resource utilization. Through experiments using workload traces from Google production clusters, we find that simply using the mean of task usage can generate synthetic workload traces that accurately reproduce resource utilizations and task waiting time. This seemingly surprising result can be justified by the fact that resource usage for CPU, memory and disk are relatively stable over time for the majority of the tasks. Our work not only presents a simple technique for constructing realistic workload benchmarks, but also provides insights into understanding workload performance in production compute clusters.

2.3 An Analysis of Traces from a Production Map reduce Cluster

Map reduce is a programming paradigm for parallel processing that is increasingly being used for data-intensive applications in cloud computing environments. An understanding of the characteristics of workloads running in Map reduce environments benefits both the service providers in the cloud and users: the service provider can use this knowledge to make better scheduling decisions, while the user can learn what aspects of their jobs impact performance. This paper analyzes 10-months of Map reduce logs from the M45 supercomputing cluster which Yahoo! made freely available to select universities for academic research. We characterize resource utilization patterns, job patterns, and sources of failures. We use an instance-based learning technique that exploits temporal locality to predict job completion times from historical data and identify potential performance problems in our dataset.

2.4 Towards Characterizing Cloud Backend Workloads: Insights from Google Compute Clusters

The advent of cloud computing promises highly available, efficient, and flexible computing services for applications such as web search, email, voice over IP, and web search alerts. Our experience at Google is that realizing the promises of cloud computing requires an extremely scalable backend consisting of many large compute clusters that are shared by application tasks with diverse service level requirements for throughput, latency, and jitter. These considerations impact (a) capacity planning to determine which machine resources must grow and by how much and (b) task scheduling to achieve high machine utilization and to meet service level objectives. Both capacity planning and task scheduling require a good understanding of task resource consumption (e.g., CPU and memory usage). This in turn demands simple and accurate approaches to workload classification-determining how to form groups of tasks (workloads) with similar resource demands. One approach to workload classification is to make each task its own workload. However, this approach scales poorly since tens of thousands of tasks execute daily on Google compute clusters. Another approach to workload classification is to view all tasks as belonging to a single workload. Unfortunately, applying such a coarse-grain workload classification to the diversity of tasks running on Google compute clusters results in large variances in predicted resource consumptions.

III. PROPOSED SCHEME

In the present system, a new approach called structured data maintenance with proper workload based schema is introduced, so the data maintenance pattern and retrieval of resource is simpler than the existing approaches. This system maintains the fixed standardization schema for maintaining the resource into the server, so that the resource is maintained into the server at some fixed measurement, with proper arrangement, it automatically cause the data retrieval so easier. Clients feel interesting to retrieve the required records without any unwanted delays and controversies found during search. Time consumption for data retrieval is really low in the Client end, so that the performance of the server is automatically high. The server maintenance is in a structured workload approach so that the system requires low cost to maintain and manipulate the resource into the server.

To experimentally analyze the present cloud schema with the all existing models and forming the new workload based creative architecture in Large-Scale cloud services. In the present system, a new approach called structured data maintenance with proper workload based schema is introduced, so the data maintenance pattern and retrieval of resource is simpler than the existing approaches. This system maintains the fixed standardization schema for maintaining the resource into the server, so that the resource is maintained into the server at some fixed measurement, with proper arrangement, it automatically cause the data retrieval so easier.

- Workload in Cloud data centers is determined not only by tasks characteristics but also by user behavioral models.
- User patterns tend to be significantly more diverse than task patterns across different observational periods, so it requires only low Cost to manipulate the data and it is comparatively good to extract the resource from the server.
- This type of Cloud is fully independent from the Third Party service providers like Google, yahoo and other mediums, because it operates with its own Mail Server.
- Stable approach is followed to maintain and handling the data, so the execution time is less and easy to maintain the data.
- File length and Execution time can be seen.

Resolving the Security Issues by means of maintain their identity as well as crypto text privacy.

IV. ARCHITECTURAL DESIGN

Generally algorithms shows a result for exploring a single thing that is either be a performance, or speed, or accuracy, and so on. In this protocol, a source can find trusted paths to a destination in a single route discovery round. New route discovery is needed only when all paths break or fail to meet the trust requirement. This protocol provides a flexible and feasible approach to choose a trusted path and the routing protocol is tested against denial of service attack. This system can be extended to defend against several attacks and finding secured paths. Again this can be extended to find shortest path from the list of multiple path. Multiple paths can also be used to balance load by forwarding data packets on multiple paths at same time. Many approaches are handled to solve these issues but the result creates a controversy at any case, because all the approaches basically follow the existing models only. There are no standardizations into the data maintenance schema so that the resource is maintained into the server at any dimension, without any proper arrangement, it automatically cause the data retrieval so difficult. Users feel tired to retrieve the unwanted records found during search for the specific record. Time consumption for data retrieval again creates a big issue in the user end, so that the performance of the server is automatically down. The server maintenance is not in a structured manner so that the system requires high cost to maintain and manipulate the resource into the server.

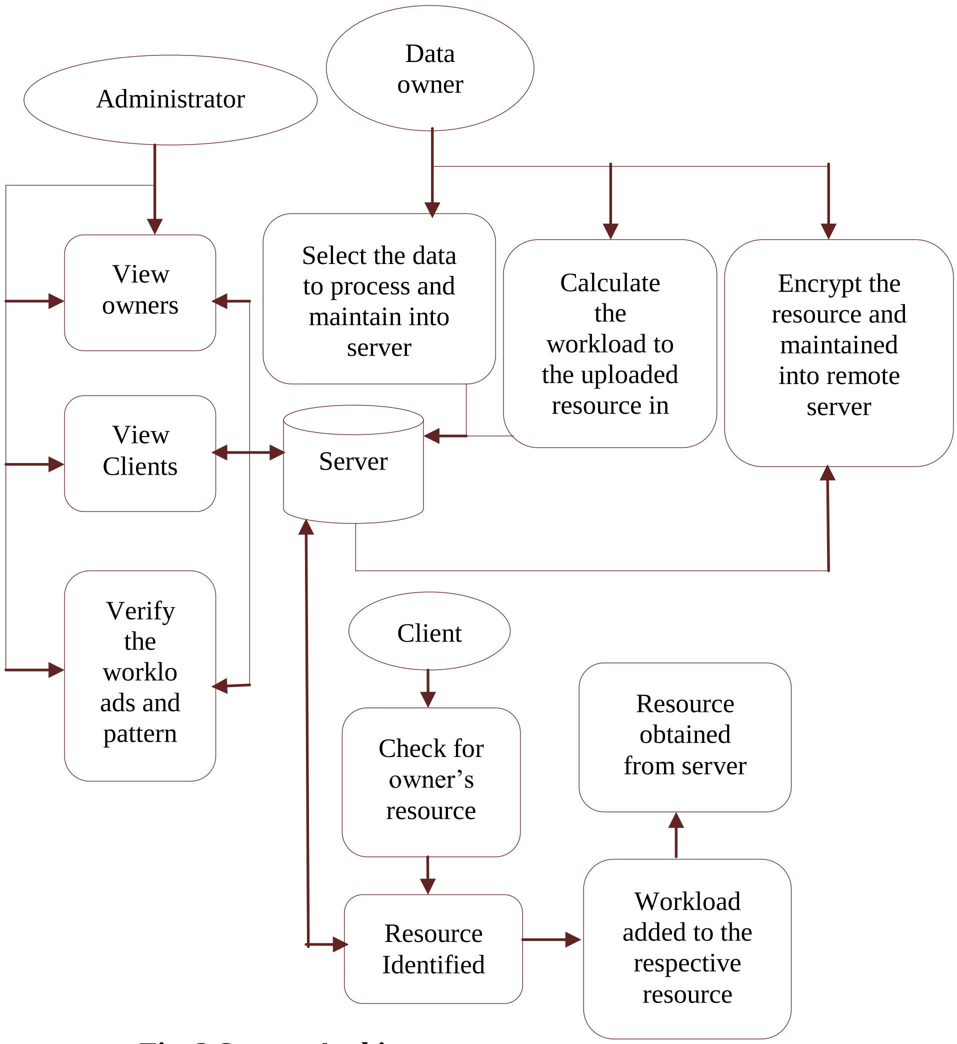

Fig. 3 System Architecture

V. METHODOLOGY

Following are the most frequently used project management methodologies in the project management practice:

1. Cloud service signing
2. Energetic signature generation
3. Workload computation
4. Pay-based service scheduling
5. Data searching process

5.1 Cloud service signing

In cloud service signing process there are two actors to participate they are owner and user .owner is a responsible for uploading the resource into the remote server until the data uploading limit is over. At the same time the users can view the files in the user has also registered in the cloud .the registered users only view the data. In the data owner has upload a three files in without cost after that uploading will be payable service. In the data owner has may upload the resource in every time to pay the cost of data in the server

5.2 Energetic Signature Generation

In the energetic signature generation is the security check for this process if the user can registered the server the security password has been generated through the mail port if the security key is used to register the user and the user can view the data in secure way. In the energetic signature generation algorithm is helps to generate the security check for the user mail id and it is helps to avoiding the unauthorized users can view the data. And the energetic signature generation algorithm helps to improving the security for the resources.

5.3 Workload Computation

In the workload computation process is used to calculate the file size. In the admin side it will be used to calculate the overall size of the uploaded data and how many files are uploaded in each and every owner are calculated and view by the admin. At the owner side calculate individual size of data and how many files can upload the server these things are view by the owner side .the user can view the current view of the data size.

5.4 Pay-Based Service Scheduling

In the data owner has upload the three resources in without cost until the owner may upload more than files it will pay based service because the fee of cost will be increasing unimportant data and large amount of data has been uploaded in the server so the data owner upload three files without cost in first time, next time the data owner used the pay based services

5.5 Data Searching Process

In the data owner upload the file completely, after that the user can search the data in the server the server has a many types of data what the user need data to search and find the data.

VI. CONCLUSION

The new system analyzes and enumerates the range of Cloud workloads and obtains a workload representation from a large scale construction Cloud data center. The obtainable investigation and representation captures the individuality and behavioral samples of user and task inconsistency across the whole system as well as different observational stages. There are number of things obtained via this proposed approach they are listed below: Workload in Cloud data centers is determined not only by tasks characteristics but also by user behavioral models, User patterns tend to be significantly more diverse than task patterns across different observational periods, so it requires only low Cost to manipulate the data and it is comparatively good to extract the resource from the server, this type of Cloud is fully independent from the Third Party service providers like Google, yahoo and other mediums, because it operates with its own Mail Server, Stable approach is followed to maintain and handling the data, so the execution time is less and easy to maintain the data. For all the entire system fully resolves and proves the proposed approach is efficient to maintain the resource in a cloud medium as well as calculating the workloads and provides the energetic support to the client for searching the data.

VII ACKNOWLEDGMENT

We would like to sincerely thank Assistant Prof. R.Ebenezer Princy for his advice and guidance at the start of this article. His guidance has also been essential during some steps of this article and his quick invaluable insights have always been very helpful. His hard working and passion for research also has set an example that we would like to follow. We really appreciate his interest and enthusiasm during this article. Finally we thank the Editor in chief, the Associate Editor and anonymous Referees for their comments.

REFERENCES

[1] R. Buyya, R. Ranjan, and R. N. Calheiros, “InterCloud: Utility-oriented federation of cloud computing environments for scaling of application services,” Proc. 10th Int. Conf. Algorithms Archit. Parallel Process., 2010, pp. 13–31.

[2] Google Cluster Data V1 (2010). [Online] Available: http://code.google.com/p/googleclusterdata/wiki/TraceVersion1

[3] Cluster Data V2 (2011). [Online] Available: http://code.google.com/p/googleclusterdata/wiki/ClusterData2011_1

[4] Yahoo. Yahoo! M45 Supercomputing Project. (2007). [Online].Available: http://research.yahoo.com/node/1884

[5] Q. Zhang, J. Hellerstein, and R. Boutaba, “Characterizing task usage shapes in Google compute clusters,” in Proc. 5th Int. Workshop Large Scale Distrib. Syst. Middleware, 2011, pp. 2–8.

[6] S. Kavulya, J. Tan, R. Gandhi, and P. Narasimhan, “An analysis of traces from a production Map Reduce cluster,” in Proc. IEEE/ACT Int. Conf. Cluster, Cloud Grid Comput., 2010, pp. 94–103.

[7] A. K. Mishra, J. Hellerstein, W. Cirne, and C. R. Das, “Towards characterizing cloud backend workloads: Insights from Google compute clusters,” ACM SIGMETRICS Perform. Eval. Rev., vol. 37, pp. 34–41, 2010.

[8] S. Aggarwal, S. Phadke, and M. Bhandarkar, “Characterization of Hadoop jobs are using unsupervised learning,” in Proc. 2nd Int. Conf. Cloud Comput. Technol. Sci., 2010, pp. 748–753.

AUTHORS PROFILE

S.Muthukaruppan is currently a PG scholar in Computer Science from the Department of Computer Science and Engineering at Mount Zion College of Engineering and Technology, Pudukottai and Tamilnadu. He received his Bachelor Degree in Information Technology from Sudarsan Engineering College, Pudukottai and Tamilnadu. His Research areas include Cloud Computing, Grid Computing and Distributed Computing.

R.Ebenezer Princy is currently working as a Assistant Professor from the Department of Computer Science and Engineering at Mount Zion College of Engineering and Technology, Pudukkottai and Tamilnadu. Her main research interests lie in the area of Cloud Computing and Wireless Sensor Networks.