Optimizing Cloud Analysis and Load Balancing for Efficient Social Network Services

Kondragunta Rama Krishnaiah, Professor, Department of Computer Science and Engineering, R K College of Engineering, Vijayawada - 521456, Andhra Pradesh, India, email: kondraguntark@gmail.com

Abstract

In practical applications, Social Network Analysis (SNA) plays a crucial role in understanding the structures that connect diverse individuals in interacting populations. This analytical approach has proven highly effective across various scientific fields. One area where it finds extensive use is in crowd sourcing, which involves gathering, integrating, and analyzing vast and diverse datasets generated from numerous sources in urban spaces, such as sensors, devices, vehicles, buildings, and human interactions. In today's world, urban emergency events can affect any country, community, or individual. As we rely more on cloud computing, a significant challenge arises in efficiently analyzing the data due to computational and communication imbalances among computers caused by the collective behavior of humanity. Traditional load balancing techniques require substantial effort to rebalance loads on the nodes and struggle to handle stragglers, making the resource allocation process even more challenging. To address these issues, it is essential to have a cloud computing simulator to test and validate proposed solutions before implementing them in real cloud infrastructures. Researchers have developed several cloud computing simulators for this purpose, which have been widely used within the research community. In our research, we have conducted a comprehensive survey of current cloud load balancing solutions. Based on our classification, these solutions can be grouped into three categories: General Algorithm-based, Architectural-based, and Artificial Intelligence-based load balancing mechanisms. Each category offers unique approaches to tackle the load balancing challenges in cloud computing. To evaluate these solutions effectively, we have identified suitable metrics and thoroughly analyzed their pros and cons. By doing so, we aim to contribute valuable insights into the effectiveness of different load balancing approaches in cloud computing scenarios. Ultimately, our research seeks to enhance the efficiency and performance of cloud infrastructures in handling complex and dynamic workloads.

Index Terms: Cloud Computing, Load Balancing, Distributed Systems, Virtual Machine.

1. INTRODUCTION

Social network analysis acts to give ranking scores and neighbors using social datasets. It gives the complete picture to understand human communities these days advance applications on social applications and models k-NN [1]. Proximity searches, Statistical classification, Recommendation systems, Internet marketing and etc. In Social network group of data is so large hence cloud data identification serving will done Cloud computation is convey by on parallel platform in cloud [2]. Crowd sourcing is also an emerging computing paradigm that tasks everyday mobile devices to form participatory sensor networks. It allows the increasing number of mobile phone users to share local knowledge acquired by their sensor enhanced devices to monitor pollution level or noise level, traffic condition [3]. The sensing data from volunteer contributors such as social network users can be further analyzed and processed, and leveraged in many areas such as environment monitoring, urban planning, emergency management, as well as public a popular Chinese micro blogging service similar to Twitter2 , has received much attention recently [4]. Take the widely used data set of Twitter web graph as an example, less than one percent of the vertices are adjacent to nearly half of all edges. It means that tasks hosting this small fraction of vertices may require many times more computation and

communication than average task does [5]. Cloud providers have to efficiently manage, provide, and allocate these resources to provide services to cloud consumers based on service level agreements (SLAs) which both sides agree to prior to the consumer using the services [6]. National Institute of Standards and Technology (NIST), a cloud ensures the five important features that are on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service [7]. Besides this, the goal is to provide on-demand computing services to cloud consumers with the guarantee of reliability, availability and scalability [8]. Cloud provides three basic service models termed as SaaS (Software as a Service), PaaS (Platform as a Service) and IaaS (Infrastructure as a Service) and is deployed in four ways known as Public, Private, Community and Hybrid Clouds. Fig. 1 summarizes the cloud services and deployment models along with some application domains that may consume the cloud resources [9]. To achieve these kinds of goals, improving the general performance of system maintain stability, availability and some other features for a cloud computing data center we need a mechanism which is called load balancing [10. Load balancing is one of the central issues and challenges in distributed systems like grid-based systems and cloud computing [11].

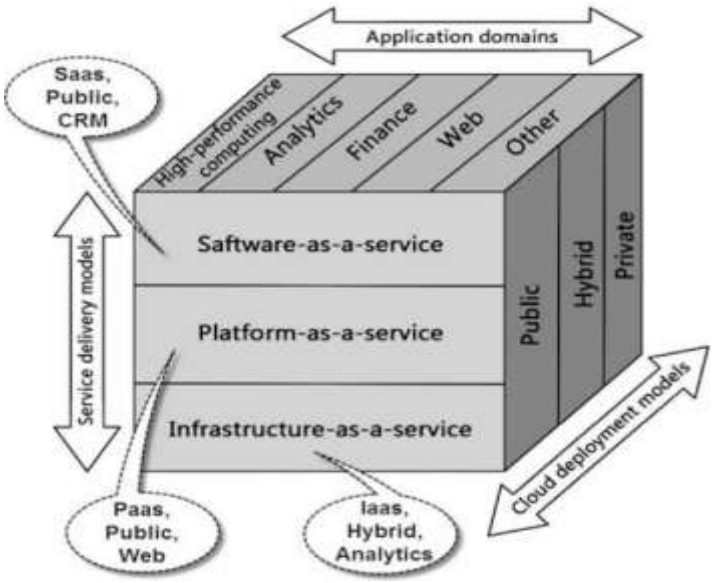

Fig. 1. Cloud computing services and deployment models

2. RELATED WORK

Advances in information technology and its widespread growth in several areas of business engineering, medical, and scientific studies are resulting in information explosion. Knowledge discovery and decision-making from such rapidly growing voluminous data are a challenging task in terms of data organization and processing [12], which is an emerging trend known as big data computing a new paradigm that combines large-scale compute, new data-intensive techniques, and mathematical models to build data analytics [13]. Big data computing demands a huge storage and computing for data duration and processing that could be delivered from on-premise or clouds infrastructures [14]. There are many kinds of load balancing mechanisms and approaches which most of the studies have been classified as two main categories static and dynamic. In static techniques there are usually prior knowledge about the global status of the system such as job resource requirements, communication time processing power of system nodes, memory and storage devices capacity [15]. A static approach is a kind of assignment from a set of tasks to a set of resources which can take either a deterministic form [16]. The performance analysis of a cloud computing system refers to evaluate it from different perspectives that include the assessment of the SLAs, resources

distribution and its complex path generation for service delivery, and secure cloud storage [16]. For evaluating the performance of a cloud computing system, one of the key performance requirements is to assure that it is a SLA- driven system performance [17].

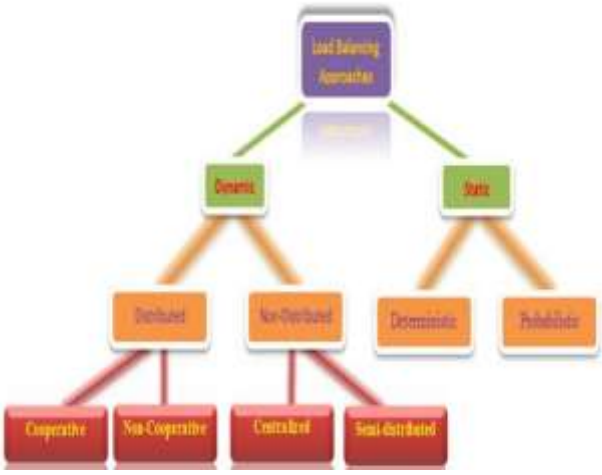

Fig.2. Load Balancing Approaches Classification.

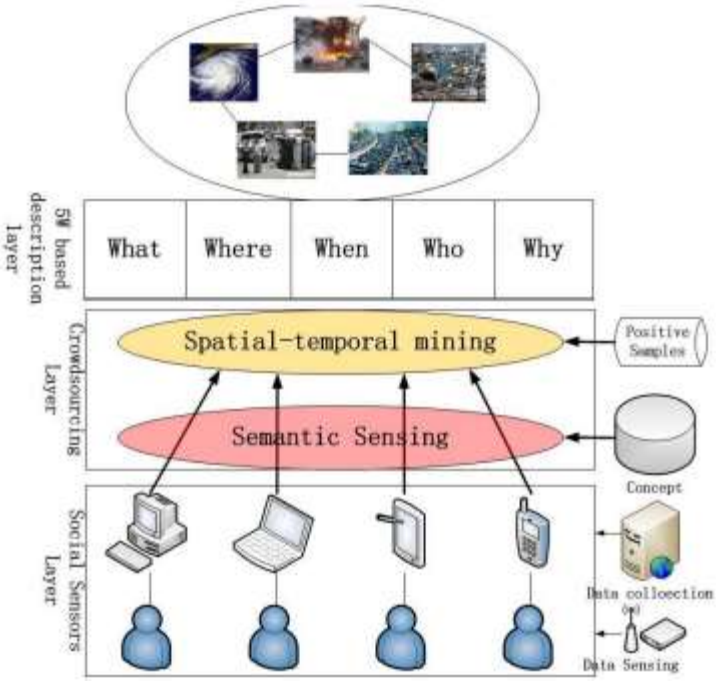

Fig. 3. The hierarchical structure of the proposed method.

4. SYSTEM MODEL

Crowd sourcing or participatory sensing may be a potential solution solving the description of urban emergency events. The proposed model aims at collecting and analyzing the information from social sensors. The social network can be seen as a sensor receiver. Usually, the social network users can be seen as social sensors [18]. The proposed model is set as a hierarchical data model including three different layers. The proposed model wants to collect the related data of urban emergency events the social network such as Weibo is sensor receiver. Usually social media provides API for downloading the real time data. In this layer basic elements of the proposed 5W model are extracted from the sensing data of the social sensors layer [19]. Knowledge base and positive samples of the urban

emergency event are implemented in this layer, which are used for improving the accuracy of this layer. In this layer, the detection and description of the urban emergency event is launched the spatial and temporal information of this event is also given. A GIS based description of the detected urban emergency event is shown [20].

5. PROPOSED SYSTEM

In this proposed system to register & login the data owners & migrate data from one to another cloud, & then it will check on which cloud the data has stored the virtual master has verify your data & find cloud efficiency based on the no attackers. The VM update the rank based while request is going to users, with down load the file [21]. Finally receive file information & view all data owners & end users. Particularly, FEP to form an effective preliminary step toward a purpose of desire on social network analysis accuracy actively contributes to the production of a result of straggling FEPs occupied into many gradual changes that lead toward a particular result of sub-processes later towards adaptation and spreads such sub-processes on systems [22]. In this category algorithms usually proposed explicitly and load balancing mechanism is represented through architecture components and the relations among them. This solution usually is an Architectural-based solution and for catching proposed goals, a special cloud computing architecture should be taken into consideration [23].

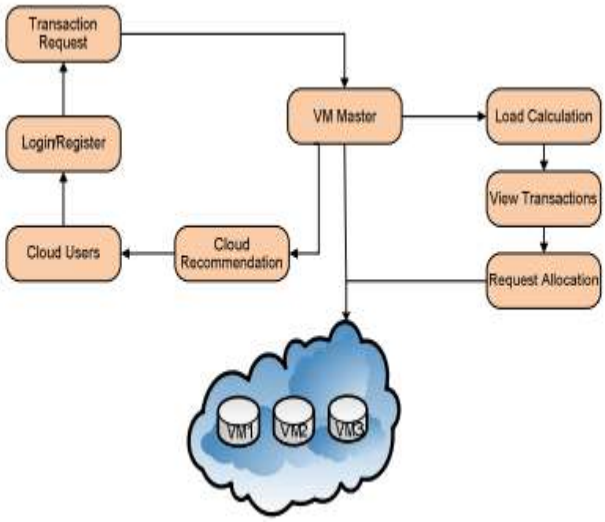

Fig. 4. Architecture of Proposed System

A. Dynamic Allocation Process

Allocation schemes provide cloud resources on the fly when the cloud user or application is requested, specifically to avoid over-utilization and under-utilization of resources. A possible drawback when needed resources are requested on the fly is that they might not be accessible [24]. The service supplier must allocate resources from different participating cloud data centers [25]. Resource allocation strategy (RAS) is related to combining cloud provider functions for utilizing and assigning scarce resources within the boundaries of the cloud system in order to suit the demand of the cloud application.

1) **Resource contention:** This situation occurs when multiple users and applications attempt to allocate the same resource simultaneously.
2) **Resource fragmentation:** This occurs when applications cannot allocate resources due to isolated resources being small items.

3) **Scarcity:** This occurs when multiple applications' requirements for the resources are high and there are limited resources, for example, requests for memory, I/O devices, CPUs, and the techniques that should serve that demand.
4) **Over provisioning:** This occurs when the users and applications obtain more resources than those that are requested to fit the quality of service (QoS) requirements.
5) **Under provisioning:** This occurs when the users and applications obtain fewer resources than those requested to fit the QoS requirements

B. D-Cloud

D-Cloud facilitates with a parallel software testing environment for reliable distributed systems that uses cloud computing technology and VMs with the facility of fault instillation [26]. D-Cloud supports the fault tolerance analysis related to the failures of hardware that happen in the computing machine. For this, the virtual machine layer of D-Cloud offers the facility of fault injection. Furthermore, it enables to manage computing resources flexibly and automatically, for instance, simulation test can be performed quickly by simultaneous use of resources if available. It automates the process of system setup including fault instillation based on test scenario provided by the tester. Additionally, it automates testing phenomena by utilizing the descriptions for system configuration, and test-cases to perform tests on cloud computing systems [27].

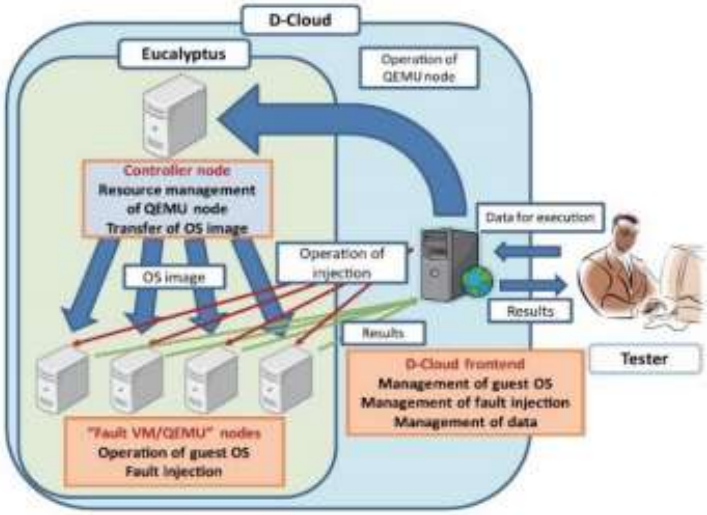

Fig. 5. Architecture of D-Cloud

6. CLOUD ANALYST

Cloud Analyst is a visual tool for analyzing the cloud computing environment and applications. It is developed based on Cloud Sim [28]. The motivation behind Cloud Analyst was the unavailability of tools that can help to estimate the requirements related to the workload on computing servers and user for geographically distributed cloud applications [29]. This requirement and performance analysis is important because cloud contains a distributed infrastructure and applications may run in different geographical locations. Cloud Analyst enables to analyze the performance of extensive cloud applications based on various deployment setups by simulating them. The Cloud Analyst is built with the extensions of Cloud Sim Toolkit It extends the GUI package to ease with separation of programming and simulation exercises. The existing Cloud Sim libraries are used to model the simulation and analysis of applications behavior. Cloud Analyst has the following main components [30].

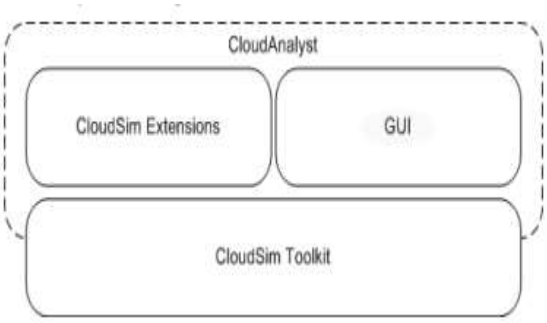

Fig. 6. Cloud Analyst architecture

1) **GUI Package:** The front-end is the graphical user interface to control screen transitions and related functionalities.
2) **Simulation:** This important component enables the development and execution of simulation by retaining the simulation parameters.
3) **User Base:** It is used to model the users and users' traffic.
4) **Data Center Controller:** This module tackles with the activities related to the data center.
5) **Internet:** This is used to exhibit the Internet and traffic routing.
6) **Internet Characteristics:** This is used to define the Internet characteristics that are used for simulation including latency, bandwidth, and region etc.
7) **Vm Load Balancer:** Used to implement the load balancing policies for data centers.
8) **Cloud App Service Broker:** This defines the cloud service broker who is responsible for managing traffic and service delivery between user and service provider.

7. RESULT

A great deal of research has been done, and many solutions have been presented in the area of cloud computing in respect to the RA problem; however, there are still some issues and challenges that need further research, and an optimal solution that is practical for most cloud environments. Cloud Sim focuses the simulation of large-scale data centers host server virtualization with customizable policies for resource provisioning, topologies for data centers, applications that use MPI, and federated clouds. those solutions that have considered the time migration in their load balancers and using some mechanisms for reducing the migration time and therefore reducing the service response time can be more successful than other approaches those using migration techniques.

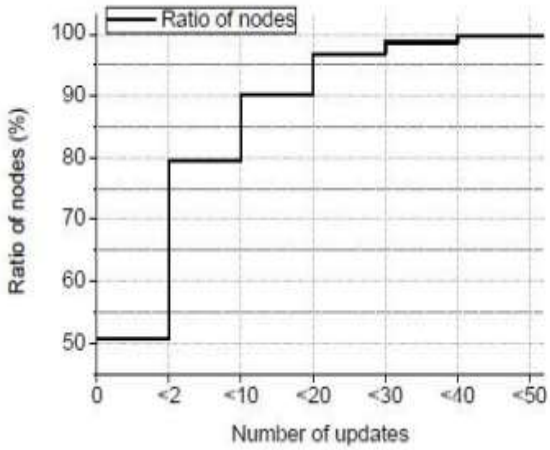

Fig. 7. Update counts distribution for age Rank

8. CONCLUSION AND FUTURE DIRECTIONS

Crowd sourcing is a process of acquisition, integration, and analysis of big and heterogeneous data generated by a diversity of sources in urban spaces, such as sensors, devices, vehicles, buildings, and human. Cloud computing technology is increasingly being used in enterprises and business markets. Subsequently, an effective required for achieving user satisfaction and maximizing the profit for cloud service providers. Load balancing in cloud computing data centers has been a main challenge and an active area of research in recent years. In this paper we have presented a survey on current load balancing techniques and solutions which have proposed only for cloud computing environments. The full multi cloud simulation refers to simulating the multiple clouds in a single environment with complete privileges. Such privileges include the inter-connection among multiple clouds with administrative control and with running different administrative, access and security policies at multiple cloud. Further, it includes the simulation of multiple clients of multiple clouds that accesses the resources in real time. We will consider more cloud load balancing solutions according to our three levels classification and survey the load balancing solutions' trend.

REFERENCES

[1] Z. Song and N. Roussopoulos, "K-nearest neighbor search for moving query point," Lecture Notes in Computer Science, vol. 2121, pp. 79–96, July 2001.

[2] Z. Song and N. Roussopoulos, "K-nearest neighbor search for moving query point," Lecture Notes in Computer Science, vol. 2121, pp. 79–96, July 2001.

[3] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp.881–892, July 2002

[4] Y. Zheng, F. Liu, and H. Hsieh. U-Air: When Urban Air Quality Inference Meets Big Data. In Proceedings of the 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 1436-1444, 2013.

[5] X. Liu, Y. Yang, D. Yuan, and J. Chen. Do we need to handle every temporal violation in scientific workflow systems. ACM Transactions on Software Engineering and Methodology, 2013.

[6] L. Wang, J. Tao, et al. G-Hadoop: MapReduce across distributed data centers for data-intensive computing. Future Generation Computer Systems, 29(3):739-750, 2013.

[7] A. Waqas, A. W. Mahessar, and N. Mahmood, "T RANSACTION M ANAGEMENT T ECHNIQUES AND P RACTICES IN C URRENT C LOUD C OMPUTING E NVIRONMENTS: A S URVEY," vol. 7, no. 1, pp. 41–59, 2015.

[8] A. Waqas, Z. M. Yusof, A. Shah, and N. Mahmood, "Sharing of Attacks Information across Clouds for Improving Security: A Conceptual Framework," in IEEE 2014 International Conference on Computer, Communication, and Control Technology (I4CT 2014), 2014, pp. 255– 260.

[9] A. Waqas, Z. M. Yusof, and A. Shah, "A security-based survey and classification of cloud architectures, state of art and future directions," in Proceedings - 2013 International Conference on Advanced Computer Science Applications and Technologies, ACSAT 2013, 2014, pp. 284– 289.

[10] Foster, I., et al. Cloud computing and grid computing 360- degree compared. in Grid Computing Environments Workshop, 2008. GCE'08. 2008. Ieee.

[11] Rastogi, G. and R. Sushil, Cloud Computing Implementation: Key Issues and Solutions, in 2nd International Conference on Computing for Sustainable Global Development (INDIACom)2015, IEEE. p. 320-324.

[12] Lee Badger, T.G., Robert Patt, Corner JeffVoas. DRAFT Cloud Computing Synopsis and Recommendations. 2011;

[13] Raghavendra Kun, Pramod Kumar Konugurthi, Arun Agarwal, Raghavendra Rao Chillarige and Rajkumar Buyya, "The anatomy of big data computing".

[14] Yu Zhang, "SAE: Toward efficient cloud Data Analysis service for large scale Social Networks.

[15] Puthal, D., et al., Cloud Computing Features, Issues, and Challenges: A Big Picture, in International Conference on Computational Intelligence and Networks (CINE)2015, IEEE. p. 116-123.

[16] Jadeja, Y. and K. Modi. Cloud computing-concepts, architecture and challenges. in Computing, Electronics and Electrical Technologies (ICCEET), 2012 International Conference on. 2012. IEEE.

[17] A. Waqas, Z. M. Yusof, and A. Shah, "A security-based survey and classification of cloud architectures, state of art and future directions," in Proceedings - 2013 International Conference on Advanced Computer Science Applications and Technologies, ACSAT 2013, 2014, pp. 284– 289.

[18] P. Mell and T. Grance, "The NIST Definition of Cloud Computing (Draft) Recommendations of the National Institute of Standards and Technology," in NIST Special Publication 800-145 (Draft), Computer Security Division, Information Technology Laboratory (ITL), U.S. Department of Commerce, Gaithersburg, MD, USA., 2011.

[19] X. Luo, Z. Xu, J. Yu, and X. Chen. Building Association Link Network for Semantic Link on Web Resources. IEEE transactions on automation science and engineering, 8(3):482-494, 2011.

[20] Y. Zheng. Tutorial on Location-Based Social Networks. In Proceedings of the 21st International World Wide Web conference, 2012.

[21] D. Haddow, A. Bullock, and P. Coppola. Introduction to Emergency Management, 2010

[22] Shailesh H. Dinde1 , Arati M. Dixit, "On Sharing Infrastructure Resources using Online Social Networks", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Vol 4 Issue 1, 2015.

[23] H. H. Song, T. W. Cho, V. Dave, Y. Zhang, and L. Qiu,"Scalable proximity estimation and link prediction in onlinesocial networks," in Proceedings of the 9th ACM SIGCOMMconference on Internet Measurement Conference. ACM, 2009, pp.322–335.

[24] Y. Zhang, X. Liao, H. Jin, and G. Min, "Resisting Skewaccumulationfor Time-stepped Applications in the Cloud viaExploiting Parallelism," IEEE Transactions on Cloud Computing,

[25] D. Kliazovich, P. Bouvry, and S. U. Khan, "GreenCloud: a packet-level simulator of energy-aware cloud computing data centers," J. Supercomput., vol. 62, no. 3, pp. 1263–1283, Nov. 2010.

[26] ARCOS, "iCanCloud," 2015. [Online]. Available: http://icancloud.org/Home.html. [Accessed: 27-Apr-2015].

[27] A. Núñez, J. L. Vázquez-Poletti, A. C. Caminero, G. G. Castañé, J. Carretero, and I. M. Llorente, "iCanCloud: A Flexible and Scalable Cloud Infrastructure Simulator," J. Grid Comput., vol. 10, no. 1, pp. 185–209, Apr. 2012.

[28] QEMU, "QEMU, open source processor emulator."

[29] D. Nurmi et al., "The Eucalyptus Open-source Cloud-computing System," in 9th IEEE/ACM International Symposium on Cluster Computing and the Grid, 2009, pp. 124–131.

[30] Eucalyptus, "Eucalyptus|Open Source Private Cloud Software.