A Survey for Dynamic Balancing of workload & Scalability in Cloud Environment

Pranjali D. Dhore and Dr. Kishor R. Kolhe

MITCOE, Pune

Abstract— **Cloud Computing is an emerging technique works on distributed computing mechanism. Cloud computing has reengineered the concepts in information Technology infrastructure. Cloud Computing provides steadfast, on requirement infrastructure over the globe and within shortest possible time span and reduces investment in IT infrastructure. In cloud computing Scalability & Load Balancing are key factors as huge data is created and maintained & better utilisation of in such dynamic environment. Achieving scalability with load balancing is pre requisite in such Big Data environment. Various techniques have been developed in the field so far and are in process of development. In this paper we studied and evaluated various techniques for load balancing and scalability and presented the methodologies explored so far. There are various load balancing & Scalability techniques developed so far and research on the same is increasing significantly. This paper explores different techniques with snapshot on methodologies used, key parameters to be considered and limitations and future scope and development in the load balancing techniques with maintaining scalability in details. This paper reveal the research on the load balancing techniques, will surely be useful for future development.**

Keywords: *Load Balancing; Cloud Computing; Scalability; Data Shedding;*

I. INTRODUCTION

With the Advancement in technology and huge database are developed and query processing has become challenging task in distributed Database Systems & ERP. Managing database and maintaining database at various locations and processing of queries related to database is getting immense importance. With the rapid increase in Data Warehouse, Database servers, distributed database locations and processing at various nodes processing of queries has become complex procedure. Researchers in the field have proposed various techniques and methodologies for reducing turnaround time and enhancing scalability. In most of the scenarios focus is on database query processing and balancing the stream workloads to avoid overload situations. Load Balancing in Cloud environment is essential for maximum utilization of resources and processing distribution over nodes available. Scalability refers to functional capacity of hardware or application to process the query or function in different environment such as change in volume of Processing data, available resources etc. In cloud computing Load Balancing is key factor and cloud functioning is directly depend on Load Balancing mechanism. Thus research and development in Load Balancing is pre-requisite and pervasive. In this paper we studied various techniques of load balancing of Database and maximum throughput for resolving the queries in an Cloud Computing wherein Database engine need to process the queries at various nodes.

The important things to consider while developing such algorithm are : estimation of load, comparison of load, stability of different system, performance of system, interaction between the nodes, nature of work to be transferred, selecting of nodes and many other ones. This load considered can be in terms of CPU load, amount of memory used, delay or Network load. Load Balancing algorithms are of two types depending on current state of

state of system as Static and Dynamic. In static algorithm prior knowledge of system components is pre requisite and thus does not depend on current system. On the other hand Dynamic Algorithm is dependent on current state of system component and is better than static mechanism.

This paper includes Section II for existing techniques of Load Balancing, Section III Advantages, Section IV contain analysis and Future Scope, Section V provides Conclusion.

II. EXISTING LOAD BALANCING TECHNIQUES IN DISTRIBUTED SYSTEM ENVIRONMENT

Various techniques are developed for load balancing in cloud and distributed environment. Godfrey B. [1] proposed a algorithm in dynamic peer to peer system and other hybrid environment. In this methodology a concept of Virtual Server with Greedy heuristic algorithm is used to schedule the load information and proper utilization of nodes. In P2P system objects and space of nodes are always subject to change due to changes in nodes infrastructure. This impact the performance of Node, Thus this methodology used the Virtual Server methodology. In this method load information of various peer nodes is stored in various directories and then reassignment of processing is done to virtual servers. Large number of Virtual Server helps in maximum utilization thereby load is spitted in to virtual server. A Virtual server is like peer but one physical node may be responsible for any number of virtual server. This approach increase the scalability and helps to utilize the node capacity at the maximum possible potential. The key problem arise in this method is reassignment of virtual server is quite complex and difficult.

Further Yunhua Deng, Rynson W.H. Lau [2] proposed Heat Diffusion based Dynamic Load Balancing model in Distributed Virtual Environment (DVE's). This approach focus on problem of Load Balancing when number of users are accessing the data. Thus Efficient Cell Selection Scheme is implemented for appropriate cell selection and load balancing through Load migration from one node to another node. In this algorithm virtual environment is parted in various square cells with object. Every node in the cell transfer node to neighbouring node after every iteration. In this algorithm heat transfer diffusion concept of transferring heat from high to low object happen when two objects are place nearby. Same concept is put forward in the algorithm proposed and high loaded node transfer the load to neighbour node and it is repeated per iteration. In this technique two types of algorithm are developed as Local Diffusion and Global Diffusion algorithm. In Load diffusion algorithm simple caparison between nodes is carried out and load is transferred. In Global Diffusion algorithm scheduling is carried out first and then load transfer. From various research carried out Global Diffusion algorithm provide optimal results. The approach provided is simple and need least complexity regarding computation and provide scalability with load balancing. But due top lot of iterations chances for network delay subs its and it is time consuming.[2]

Going ahead Markus [3] explored the concept of Overlay network in virtual online world environment. With global scenario changing constantly and concept of 3D Web are in process to outcome various decentralised techniques are getting importance. Markus hereby introduced the concept of self organised & Scale free overlay network for interconnected machines in global online environment. This system developed Hyper verse architecture, responsible for the proper hosting of the virtual world. Through load balancing methods world surface is subdivided in to small cells and managed by a public server. The Public server compute the absolute mass of the object in the cell. The proposed system cannot transfer the node load but find out the links assigned to each node in network formation. Due to this application network becomes efficient, reliable and fault tolerant. As Public server are randomly placed substantial time is required for load balancing.

B.Dong [4] came up with Self Acting Load Balancing Algorithm for parallel file system based on distributed architecture. In Parellel file system Data management plays vital role as data is constantly transferred between memory and storage devices. Various factors like scalability, availability of system, Load Migration and Data Transmission followed by continuous workload variation. Thus this calls for central decision maker and to explore decision making algorithm. Thus one node act as decision maker and others follow the assigned task accordingly. If central node fails then total system will fail hence this method is not followed dominantly. This mechanism provided load balancing mechanism but as task migration will frequently takes place, possibility of

system degradation increases. Self Acting Balancing Algorithm (SALB) provides better load forecasting algorithm, Load collection, Distribution and migration on in effective manner. [5]

Dhinesh Babu [5] proposed another Load Balancing mechanism based on Honey Bee Behaviour. The method is derived from behaviour of Honey bees. The honey bees are of two types as scout bees and forager bees. Scout bees find out the food and gives message to forager bees through dance. After this forager bees follow the scout bees and track the location of food. Same concept of utilised for load balancing and overloaded Virtual Machine (VM) act as honeybees. On assignment of any task the under loaded Vm will update the information and if high loaded or priority task is submitted for processing it will be assigned to VM having minimum number of high priority task. As the task assignment is on the basis of ascending order under loaded VM's will get the high priority task and thus load balancing is carried out. Load of specific server is computed on the basis if information accessed from Data centre/server. Honey bee approach reduced waiting time and high priority task are processed quickly and scalability is achieved without further extension of resources. This suffers from drawback as high priority task are processed first it forms que and low priority task remain in queue for further processing for quite long period.[5]

Raj, G. Punjab Tech. Univ., Jalandhar, India   Singh, D. ; Bansal, A [6] .proposed a mechanism in which load balancing is managed to achieve better response time. The technique transfers load of one server to another server in round robin fashion and then scheduling is done in jobs. Round-Robin is scheduling algorithm, based on time sharing among jobs in equal slice / quantum and in circular queue without priority so it is simple and easy to implement, it focuses on fairness between jobs. The advantage of this algorithm is that no job has to wait for another one to be completed. Again this methodology is not a good choice for jobs that are largely varies in their size and requirements, by means, a job is never been satisfied which in turn leads to starvation or indefinite blocking. With Round - Robin scheduling we have proposed a modified approach of load balancing mechanism which is a combination of 2 types of scheduling i.e. round robin scheduling and Batch Mode Heuristic priority Algorithm. Thus this technique is used to enhance the round robin scheduling and boosted the system by adding Heuristic Priority to optimize throughput. [6]

Pedro Martins, Maryam Abbasi, Pedro Furtado [7] further extended the architecture for load balancing and scaling with queries and heavy data processing. The paper focus on identifying overload problem, rescheduling queries controlling query processing and attaining scalability. This paper proposed an automated method for scalability, elasticity, shedding policies and   admission control. Thus this paper tries to integrate all the factors essential for speedy processing of queries without adding up of additional nodes. It aims at attending maximum throughput with available nodes and resources. It is based on the concept of complex event processing. Scheduler is used for queuing the queries and passing them to nodes specified as per load criteria. In this technique algorithms are developed to attain the scalability with high rate CEP. Rescheduling of queries
is possible through Round Robin, Least Work or least weight mechanism. Automatic dynamic balancing with least weight supports a much higher throughput and provides better results. This technique provides integrated mechanism to work in stream workload environment.
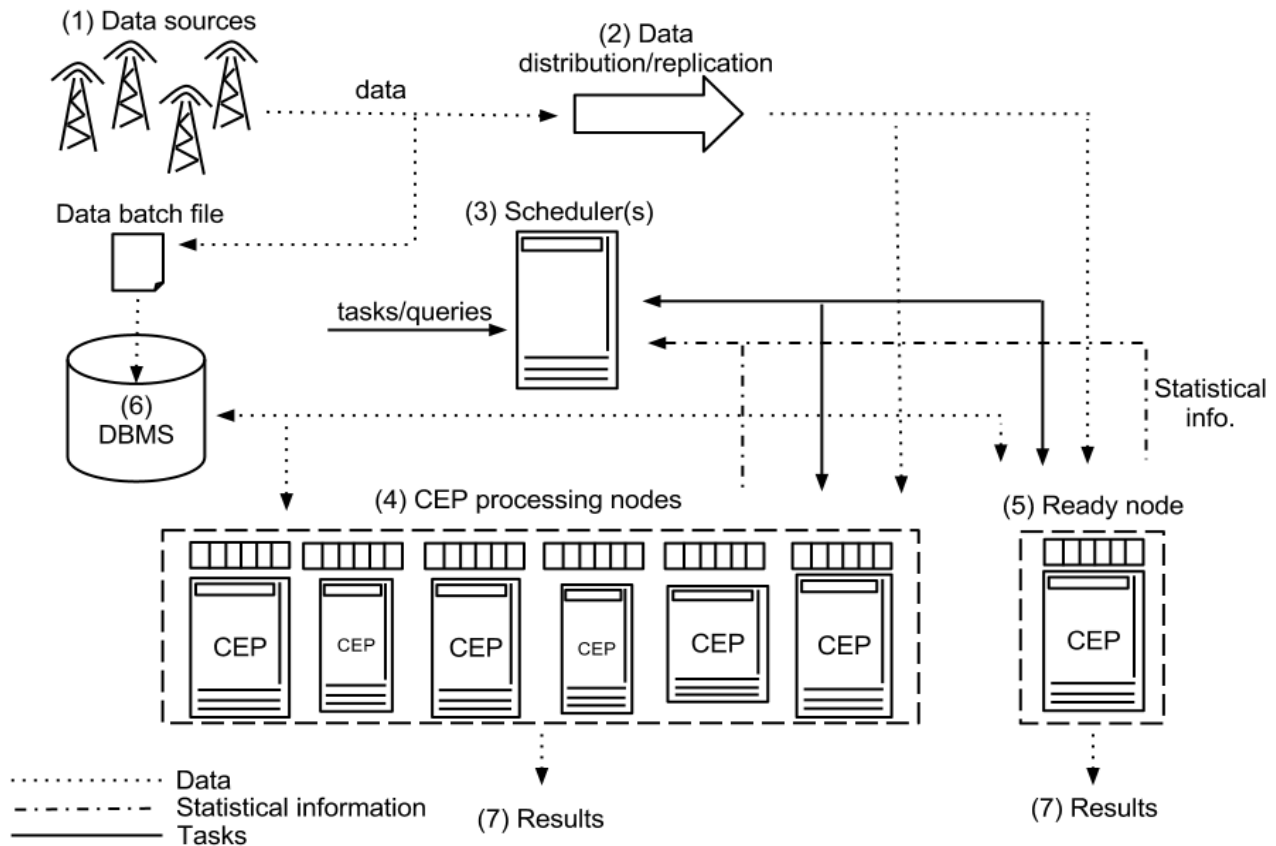
Fig 1. AuDy Architecture

III. ADVANTAGES OF DYNAMIC LOAD BALANCING

- **Redundancy**

Redundancy describes the method of running two or more identical servers, providing a failsafe in the event that one server becomes unavailable. With setups that don't use load balancers, it's difficult to ensure redundancy, because each server in the cluster is essentially a single point of failure. For instance, if a server becomes unresponsive due to a rogue process, all the users served by that machine will get an error message and most likely, they will flood support channels with complaints. With load balancing in place, the system "senses" when a server becomes unavailable and instantly reroutes traffic to the other machines, ensuring that another node in your cluster handles the traffic. Even better, our load balanced architecture always includes a secondary load balancing device, guaranteeing full redundancy across every component of your network.

- **Scalability**

Even if you have modest resource requirements right now, scalability should always be a consideration when finding the right hosting solution. In a few months, you may find that you need to add more horsepower to your cluster. The smart thing to do is to opt for a load balancing system, which not only lets you get the most out of your current machines, but lets you scale your application when needed. You can have a new machine (or ten new machines) go live, without having to take the entire application out for maintenance.

- **Resource optimization**

With a load balancer in place, you can optimize how traffic is distributed to the server cluster to ensure the best performance possible. Load balancers can also speed up an application, by taking over time-consuming operations such as SSL encryption and decryption. The specialized components in F5's BIG-IP products decrypt the web traffic before dispatching it to the cluster, thus freeing up resources on servers and reducing overhead.

- **Security**

The security of the web application is also improved dramatically. With load balancing, only one IP is exposed to the web, which greatly reduces the number of breach points in case of an attack. The internal topology of your network is hidden, which also improves the safety of your setup. The servers from the cluster receive their own virtual IP, which makes it possible for the load balancer to direct traffic as needed without exposing your address to hackers. This adds an additional layer of security onto our 5-tier security platform.

IV. ANALYSIS & FUTURE SCOPE

The Automated Load balancing mechanism [7] for stream workload scalability is distributes workload among various nodes, thereby enhancing scalability and throughput time. It is very much useful in Distributed Database to achieve better results for query processing. Technology can be utilized in telecommunication corporate wherein extensive communication links and external networks are established. As use of cloud technology is at peak, the methodology will enhance the turnaround time to process the query and additional cost on nodes can be optimized by utilizing the inherent capacity of nodes.

In future focus can be extended to cover more parameters and to develop integrated approach for query processing in dynamic environment. Heavy Queries can be processed with lowest turnaround time and research can be directed to take quick measures to node crash and reducing disaster recovery time and achieving results. This technique can be extended with large data warehouses and cloud services on large scale. Enhancing scalability and reducing throughput time with inbuilt testing mechanism with backup emergency plan can be worked out with more relevant parameters such as storage capacity, processing parameters and many more factors which plays significant role in network designing, distribution of database, processing of queries and cloud technology.

V. CONCLUSION

In this paper we have studied Various approaches which are proposed to increase scalability of query processing with automated architecture for overload rescheduling, automated elasticity, Query processing control, Alerts in case of overload situation and node failure. Various researchers have to put forward mechanism for load balancing in networking and cloud environment. On the basis of Survey conducted and research literature available so far *Automatic Dynamic Least-Weight Balancing for Stream Workloads Scalability* approach provides unique and integrated approach and considers many factors in unison to provide best possible results.

VI. REFERANCES

[1] *Godfrey, B. ; Lakshminarayanan, K. ; Surana, S. ; Karp, R, "Load balancing in dynamic structured P2P systems", IEEE.*

[2] *Yunhua Deng, Rynson W.H. Lau, "Heat diffusion based dynamic load balancing for distributed virtual environments", in: Proceedings of the17th ACM Symposium on Virtual Reality Software and Technology, ACM, 2010.*

[3] *Markus Esch, Eric Tobias, "Decentralized scale-free network construction and load balancing in Massive Multiuser Virtual Environments", in: Collaborative Computing: Networking, Applications and Worksharing, Collaborate Com, 2010, 6th International Conference on, IEEE, 2010.*

[4] *Bin Dong, Xiuqiao Li, Qimeng Wu, Limin Xiao, Li Ruan, "A dynamic and adaptive load balancing strategy for parallel file system with large-scale I/O servers", J. Parallel Distribution Computing,2012.*

[5] *Dhinesh Babu L.D, P. VenkataKrishna, "Honey bee behavior inspired load balancing of tasks in cloud computing environments",Applied Soft Computing 13,2013.*

[6] Raj, G. Punjab Tech. Univ., Jalandhar, India   Singh, D. ; Bansal, A. *"Load balancing for resource provisioning using Batch Mode Heuristic Priority in Round Robin (PBRR) Scheduling"* IEEE Confluence 2013: The Next Generation Information Technology Summit (4th International Conference).

[7] Pedro Martins, Maryam Abbasi, Pedro Furtado , "AuDy: Automatic Dynamic Least-Weight Balancing for Stream Workloads Scalability", 2014 IEEE International Congress on Big Data.

[8] U. Karthick Kumar , "A Dynamic Load Balancing Algorithm in Computational Grid Using Fair Scheduling", IJCSI International Journal of Computer Science Issues, Vol. 8, September 2011.