Special Issue Published in International Journal of Trend in Research and Development (IJTRD), ISSN: 2394-9333, www.ijtrd.com

A Comparative Study over Parallel and Distributed Cloud Frameworks

[1]Ahthasham Sajid, [2]Amara Akram, [3]Ihsanullah and [4]Atiq Ahmed,
[1,2]Baluchistan University of Information Technology Engineering and Management Sciences Quetta, Baluchistan, Pakistan
[3,4]Department of Computer Science, University of Baluchistan, Quetta

Abstract: This research article presents an elementary survey on cloud computing facts center architecture, discover key parameters and simulate the scalability of the cloud by means of varying variety of nodes and users to measure the put-offs amongst legacy two and three tier records centers. Furthermore, this research puts forward a detail in depth survey on cloud adoption in social networks. There are few research work done in existing literatures on exceptional dimensions of social networks; however, they are more targeted on users security problems therefore, leaving open the scalability troubles for researchers to explore and analyze. This article discusses the cloud architectures in terms with their important traits in social networks; the social community architecture is categorized as: (a) decentralized, (b) centralized, and (c) hybrid. In the end, to illustrate the effectiveness of cloud data middle architecture, a simulation conducted using OMNET + + simulator by varying no. of customers and computing nodes in the cloud structure.

Keywords: *OMNET, Hadoop, Cloud Framework, OMNET ++*

I. INTRODUCTION

Development and adoption of computing technologies are growing daily. To begin with the notion of computing has started out with a sequential implementation on a single threaded gadget, however to meet the continuous growing demand of users makes it extra difficult. The experts explored several techniques and proposed feasible solutions that are expected to be profitable. Primarily based on clinical improvement in machine hardware, notion of parallel computing has been established in 90's. The preliminary computing techniques laid the muse for advanced computing paradigm such a cluster, grid and cloud computing. Served as the basis for cluster, grid and later cloud computing, and this similarly provide a simple basis for social networks. A social network is typically based totally on a cloud. It provides an internet-based platform that permits humans to attach through social networking sites [01]. The introduction of net 2.zero has changed the manner group of people can have interaction with each other. In Social networks, customers are the purchasers of on-line statistics but at the same time take an active part inside. The introduction of statistics within the shape of motion pictures, snap shots, and blogs. As an end result, networking sites like MySpace, Twitter, Google Plus, fb and LinkedIn have witnessed outstanding growth. As in line with the given stats, in January 2013, the variety of Facebook customers exceeded 1000 million. Consequently this rapid boom of social networks offers upward push to further challenges like fault tolerance, scalability, data replication, migration and so forth [1]. Even though the term cluster and grid computing can be used interchangeably but there's a distinction among these two tactics. In easy term we define cluster computing is homogeneous connected nodes in an unmarried administrative domain whilst grid make a specialty of dispensed computing over the network of heterogeneous resources. There are many ways for the implementation of cluster. Beowulf cluster is the instance of imposing the same commodity computers in a community [2].Cluster computing can only be obtained through the use of the cluster of computers whereas Grid computing is something much like cluster computing, it makes use of several computer systems related is some manner, to remedy a massive problem. There may be regularly a little confusion approximately the difference between grids vs. cluster computing. The big distinction is that grid specializes in allotted heterogeneous sources used as a platform for excessive performances computing at the same time as cluster computing specializes in platform such as homogeneous interconnected nodes in a single administrative area. There is not any centralized control module in grid computing structure [3].

1.1.1 Issue of Social Networks:

Social networks having tens of heaps of users having access to records require massive infrastructure to provide easy offerings without any interruption. These servers are needed to assist diverse offerings like records dissemination and to satisfy the performance necessities. Facebook currently has five statistics centers; Four of the information facilities are positioned in extraordinary cities of America and one is located in Sweden [60]. In addition, Twitter has facts centers located at distinct places of the United States. The use of this sort of gigantic infrastructure may be very expensive for the social network service carriers. Following are the important thing elements contributed to the free factor of big information canters.

1.1.2 Cost of Equipment's:

Due to the speedy growth of the social community customers, the underlying infrastructure of the social network also needs to be constantly upgraded. Therefore, the quantity of servers that offer social network offerings vary from few servers to thousands of servers. For example, a Facebook has greater than 180,000 servers and LinkedIn has 30,000 servers until September 2013. Moreover, the price of community system and cabling additionally desires to be considered. Being in line with Greenberg et al., 45% of the statistics center cost is attributed to the servers, 25% to the infrastructure, 15% to the network equipment, and 15% to the strength structures [61].

1.1.3 Maintenance Cost:

The massive scale structures with hundreds of servers are linked to community switches and links. Therefore, those components fail at ordinary duration. The massive infrastructures require continuous troubleshooting and protection to diagnose and get control of the faults. The well-Trained and professional team of workers is needed to control the sort of massive infrastructure. In line with Koomey et al., 1/4 of the annual charge of information facilities are attributed to the operational expenses and three/four is attributed to capital fee that encompasses infrastructure and server expenses [62]. For instance, the daily fee of YouTube for server bandwidth is one million greenback.

1.1.4 Energy Consumptions:

The centralized social community infrastructure holds a considerable quantity of servers that devour enormous amount of power [1]. Further to the strength, this is used to power the servers, approximately identical amount of electricity is consumed by means of the Heating, air flow, and air con (HVAC) gadget for cooling the servers and networking equipment. Furthermore, the community gadget makes use of 10% of the energy fed on with the aid of the IT infrastructure that includes servers and storage. Therefore, the price of energy required to power the centralized social network infrastructure is an important element. Consistent with an examine carried out through Gartner group estimates that approximately 10% of the operational prices of a facts center is attributed in the direction of strong consumption and within the next 5 years that is expected to reach 50% [63]. For example, in keeping with estimates taken in October 2008, Facebook spent roughly two million bucks in step with month to power the servers and cooling. In addition, fb records facilities fed on 678 million Kwh of energy within the 12 months of 2012 that is 33% boom compared to the energy intake in 2011 [60]. Furthermore, the carbon footprint of Facebook, due to CO_2 emissions, elevated 52% in 2012 compared to 2011.

1.2 Social Networks Challenges:

Popular social networking websites faces many demanding situations inside the provision of included offerings to purchasers. Millions of extra users concurrently get rights of entry to those pages. Get admission to through the consumer method massive number of requests which normal internet architecture cannot efficiently take good care of. Some of those demanding situation are fixed as under.

i.Permit near real-time communication.
ii. Aggregate contents on the fly from more than one resource.
iii. Be able to fraction popular contents.
iv. Scale is to method hundreds of thousands of personal requests in keeping with 2d

Those demanding situations are immediately related to abandon consumer. Without a proper answer the delay will increase growth in a wide variety of customers. Right here, awareness is on how distinct social networking websites deal with such problems.

II. RELATED WORK

This segment covers the contemporary painting that explains the inner running of cloud computing paradigm that allows users to apprehend the various architectures and paradigms to be had. In addition, in [12] the authors have spoken of overall resemblance of social networks like Facebook and Google Plus. In rest of this bankruptcy, this section covers the studies painting accomplished below the cloud paradigm and based totally on MapReduce or other such frameworks. The DAvinCi is one of the frameworks this is based on Hadoop and it gives computing framework that is scalable, robust and capable of coping with massive datasets. Its miles primarily based on FastSLAM algorithm and mentioned a substantial overall performance improvement compared to conventional Hadoop map reduces model [13]. The ORCA framework is based on federated cloud version wherein compute capacity is fulfilled throughout the statistics centers. M Xin et al [14] proposed ORCA framework that runs Hadoop cluster across geographically disbursed information centers. The authors have finished this by giving a digital layer on the top of Hadoop cluster. This sediment provides assets on-call for. Moreover, the writers have assessed the performance of intra-Hadoop communication. The Cloud computing paradigm is based on virtualization and vehicle scaling techniques. Further, in [15] the authors have discussed the impact of virtualization on diverse sources including network. For experimental assessment Amazon EC2 is utilized. It monitors that small EC2 instances get forty-50% processing strength. Shufen Zhang et al. in [16] defined the cloud scalability. The cloud offers scalability and availability of the provider even in case of failure internal records facilities. Virtualizes sources of Cloud are managed and maintained by using an administrator, the consumer needs no longer to worry approximately node screw ups. S. M. et al. [17] proposed a framework to apply Cloud computing for actual time programs. The proposed framework marked nodes as dependable as they may be processing actual time software and on a hit completion of the task the concern of node receives expanded. Alternatively the priority receives decreased on failure of the assigned task. It's for new vicinity of studies and really restrained work, exists that talk about the use of cloud computing for real-time programs. Eucalyptus provides a framework to execute very specific protocols for cloud computing environment. Its miles freely available framework constructs to assist researchers and developers. The interface of Eucalyptus is exactly how Amazon EC2 and it permits VM management through its personal well defined interface. Its miles based totally on cleaning soap framework [15]. Eucalyptus is part based layout. The additives are linked through properly define internet interface. Carlos R. Senna et al. offered a framework based totally mostly on Hadoop orchestration so it needs to be able to run packages in federated clouds. It gives an internet based totally interface for manner submission that is received at the orchestration engine. The engine automates the cross-domain clusters, and done document provisioning [18]. Hadoop cloud based fashions are presented in [19]; few relevant freely to be had and business cloud structures are presented in table 1. Google and Amazon have designed their own business cloud computing machines over Hadoop system [25].

Table 1: Commercial vs. open source Cloud Platforms

Open Source Cloud Platforms	Commercial Cloud Platforms
Hadoop HDFS	Google GFS
Hadoop MapReduce	Google MapReduce
Hadoop HBase	Google Bitable
Hadoop Pig	Google SawZall
Eucalyptus	Amazon S3

A. R. Khan et al. has provided a detailed survey on mobile Cloud fashions and packages; in this survey, mobile cloud constraints and their utility fashions are offered [20]. Every other Hadoop like structure, Aurora is designed and implemented to execute PDES codes over Cloud environment [21]. In addition, TW-SMIP is designed to efficaciously improve overall performance of PDES codes over Cloud [22]. Similarly, Hadoop is used in almost every field; companies offer services based on Hadoop are listed below [24]: GBIF (Global Biodiversity Information Facility), Facebook, IBM, Rackspace/Mailtrust, AOL, Fox interactive media, Google, New York Times, PowerSet (now Microsoft), Yahoo!, Amazon/A9, Cooliris, devdaily.com, Ebay, Rackspace. Evaluations affordable for huge social networks MapReduce schemes are followed. In recent Wan X. et. at. [26] Presented X-RIME based on Hadoop MapReduce parallel programming version that permits users to carry out analysis on social networks [25].

Every other vital area in which researchers are Hadoop MapReduce parallel programming is identification of rumors. Rumors are unavoidable in social networks and spreading thru social networks in maximum of the instances. In [26] Hadoop primarily based mobile automate liner set of rules is presented that help people to come across the rumors. Consequently three roles are described for rumors algorithm, rumor makers are folks that distribute rumors in social networks. If you want to propagate rumor, assistance of different people is required. As a result, rumor maker cannot achieve this purpose without others help. Then there are harmless folks who do now not percentage assistance for rumors spreading. In order to analyze Hadoop primarily based cellular automate algorithm is deployed on open source Hadoop platform, the targeted working and take a look at outcomes [26]. Social recommended gadget is some other region especially more explored based on social networks. As Social Networks collects a big quantity of consumer-generated facts. The generated records are of no need until positive analysis is performed to be had records. This record is used to generate advice to satisfy customers want. Web Social advice Systems (SRSs) can help customers discover records they're fascinated. In SRS machine advice are generated based on big quantity of data; therefore. Hadoop MapReduce parallel programming framework is followed through Chaobo He et. al [27].

III. SIMULATION SETUP

In simulation scenario, different architecture used inside data centers is compared. The comparison is done by varying the number of users and computing nodes. To perform this study. Discrete Event Simulator is used. OMNeT++ is a C++ discrete communications networks simulation and modeling of multi-processor distributed or parallel networks. OMNeT ++ is open source, available below the GNU general Public License. It's miles discrete occasion simulation device. The information of OMNeT++ can be located in [67].OMNeT++ is primarily based on modules that change message by using exchanging pointers of the share memory. Those components are termed as easy modules, completely written in C++. All the user defined lessons must be inherited from easy module furthermore for improving simulation, is accomplished via properly defined ports referred as gates The hierarchy of modules isn't constrained. User can outline much range of nested modules. The concept of modules is derived from DEVS offered in [66]. Two and three tier architecture used within a data center is compared. The comparison is performed on variable users count and computing servers. To perform this study, following network properties are in used.

To degree performance primarily based at the postpone, one to all simulation version is used. In this model, the consumer selects the vacation spot node randomly. Every of the experiments have been executed on 3 instances to normalize reading [65].

Table 2: Simulation Properties

Simulation Properties	Values
Core to Aggregate switch – data rate	1Gbps
Aggregate to Access switch – data rate	1 Gbps
Access switch to Compute Node	100 Mbps
User node to Core switch	100 Mbps
Core to Aggregate switch – BER	$5\,x\,10^{-12}$
Aggregate to Access switch – BER	$5\,x\,10^{-12}$
Access switch to Compute Node	$5\,x\,10^{-5}$
User node to Core switch	$5\,x\,10^{-5}$
Number of Users	Vary
Number of Compute Nodes	Vary
Number of racks	03

Figures 3-7 and determine 3-eight shows two and three tier structure models in OMENT++; below graphs suggest that in 3-tier architecture, because of an additional layer of switches. Discern three nine examine the two architectures by means of changing the nodes remember and preserving users constant i.e.2 in all the simulation runs. Figure 3.10 shows the as compared outcome puts -off variable, retaining 20 customers within the simulation setup. Further figure 3.5 shows the outcomes of 40 customers. In every of the figures, 2-tier dominates the 3-tier architecture because of the combination of the transfer layer. Alternatively, three tiers is broadly utilized architecture because of its scalability. It can accommodate extra variety of nodes compared to a two-tier structure. In Figures, the cut off is at the growing side, this delay consists of the ARP solve time. If on every occasion a user sends a packet to a specific compute node, the cut off increases due to ARP, but once all of the paths are recognized, the put off decreases as proven inside the figures. The motive for surprising dip in graphs under is owing to random choice of nodes. Every time person selects a random vacation spot node the query is resolved via address resolution protocol.

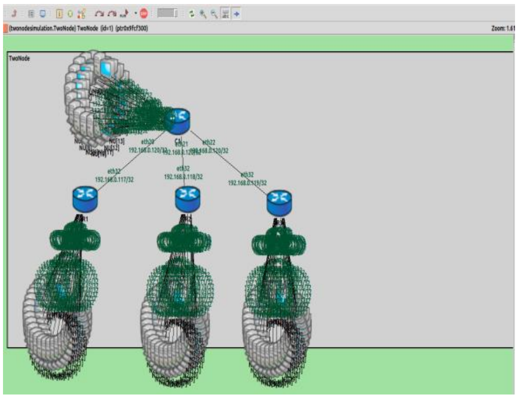

Figure 3-1: Running simulation snapshot: 2-T Model

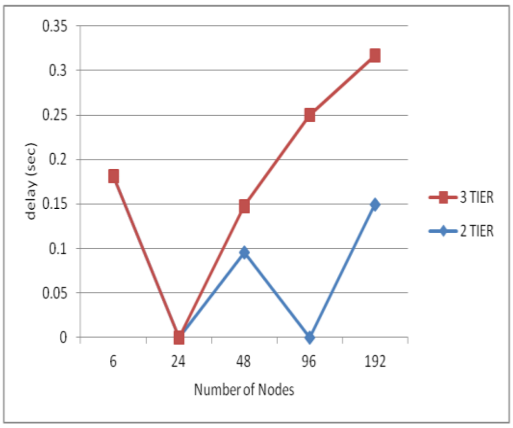

Figure 3-2: Three Tier architecture: running simulation snapshot

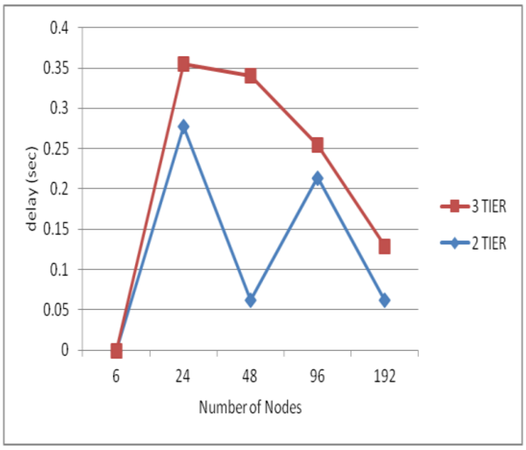

Figure 3-3: Architecture comparison, varying number of nodes, users = 20

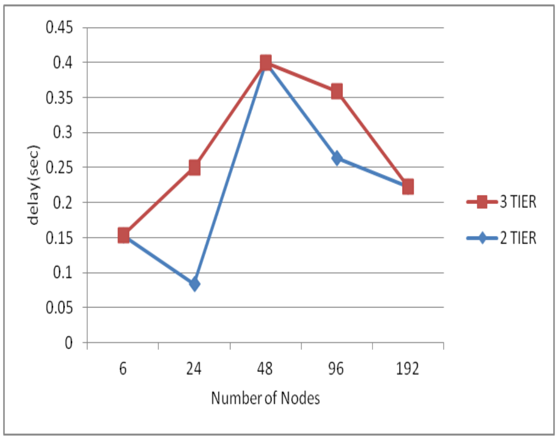

Figure 3-4: Architecture comparison, varying number of nodes, users = 30

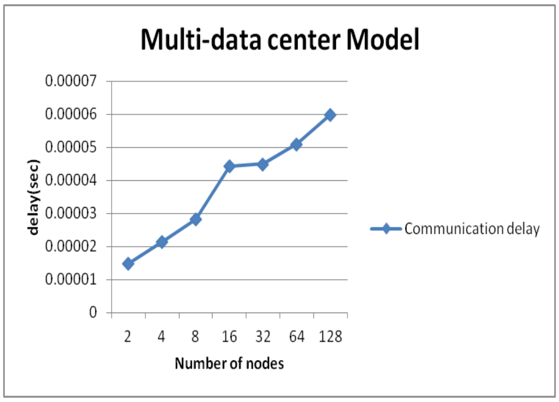

Figure 3-4: Architecture comparison, varying number of nodes, users = 40

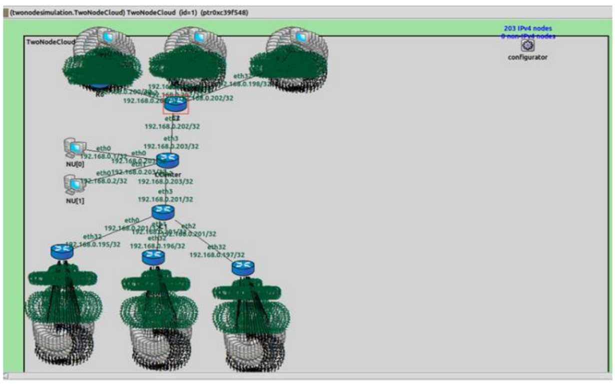

Figure 3-6: Communication delay between multiple data canter model by varying number of users and kept number of nodes fixed

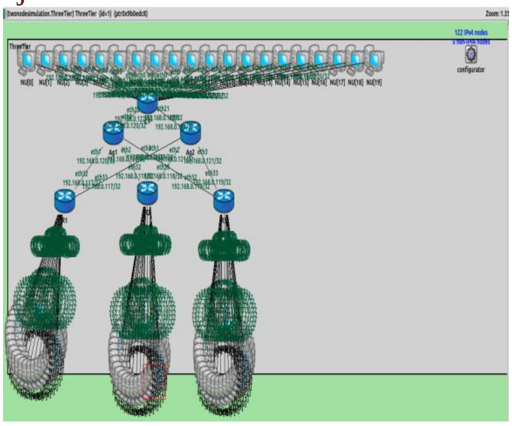

Figure 3-7: Graphical user interface of multi data centre simulation model

IV. RESULTS DISCUSSION

To build a more realistic simulation, multiple data centers cloud model connected through high speed communication channel is developed. The delay is measured in multi data center on i) variable nodes and ii) variable users. Fig 3.2 and Fig 3.3 describes the screen shot of simulation over 2-tier and 3-tier architecture respectively. The outcomes are taken into account in Fig. 3.5 and Fig. 3.6. In Fig. 3.6, the delayincreases by increasing the number of nodes, but there is certainly sudden variation at 16-32, again that is due to random selection of destination nodes. The comparison of the graph of Figure 3.6 It clearly shows that the delay is very high with the increase in users. The Fig. 3.7 displays the graphical layout of multiple data center simulation model. Over the years, 2-T architecture has widely used inside data centers. The 2-T data center can accommodate five thousand five hundred nodes; whereas, 3-tier architecture supports more than 10,000 computing machines using inexpensive TOR switches. This is achieved due to the extra layer of switches which are not available in two-tier architecture. So, selection between two and three-tier architecture is a trade-off between delay and scalability. There are advance techniques available that reduce the delays by using context-aware server replication techniques; but on the other hand all the existing social networks cannot compromise on scalability. Therefore, three-tier architecture always dominates over two-tier architecture due to its scalability.

CONCLUSION

This research paper presented an in-intensity evaluation of social network traits at the side of one-of-a-kind social network architecture. This research paper offered the perception of various cloud providers, mainly the architecture and technology used to facilitate the end user. The idea of cloud is primarily based on virtualization and computing energy is supplied through virtual servers. Diverse unique parameters are mentioned in this research paper. The discussion includes the working of Hadoop, together with Facebook and Google and the architecture of Facebook, MySpace and Google. They all are using similar techniques to facilitate stop customers. In this research performance of two and three tier architecture, as they may be maximum adopted strategies used a facts centers. The assessment is finished through varying the quantity of customers and computing nodes. In this carry out look at numerous community parameters is used and one to all simulation models is followed. The outcomes indicates extra put off in 3 tier architecture this is due to the addition of combination layer but then again it a affords extra flexibility to three tier architecture to add greater compute nodes with commodity hardware switches, whereas in two tier structure the put-off is less however it has a restriction in placing of computer node.

VI. FUTURE WORK

In Future, this work can be prolonged by using going into similarly element of scheduling algorithms and green and brown electricity consumption strategies, exceptional cloud vendors are the usage, as energy intake is one of the actual difficulty and due to that maximum of the companies moving their data centers to Iceland. Secondly the interchangeability among special cloud carriers is every other vicinity to explore in density.

Acknowledgements

I am really thankful for my parent organization BUITEMS for the financial support I have received from them and by Higher Education Commission in Pakistan to attend this prestigious academic event to explore latest research ideas.ICTIMESH provided a good platform for researchers across world to get connected.

References

[1] Kryczka, M., et al., *A first step towards user assisted online social networks*, in *Proceedings of the 3rd Workshop on Social Network Systems*. 2010, ACM: Paris, France.

[2] Yang, Z., et al., *E-SmallTalker: A Distributed Mobile System for Social Networking in Physical Proximity*, in *Proceedings of the 2010 IEEE 30th International Conference on Distributed Computing Systems*. 2010, IEEE Computer Society.

[3] http://www.facebook.com/notes/facebook-data-team/anatomy-of-facebook/10150388519243859. [cited 2014 14 Dec].

[4] O'Neill, N.F.s.r.s.p.h.a.c.f.-r.-s.-p.b.O.a.-D.-.

[5] Korpela, E., et al., *SETI@home-massively distributed computing for SETI*. Computing in Science & Engineering, 2001. **3**(1): p. 78-83.

[6] Rohit, E., et al. *A Robust Communication Framework for Parallel Execution on Volunteer PC Grids*. in *Cluster, Cloud and Grid Computing (CCGrid), 2011 11th IEEE/ACM International Symposium on*. 2011.

[7] Liu;, H. and D. Orban;. *GridBatch : Cloud computing for large-scale data-intensive batch applications*. in *IEEE international symposium on cluster computing and the grid* 2008.

[8] Park, A. and R. Fujimoto. *A scalable framework for parallel discrete event simulations on desktop grids*. in *2007 8th IEEE/ACM International Conference on Grid Computing, .* 2007.

[9] Armbrust;, M., et al., *Above the Clouds: A Berkeley View of Cloud Computing*. 2009, EECS Department University of California Berkeley Tech Rep UCBEECS200928 (2009): California. p. 25.

[10] Armbrust, M., et al., *A view of cloud computing.* Commun. ACM. **53**(4): p. 50-58.

[11] Zhao, W., et al. *Modeling and simulation of cloud computing: A review*. in *Cloud Computing Congress (APCloudCC), 2012 IEEE Asia Pacific*.

[12] Bo-Wen, Y., et al. *Cloud Computing Architecture for Social Computing - A Comparison Study of Facebook and Google*. in *Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on*. 2011.

[13] Enti et. al. *DAvinCi: A cloud computing framework for service robots*. in *(ICRA), 2010 IEEE International Conference on Robotics and Automation*. 2010.

[14] H. Alipour,et al, *"Model Driven Performance Simulation of Cloud Provisioned Hadoop MapReduce Applications,"* 2016 IEEE/ACM 8th International Workshop on Modeling in Software Engineering (MiSE), Austin, TX, 2016, pp. 48-54.

[15] Wang;, G. and T.S.E. Ng;. *The Impact of virtualization on Network Performance of Amazon EC2 Data Center*. in *INFOCOM*. 2010. Houston TX USA: IEEE.

[16] Ajaltouni, E.E., A. Boukerche, and M. Zhang, *A grid-based DEVS approach to dynamic load balancing for large scale distributed simulations*, in *Proceedings of the 2009 Spring Simulation Multiconference*. 2009, Society for Computer Simulation International: San Diego, California.

[17] Malik, S. and F. Huet. *Adaptive Fault Tolerance in Real Time Cloud Computing*. in *Services (SERVICES), 2011 IEEE World Congress on*.

[18] Senna, C.R., L.G.C. Russi, and E.R.M. Madeira. *An Architecture for Orchestrating Hadoop Applications in Hybrid Cloud*. in *Cluster, Cloud and Grid Computing (CCGrid), 2014 14th IEEE/ACM International Symposium on*. 2014.

[19] Lu, H., C. Hai-Shan, and H. Ting-Ting. *Research on Hadoop Cloud Computing Model and its Applications*. in *Networking and Distributed Computing (ICNDC), 2012 Third International Conference on*. 2012.

[20] Khan, A.R., et al., *A Survey of Mobile Cloud Computing Application Models.* Communications Surveys & Tutorials, IEEE. **16**(1): p. 393-413.

[21] Park, A. and R.M. Fujimoto, *Aurora: An Approach to High Throughput Parallel Simulation*, in *Proceedings of the 20th Workshop on Principles of Advanced and Distributed Simulation*. 2006, IEEE Computer Society. p. 3-10.

[22] Malik;, A., A.A. Park;, and A.R. Fujimoto;. *Optimistic Synchronization of Parallel Simulations in Cloud Computing Environments*. in *Proceedings of the 2009 IEEE International Conference on Cloud Computing*. 2009. India: IEEE Computer Society.

[23] Metzker, M.L., *Sequencing technologies the next generation.* Nat Rev Genet. **11**(1): p. 31-46.

[24] http://wiki.apache.org/hadoop/PoweredBy, H.U.

[25] Wei, X., S. JuWei, and Y. Bo. *X-RIME: Cloud-Based Large Scale Social Network Analysis*. in *Services Computing (SCC), 2010 IEEE International Conference on*.

[26] Hui, Z., L. Ji, and X. Yueliang. *Hadoop Cellular Automata for Identifying Rumor in Social Networks*. in *Information Science and Cloud Computing Companion (ISCC-C), 2013 International Conference on*.

[27] He;, C., et al., *SRSH: A Social Recommender System based on Hadoop.* International Journal of Multimedia and Ubiquitous Engineering, 2014. **9**(6).

[28] Akioka, S. and Y. Muraoka. *HPC Benchmarks on Amazon EC2*. in *Advanced Information Networking and Applications Workshops (WAINA), 2010 IEEE 24th International Conference on*. 2010.

[29] Alam, A. and J. Ahmed. *Hadoop Architecture and Its Issues*. in *Computational Science and Computational Intelligence (CSCI), 2014 International Conference*

on.

[30] *Hadoop documentation* Available at : *http://www.hadoop.apache.org*.

[31] Mackey, G., et al. *Introducing map-reduce to high end computing*. in *Petascale Data Storage Workshop, 2008. PDSW '08. 3rd*. 2008.

[32] Patil;, V.S. and P.D. Soni;, *Hadoop Skeleton and Fault Tolerance in Hadoop Clusters*. Journal of Application on Innovation in Engineering and Management, 2013. **2**(2).

[33] *Hadoop architecture http://hadoop.apache.org/core/* .

[34] Delmerico, J.A., et al. *Comparing the performance of clusters, Hadoop, and Active Disks on microarray correlation computations*. in *High Performance Computing (HiPC), 2009 International Conference on*. 2009.

[35] Shvachko, K., et al. *The Hadoop Distributed File System*. in *Mass Storage Systems and Technologies (MSST), 2010 IEEE 26th Symposium on*.

[36] Khan, M.A., Z.A. Memon, and S. Khan. *Highly Available Hadoop NameNode Architecture*. in *Advanced Computer Science Applications and Technologies (ACSAT), 2012 International Conference on*.

[37] Amazon. *Amazon Elastic Compute Cloud (Amazon EC2)," http://aws.amazon.com/ec2/*
[cited 2014 30-12-2014].

[38] Aradhana et al. On the Go, a Cloud application for social networking, *4th International Conference on New Trends in Information Science and Service Science*, Gyeongju, 2010, pp. 108-112

[39] Chard, K., et al. *Social Cloud: Cloud Computing in Social Networks*. in *Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on*.

[40] Guo, C., et al., *Dcell: a scalable and fault-tolerant network structure for data centers*, in *Proceedings of the ACM SIGCOMM 2008 conference on Data communication*. 2008, ACM: Seattle, WA, USA.

[41] Dzmitry Kliazovich, Pascal Bouvry, and S.U. Khan, *GreenCloud: a packet-level simulator of energy-aware cloud computing data centers.* The Journal of Supercomputing 2012. **62**(3): p. 1263 - 1283.

[42] Yao;, Y., et al. *HaSTE: Hadoop YARN Scheduling Based on Task-Dependency and Resource-Demand*. in *7th IEEE International Conference on Cloud Computing, Anchorage*. 2014.

[43] Murthy;, A.C., et al., *Apache Hadoop YARN*, ed. Addison-Wesley. 2014.

[44] Haque, M.E., et al. *Providing green SLAs in High Performance Computing clouds*. in *Green Computing Conference (IGCC), International*. 2013.

[45] Dean, J. and S. Ghemawat, *MapReduce: simplified data processing on large clusters.* Commun. ACM, 2008. **51**(1): p. 107-113.

[46] *Google Protocol buffer: https://developers.google.com/protocol-buffers/ Last accessed: 15-Dec-2104*.

[47] Lam, C.F. *Datacenter networks*. in *OptoeElectronics and Communications Conference (OECC), 2011 16th*.

[48] Shvachko, K., et al. *The Hadoop Distributed File System*. in *Mass Storage Systems and Technologies (MSST), 2010 IEEE 26th Symposium on*. 2010.

[49] Dwyer, C.H., Starr Roxanne; and Passerini, Katia, . *"Trust and Privacy Concern Within Social Networking Sites: A Comparison of Facebook and MySpace"*. in *Proceedings of AMCIS* 2007.

[50] Heymann, Paul et al. Fighting Spam on Social Web Sites: A Survey of Approaches and Future Challenges, IEEE Internet Computing volume 11, num 6, pp 36-45, 2007.

[51] Narendula, R., T.G. Papaioannou, and K. Aberer. *Privacy-Aware and Highly-Available OSN Profiles*. in *Enabling Technologies: Infrastructures for Collaborative Enterprises (WETICE), 2010 19th IEEE International Workshop on*. 2010.

[52] Furht, B., et al., *Decentralized Online Social Networks*, in *Handbook of Social Network Technologies and Applications*. 2010, Springer US. p. 349-378.

[53] Anh Tuan, N., et al. *Stir: Spontaneous social peer-to-peer streaming*. in *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*. 2011.

[54] Sarigol, E., O. Riva, and G. Alonso. *A tuple space for social networking on mobile phones*. in *Data Engineering (ICDE), 2010 IEEE 26th International Conference on*. 2010.

[55] Tuan Anh, D.T., et al., *Mosco: a privacy-aware middleware for mobile social computing.* Journal of Systems and Software, 2014. **92**(0): p. 20-31.

[56] Jinzhou, H. and J. Hai. *SoMed: A Hybrid DHT Framework towards Scalable Decentralized Microblogging Services*. in *Parallel and Distributed Systems (ICPADS), 2013 International Conference on*. 2013.

[57] Pujol, J.M., et al., *The little engine(s) that could: scaling online social networks.* SIGCOMM Comput. Commun. Rev., 2010. **40**(4): p. 375-386.

[58] Sakaki, T., M. Okazaki, and Y. Matsuo, *Earthquake shakes Twitter users: real-time event detection by social sensors*, in *Proceedings of the 19th international conference on World wide web*. 2010, ACM: Raleigh, North Carolina, USA.

[59] *Hoff, T. (2009). Why are Facebook, Digg, and Twitter so hard to scale? http://highscalability.com/blog/2009/10/13/why-are-facebook-digg-and-twitter-so-hard-to-scale.html. [Online; accessed 16-Dec-2014]*.

[60] *Miller, R. (2013). Facebook's power footprint growing, moving east. <http://www.datacenterknowledge.com/archives/2013/07/22/facebooks-shifting-power-footprint/> . [Online; accessed 18-Dec-2014]*.

[61] Greenberg, A., et al., *The cost of a cloud: research problems in data center networks.* SIGCOMM Comput. Commun. Rev., 2008. **39**(1): p. 68-73.

[62] *Koomey, J., Brill, K., Turner, P., Stanley, J., and Taylor, B. (2007). A simple model for determining true total cost of ownership for data centers. Uptime Institute White Paper, Version, 2*.

[63] Bilal, K., et al., *A taxonomy and survey on Green Data Center Networks*, in *Special Section: Intelligent Big Data Processing Special Section: Behavior Data Security Issues in Network Information Propagation Special Section: Energy-efficiency in Large Distributed Computing Architectures Special Section: eScience Infrastructure and Applications*. 2014. p. 189-208.