ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Random Probability Distribution Method of Scheduling Tasks in Large Dispersed Data Centers for Delay Tolerant Activities

Hareesh Chatarasupalli
MTech Student,
Department of CSE,
Nova College of Engineering and Technology,
Jangareddigudem, West Godavari.

Revathi Mallipudi, M.Tech
Assistant Professor,
Department of CSE,
Nova College of Engineering and Technology,
Jangareddigudem, West Godavari.

Abstract:

A data center is a facility used to house computer systems and associated components, such as telecommunications and storage systems. It generally includes redundant or backup power supplies, redundant data communications connections, environmental controls and various security devices. With the vital objective of removing downtime and sharing data across regions, enterprises are deploying geographically dispersed data centers to minimize planned or unplanned downtime. Large dispersed data centers are industrial scale operations using as much electricity as a small town. Power is the largest recurring cost to the user of a large dispersed data center. This paper focuses on reducing cost of power for such data centers. In this paper we implement a Random probability distribution method of scheduling tasks in large data centers for delay tolerant activities.

Introduction:

Data centers have their roots in the huge computer rooms of the early ages of the computing industry. The availability of inexpensive networking equipment, coupled with new standards for network structured cabling, made it possible to use a hierarchical design that put the servers in a specific room inside the company. The use of the term “data center”, as applied to specially designed computer rooms, started to gain popular recognition in 1980s. The boom of data centers came during the dot-com bubble of 1997–2000. Companies needed fast Internet connectivity and non-stop operation to deploy systems and to establish a presence on the Internet. Installing such equipment was not viable for many smaller companies. Many companies started building very large facilities, called Internet data centers (IDCs), which provide commercial clients with a range of solutions for systems deployment and operation. New technologies and practices were designed to handle the scale and the operational requirements of such large-scale operations.

These practices eventually migrated toward the private data centers, and were adopted largely because of their practical results. Data centers for cloud computing are called cloud data centers (CDCs). But nowadays, the division of these terms has almost disappeared and they are being integrated into a term “data center”.Globalization, security and disaster recovery considerations are driving business to diversify locations across multiple regions. In addition, organizations are looking to distribute workloads between computers, share network resources effectively and increase the availability of applications. With the ultimate goal of eliminating downtime and sharing data across regions, enterprises are deploying geographically dispersed data centers to minimize planned or unplanned downtime.Datacenters use a lot of power, consumed by two main usages: the power required to run the actual equipment and then the power required to cool the equipment.

The first category is addressed by designing computers and storage systems that are increasingly power-efficient. To bring down cooling costs datacenter designers try to use natural ways to cool the equipment. Many datacenters are located near good fiber connectivity, power grid connections and also people-concentrations to manage the equipment, but there are also circumstances where the datacenter can be miles away from the users and don’t need a lot of local management. Examples of this are the ‘mass’ datacenters like Google or Facebook: these DC’s are built around many standardized servers and storage-arrays and the actual users of the systems are located all around the world. After the initial build of a datacenter staff numbers required to keep it running are often relatively low: especially datacenters that provide mass-storage or computing power which don’t need to be near population centers. Datacenters in arctic locations where outside air provides all cooling are getting more popular as cooling and electricity are the two main variable cost components.

Load balancing and energy consumption is the biggest issue in cloud. Payment of a data centre for energy and cooling may be larger than the overall investment in the computing system . Therefore, minimize energy consumption with balancing the workload of resources is a hot topic not only cloud computing but also in other areas. The propose technique is suitable in delay tolerent workload tasks. Their goal is to save maximum energy of the data center. In this paper we are dealing with workload that can be generally classified as delay-tolerant. Delay-tolerant tasks include compute-intensive or data-intensive jobs that require a relaxed service delay, such as scientific computing applications and web index updating. Recent reports showed that supporting delay tolerant workloads in data centers improved the utilization of computer systems and saved computing costs significantly in cloud computing enterprises.

Related Work:

Yuan Yao [1], From this work they focus on a stochastic optimization based approach to make distributed routing and server management decisions in the context of largescale, geographically dispersed data centers, which suggestions important possible for exploring power cost reductions. Their work reflects such decisions at diverse time scales and offers demonstrable power cost and delay characteristics. The usefulness of their method and its robustness are also illustrated through simulation-based experiments under delay tolerant workloads. Their proposed solution exploits temporal and spatial variations in the workload arrival process (at the front end servers) and the power prices (at the back end clusters) to decreases power cost. It also simplifies a cost vs. delay trade-off which permits data center operators to reduce power cost at the expense of increased service delay. Hence, their work is suited for delay tolerant workloads such as massively parallel and data intensive Map Reduce jobs.

DzmitryKliazovich et al, [2] analyze the Cloud computing data centers are becoming increasingly popular for providing computing resources. Hence, the overheads of these data centers has skyrocketed through the increase in computing capacity with large percentage of the operational expenses due to energy consumption, specifically in data centers that are used as backend computing infrastructure for cloud computing. This chapter highlights the part of the communication material in energy consumption and presents solutions for energy efficient network aware resource allocation in clouds.

There are two main alternatives for reducing the energy consumption of data centers: (a) shutting down devices or (b) scaling down performance. Daniel Gmach et al [3], study the advances in virtualization technology are enabling the creation of resource pools of servers that permit multiple application workloads to segment each server in the pool. Sympathetic the nature of initiative workloads is crucial to properly designing and provisioning current and forthcoming services in such pools. This paper considers issues of workload analysis, performance modeling, and capacity scheduling. Their goal is to systematize the effectual use of resource pools when hosting large numbers of enterprise services. A workload examination reveals the business and repetitive nature of initiative workloads. Workloads are robotically classified permitting to their episodic behavior.

Ankita Sharma et al [4], study the Cloud computing is an expanding area in investigation and industry today, which comprises virtualization, circulated computing, internet, and software and web services. Author's presents a method for scheduling algorithms that can maintain the load balancing. From this work they have established power optimization algorithm which over comes the limitations of the previous algorithms Round Robin, Equally Spread Current Implementation Algorithm, Throttled Load Balancing which are used for the over load management of the data leading to positive consequences in terms of overall power consumption of the data center thus helping in green computing. As due to undue excess of traffic flow and then overhead due to mitigation and migration of the virtual machines to balance out the operations there is always an impact on the power consumption, if there is more overload, there is bound to be more power ingesting, and if balancing works well, there is bound to be an optimized trade-off for energy consumption. Results have shown that overall impact of power consumption is reduced by using the proposed algorithm.

Vijaya -Kumar-C et al [5] their development is based on, difficulty data exhaustive applications are increasing in cloud computing. We know that it can reduce investments, human resources and enhance productivity. Data centers play a key role with rapid growth online services of client demands in terms of providing the infrastructures as services (Iaas). For data exhaustive application needs more number of data centers and also massive amount of energy used to operating the servers. Due to increases in data centers in different locations its impact on environment in terms of increased the carbon footprint.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

We proposed virtual machine migration (VMM) technique to optimize data centers, satisfy performance resource distribution, and reduce the server disappointments and also energy consumption. To reduce the energy consumption, they are proposed virtual machine placement and dynamic load balancing algorithms.

Proposed System:

Geo distributed data center means many data centers are geo graphically distributed and connected through the WAN environment .In recently many organizations move to this geo distributed data center. Because they stored large or massive volume of datas. If they are using our own data center means only limited storage will be there so only many of them used this geo distributed data centers.

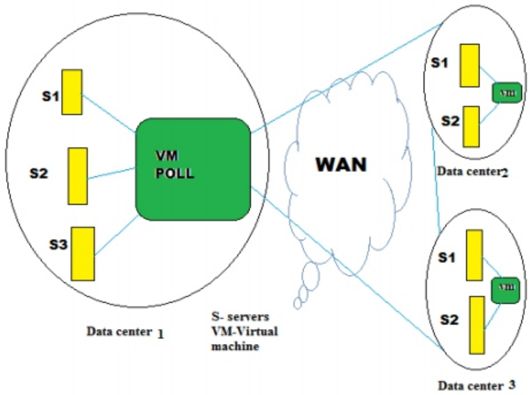

Fig Geo distributed datacenters

Stochastic optimization (SO) methods are optimization methods that generate and use random variables. For stochastic problems, the random variables appear in the formulation of the optimization problem itself, which involve random objective functions or random constraints, for example. Stochastic optimization methods also include methods with random iterates. Some stochastic optimization methods use random iterates to solve stochastic problems, combining both meanings of stochastic optimization. Stochastic optimization methods generalize deterministic methods for deterministic problems. In order to determine workload arrival process and power prices, we use temporal and spatial variations. A variant that is a function of time. Temporal variation refers to changes accrued with elapsing time and spacial to geogrphical/horizontal variations which are also denoted as vertical and horisontal variation respectively. Spatial variability occurs when a quantity that is measured at different spatial locations exhibits values that differ across the locations.

Spatial variability can be assessed using spatial descriptive statistics such as the range.A performance metric is defined for efficiency of observations. A workload prediction scheme does not deal with a generalized system model and they are not suitable for providing an accurate tradeoff between power saving and performance reduction. Therefore, stochastic techniques are required to mitigate the limitations of predictive techniques. The stochastic policies are better in terms of the power delay trade-off than heuristic policies.

Conclusion:

For the users request of data in web, Load balancing and energy consumption is the biggest issue for data centers. It contains the thousands of servers to share the data in cloud. The cost of a data center for energy and cooling may be larger than the overall investment in the computing system. Users of data centers consumed the more energy in both academic and industry. Therefore, minimize energy consumption with balancing the workload of resources is a main credit. In this paper, we study the data placement and data Assignment to minimize the power cost in geographically distributed data centers. In this paper we implement a Random probability distribution method of scheduling tasks in large data centers for delay tolerant activities.

References:

[1] Yuan Yao, Longbo Huang , Sharma A.B, Golubchik L & Neely M.J., Power Cost Reduction in Distributed Data Centers: A Two-Time-Scale Approach for Delay Tolerant Workloads, IEEE Transactions on Parallel and Distributed Systems, (Volume:25 , Issue: 1)

[2] DzmitryKliazovich, Pascal Bouvry et al, "Energy Consumption Optimization in Cloud Data Centers".

[3] Daniel Gmach et al, "Workload Analysis and Demand Prediction of Enterprise Data Center Applications".

[4] Ankita Sharma et al, Upinder Pal Singh, "Energy Efficiency in Cloud Data Centers Using Load Balancing

[5] Vijay -Kumar-C, Dr. G.A. Ramachandra, "Thrusting Energy Efficiency for Data center in Cloud Computing Using Resource Allocation Techniques"

[6] http://www.networkworld.com/article/2237134/cisco-subnet/geographically-dispersed-data-centers.html

[7]http://www.datacenterknowledge.com/archives/2014/12/11/reducing-energy-consumption-cost-data-center/

[8] Jianying Luo, Lei Rao, and Xue Liu, Temporal Load Balancing with Service Delay Guarantees for Data Center Energy Cost Optimization, http://simula.stanford.edu/~jyluo/publications/ecoIDC_tpds_2013.pdf

[9] R. Urgaonkar, B. Urgaonkar, M. J. Neely, and A. Sivasubramaniam, "Optimal Power Cost Management Using Stored Energy in Data Centers," in Proceedings of International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS). ACM, 2011, pp. 221–232.

[10] X. Fan, W.-D. Weber, and L. A. Barroso, "Power Provisioning for A Warehouse-sized Computer," in Proceedings of the 34th Annual International Symposium on Computer Architecture (ISCA). ACM, 2007, pp. 13–23.

[11] S. Govindan, A. Sivasubramaniam, and B. Urgaonkar, "Benefits and Limitations of Tapping Into Stored Energy for Datacenters,"in Proceedings of the 38th Annual International Symposium on Computer Architecture (ISCA). ACM, 2011, pp. 341–352

[12] P. X. Gao, A. R. Curtis, B. Wong, and S. Keshav, "It's Not Easy Being Green," in Proceedings of the ACM Special Interest Group on Data Communication (SIGCOMM). ACM, 2012, pp. 211–222.

[13] S. A. Yazd, S. Venkatesan, and N. Mittal, "Boosting energy efficiency with mirrored data block replication policy and energy scheduler," SIGOPS Oper. Syst. Rev., vol. 47, no. 2, pp. 33–40, 2013.

[14] J. Cohen, B. Dolan, M. Dunlap, J. M. Hellerstein, and C. Welton, "Mad skills: new analysis practices for big data," Proc. VLDB Endow., vol. 2, no. 2, pp. 1481–1492, 2009.

[15] R. Kaushik and K. Nahrstedt, "T*: A data-centric cooling energy costs reduction approach for Big Data analytics cloud," in 2012 International Conference for High Performance Computing, Networking, Storage and Analysis (SC), 2012, pp. 1–11.