

LOAD BALANCING OF VIRTUAL MACHINE SERVERS IN CLOUD COMPUTING USING GENETIC-BASED APPROCHES FOR LOAD PREDICTION

Kannan M, M.E¹, Konduru Revanth Reddy², Dasariraju Saiteja³, Konuri Harichandra Prasad⁴

¹Professor, Department of Computer Science and Engineering, Mahendra Engineering College, Mahendrapuri,

^{2,3,4}UG Scholar (B.E), Department of Computer Science and Engineering, Mahendra Engineering College, Mahendrapuri

ABSTRACT: This paper offers a two-stage genetic mechanism for the migration-based load stability of digital computing device hosts (VMHs) in cloud computing. Previous techniques generally expect this difficulty as a job-assignment optimization trouble and solely reflect on consideration on the modern-day VMHs' loads; however, except thinking about hundreds of VMHs after balancing, these techniques can solely achieve confined effectiveness in actual environments. In this study, two genetic-based strategies are built-in and presented. First, overall performance fashions of digital machines (VMs) are extracted from their developing parameters and corresponding overall performance measured in a cloud computing environment. The gene expression programming (GEP) is utilized for producing symbolic regression fashions that describe the overall performance of VMs and are used for predicting masses of VMHs after load-balance. Secondly, with the VMH masses estimated via GEP, the genetic algorithm considers the present day and the future hundreds of VMHs and decides an most beneficial VM-VMH undertaking for migrating VMs and performing load-balance. The overall performance of the proposed strategies is evaluated in a actual cloud-computing environment, Jnet, whereby these techniques are applied as a centralized load balancing mechanism. The experimental effects exhibit that our approach outperforms preceding methods, such as heuristics and records regression.

Key Words: Cloud Computing, Virtualization, Load Balancing, Migration, Genetic Algorithm, Gene Expression Programming.

INTRODUCTION:

Cloud computing is an Internet-based useful resource utility [1], [2]. Its central thought is "everything can be a service". In cloud computing, computing hardware and software program sources are capsulized as web-services that can be accessed via the Internet. Three software sorts of cloud computing fashions are described [3]: infrastructure as a provider (IaaS), platform as a provider (PaaS), and software program as a carrier (SaaS). Among them, virtualization is a consultant IaaS application, which affords computing infrastructure resources, such as computing power, records storage, networking, all in the structure of internet offerings [4], [5]. IaaS vendors buy and preserve bodily computing and storage

hardware and grant internet offerings to users. With the virtualization technology, the customers request IaaS vendors for computation or storage sources as they personal a "virtual machine"(VM) barring buying and retaining bodily hardware. The customers can make use of VMs for deploying system/application software program with a drastically decrease price of hardware procurement and possession.

An IaaS issuer operates a server farm consisting of some computing and storage hardware, whereby some web hosting servers (referring to as VMHs) furnish virtualization services. A VMH can also run one or many VMs, relying on the capability of the VMH. If VMHs are now not excellent managed in a server farm, some VMHs may additionally be busy jogging many VMs however some VMHs nearly idle with few VMs. Managing masses of VMHs through adjusting the consumption of sources held by way of VMs for higher cost-performance effectivity whilst guaranteeing carrier stage agreements (SLA) to the clients is an fundamental trouble in IaaS.

Live migration with shared storage:

Load-balance is a standard however integral lookup theme in parallel or dispensed systems, and many computer learning-based techniques have been proposed, such as [10]–[11][12]. The following discusses some current strategies for load balancing in cloud computing. Alonso-Calvo et al. introduced in [13] an application-level load balancing device for functions strolling on VMs. Doong et al. introduced in [14] a multi-kernel help vector regression mannequin for modeling VMs performance. Hu et al. developed a scheduling approach for load balancing of VM assets the usage of GA that refers to historic facts and the present day country of the device [15]. A capability allocation algorithm is introduced in [16] that coordinates a couple of allotted aid controllers executing in geographically allotted cloud sites. Pang et al. introduced in [17] a hybrid technique that employs the estimation of distribution algorithm to estimate viable options of VM hundreds and then makes use of GA to modify these solutions.

HEELS [18] is a heuristic undertaking deployment strategy primarily based on clustering and Glowworm Swarm Optimization and used for long-term load balancing of a cloud framework with side computing. A hybrid metaheuristic is proposed in [19] that hybridizing artificial bee colony and ant colony optimization for load balancing of VMs in the cloud.

LITERATURE SURVEY:

Mohammed Ala'Anzy and Mohamed Othman described an environment friendly approach for load balancing and server consolidation in cloud computing environments: a meta-study. The data-center is viewed the coronary heart of cloud computing. Recently, the developing demand for cloud computing offerings has precipitated a developing load on information centers. In phrases of machine conduct and workload, patterns of cloud computing are very dynamic; and that may serve to imbalance the load amongst statistics middle resources. Eventually, some data-center assets may want to come to be over-loaded/under-loaded, which leads to an make bigger in power consumption in addition to lowered functioning and wastage of resources. Just thinking about energy-efficiency (that can be attained correctly via consolidate the servers) may additionally now not be adequate for actual functions due to the fact it might also motive troubles such as unbalanced load for every Physical Machine (PM). Therefore, this paper surveys posted load balancing algorithms that finished with the aid of server consolidation by using a meta-analysis. Load balancing with server consolidation enriches the exploitation of useful resource utilization and can beautify Quality of Service (QoS) metrics, considering the fact that data-centers and their purposes are growing exponentially. This meta-study, opinions the literature on load balancing and server consolidation and gives a equipped reference taxonomy on the most environment friendly algorithms that reap load balancing and server consolidation. This work tries to current a taxonomy with a new classification for load balancing and server consolidation, such as migration overhead, hardware threshold, community traffic, and reliability.

Jinglin Li, Guiyang Luo and Nan Cheng introduces an end-to-end load balancer primarily based on deep gaining knowledge of for vehicular community visitors control. The infrastructure to automobile (I2V) verbal exchange boosts a giant range of prevailing vehicular services, which can furnish automobiles with exterior information, storage, and computing strength positioned at each cellular aspect server (MES) and far off cloud. However, automobile distribution is variation due to the spatial inhomogeneity and temporal

dynamics. As a consequence, the conversation load for MES is imbalanced and automobiles may also go through from negative I2V communications the place the MES is overloaded. In this paper, we endorse a novel proactively load balancing method that permits environment friendly cooperation amongst MESs, which is referred to as end-to-end load balancer (E2LB). E2LB schedules the cached records amongst MESs based totally on the anticipated avenue visitors situation. First, a convolutional neural community (CNN) is utilized to correctly study the spatio-temporal correlation in order to predict the street visitors situation. Then, we formulate the load balancing hassle as a nonlinear programming (NLP) hassle and a novel framework primarily based on CNN is adopted to approximate the NLP optimization. Finally, we join the above neural networks into an end-to-end neural community to together optimize the performance, the place the enter is the historic site visitors scenario whilst the output is the balanced scheduling solution. E2LB can warranty the real-time scheduling, considering the fact that the calling of a well-trained neural community solely requires a small wide variety of easy operations. Experiments on the trajectories of taxis and buses in Beijing reveal the effectivity and effectiveness of E2LB.

Haipeng Yao, Xin Yuan and Peiying Zhang introduced desktop mastering aided load stability routing scheme thinking about queue utilization Due to the speedy improvement of community techniques, packet-switched structures journey high-speed extend of traffic, which imposes a load and unbalanced burden on the routers. Hence, environment friendly routing schemes are required in order to gain load balance. By decoupling the manage airplane and the records plane, Software-Defined Network (SDN) suggests its flexibility and extensibility to reap the automated administration of community resources. Based on the SDN architecture, we suggest a pair of computer gaining knowledge of aided load stability routing schemes thinking about the queue utilization (QU), which divide the routing procedure into three steps, specifically the dimension reduction, the QU prediction, as nicely as the load stability routing. To the fantastic of our knowledge, it is the first time that important aspect evaluation (PCA) is used for the dimension discount of the substrate network. Furthermore, QU prediction is performed with the resource of neural community algorithms for the sake of coping with the community congestion ensuing from burst traffic. Finally, simulation consequences exhibit that our proposed routing schemes thinking about QU estimated with the aid of the computing device getting to know algorithms outperform the typical Bellman-Ford (BF) routing approach in phrases of the

common packet loss ratio, the worst throughput and the common extend.

SYSTEM STUDY

FEASIBILITY STUDY

The feasibility of the undertaking is analyzed in this segment and enterprise notion is put forth with a very ordinary format for the task and some fee estimates. During device evaluation the feasibility learn about of the proposed gadget is to be carried out. This is to make certain that the proposed device is now not a burden to the company. For feasibility analysis, some appreciation of the main necessities for the device is essential.

Three key considerations involved in the feasibility analysis are

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

ECONOMICAL FEASIBILITY

This find out about is carried out to test the monetary have an effect on that the device will have on the organization. The quantity of fund that the agency can pour into the lookup and improvement of the device is limited. The costs need to be justified. Thus the developed machine as nicely inside the price range and this used to be finished due to the fact most of the applied sciences used are freely available. Only the personalized merchandise had to be purchased.

TECHNICAL FEASIBILITY

This learn about is carried out to take a look at the technical feasibility, that is, the technical necessities of the system. Any machine developed ought to now not have a excessive demand on the handy technical resources. This will lead to excessive needs on the reachable technical resources. This will lead to excessive needs being positioned on the client. The developed gadget ought to have a modest requirement, as solely minimal or null modifications are required for imposing this system.

SOCIAL FEASIBILITY

The element of find out about is to test the degree of acceptance of the device via the user. This consists of the

procedure of education the consumer to use the machine efficiently. The person should now not experience threatened through the system, as an alternative should take delivery of it as a necessity. The degree of acceptance via the customers completely relies upon on the strategies that are employed to train the person about the device and to make him acquainted with it. His stage of self belief ought to be raised so that he is additionally in a position to make some positive criticism, which is welcomed, as he is the closing person of the system.

SOFTWARE ENVIRONMENT

NetBeans

NetBeans is an built-in improvement surroundings (IDE) for Java. NetBeans permits purposes to be developed from a set of modular software program elements referred to as modules. NetBeans can be used on Windows, macOS, Linux and Solaris. In addition to Java development, it has extensions for different languages like PHP, C, C++, HTML5, and JavaScript. Applications primarily based on NetBeans, such as the NetBeans IDE, can be prolonged by way of 0.33 birthday party developers.

The NetBeans Platform is a framework for simplifying the improvement of Java Swing laptop applications. The NetBeans IDE bundle for Java SE includes what is wanted to begin growing NetBeans plugins and NetBeans Platform based totally applications; no extra SDK is required.

Applications can set up modules dynamically. Any utility can consist of the Update Center module to permit customers of the software to down load digitally signed improvements and new aspects immediately into the going for walks application. Reinstalling an improve or a new launch does no longer pressure customers to down load the complete utility again.

The platform affords reusable offerings frequent to computing device applications, permitting builders to focal point on the common sense unique to their application. Among the features of the platform are:

- User interface management (e.g. menus and toolbars)
- User settings management
- Storage management (carries out efficient storage)
- Window management
- Wizard framework (supports step-by-step dialogs)

- NetBeans Visual Library
- Integrated development tools

NetBeans IDE

NetBeans IDE is an open-source built-in improvement environment. NetBeans IDE supports improvement of all Java utility kinds (Java SE (including JavaFX), Java ME, web, EJB and cellular applications) out of the box. Among different elements are an Ant-based challenge system, Maven support, refactorings, model manipulate (supporting CVS, Subversion, Git, Mercurial and Clearcase).

This tutorial affords a very easy and speedy introduction to the NetBeans IDE workflow through strolling you via the advent of a easy "Hello World" Java console application. Once you are accomplished with this tutorial, you will have a typical expertise of how to create and run purposes in the IDE. This tutorial takes much less than 10 minutes to complete.

After you end this tutorial, you can go on to the mastering trails, which are linked from the Documentation, Training & Support page. The getting to know trails furnish complete tutorials that spotlight a wider vary of IDE aspects and programming strategies for a variety of utility types. If you do no longer choose to do a "Hello World" application, you can pass by this tutorial and bounce straight to the mastering trails.

NetBeans IDE Bundle for Web and Java EE

The NetBeans IDE Bundle for Web & Java EE provides complete tools for all the latest Java EE 6 standards, including the new Java EE 6 Web Profile, Enterprise Java Beans (EJBs), servlets, Java Persistence API, web services, and annotations. NetBeans also supports the JSF 2.0 (Facelets), JavaServer Pages (JSP), Hibernate, Spring, and Struts frameworks, and the Java EE 5 and J2EE 1.4 platforms. It includes GlassFish and Apache Tomcat. Some of its features with javaEE includes

- Improved support for CDI, REST services and Java Persistence
- New support for Bean Validation
- Support for JSF component libraries, including bundled PrimeFaces library
- Improved editing for Expression Language in JSF, including code completion, refactoring and hints

MySQL

MySQL, the most famous Open Source SQL database administration system, is developed, distributed, and supported by using Oracle Corporation. MySQL is an Oracle-backed open supply relational database administration device (RDBMS) based totally on Structured Query Language (SQL). MySQL runs on surely all platforms, together with Linux, UNIX and Windows. Although it can be used in a broad vary of applications, MySQL is most regularly related with net functions and on line publishing. MySQL is an vital issue of an open supply corporation stack referred to as LAMP. LAMP is a net improvement platform that makes use of Linux as the running system, Apache as the net server, MySQL as the relational database administration machine and PHP as the object-oriented scripting language.

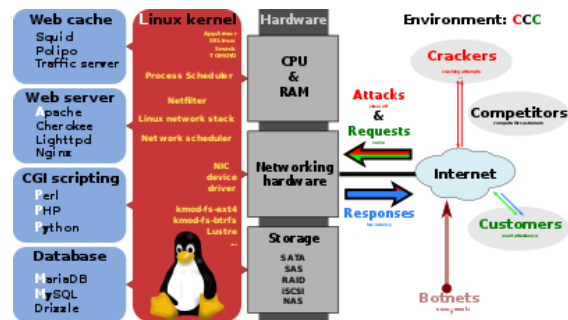


Fig 1: System Architecture

Open Source capability that it is feasible for every person to use and alter the software. Anybody can down load the MySQL software program from the Internet and use it except paying anything. If you wish, you may additionally find out about the supply code and alternate it to go well with your needs. The MySQL software program makes use of the GPL (GNU General Public License), to outline what you may additionally and may additionally no longer do with the software program in special situations. If you sense uncomfortable with the GPL or want to embed MySQL code into a industrial application, you can purchase a commercially licensed model from us. See the MySQL Licensing Overview for greater records.

MY SQL IS A DATA BASE MANAGEMENT SYSTEM:

A database is a structured series of data. It can also be whatever from a easy purchasing listing to a image gallery or the great quantities of statistics in a company network. To add, access, and manner facts saved in a laptop database, you want a database administration machine such as MySQL Server. Since computer systems are very correct at dealing

with massive quantities of data, database administration structures play a central function in computing, as standalone utilities, or as components of different applications.

MYSQL DATABASES ARE RELATIONAL:

A relational database shops information in separate tables instead than placing all the records in one large storeroom. The database buildings are equipped into bodily archives optimized for speed. The logical model, with objects such as databases, tables, views, rows, and columns, gives a bendy programming environment. You set up regulations governing the relationships between one of a kind information fields, such as one-to-one, one-to-many, unique, required or optional, and "pointers" between distinct tables. The database enforces these rules, so that with a well-designed database, your utility by no means sees inconsistent, duplicate, orphan, out-of-date, or lacking data. The SQL section of "MySQL" stands for "Structured Query Language". SQL is the most frequent standardized language used to get admission to databases. Depending on your programming environment, you may enter SQL without delay (for example, to generate reports), embed SQL statements into code written in every other language, or use a language-specific API that hides the SQL syntax.

HOW MYSQL WORKS

MySQL is primarily based on a client-server model. The core of MySQL is MySQL server, which handles all of the database directions (or commands). MySQL server is reachable as a separate software for use in a client-server networked surroundings and as a library that can be embedded (or linked) into separate applications.

MySQL operates alongside with countless utility applications which assist the administration of MySQL databases. Commands are despatched to MySQL Server by means of the MySQL client, which is set up on a computer.

MySQL used to be initially developed to deal with massive databases quickly. Although MySQL is usually established on solely one machine, it is capable to ship the database to a couple of locations, as customers are capable to get entry to it by way of extraordinary MySQL customer interfaces. These interfaces ship SQL statements to the server and then show the results.

A LARGE AMOUNT OF CONTRIBUTED MYSQL SOFTWARE IS AVAILABLE

MySQL Server has a sensible set of facets developed in shut cooperation with our users. It is very probable that your favourite software or language helps the MySQL Database Server. The reliable way to pronounce "MySQL" is "My Ess Que Ell" (not "my sequel"), however we do now not idea if you pronounce it as "my sequel" or in some different localized way.

CORE MYSQL FEATURES

MySQL allows information to be saved and accessed throughout a couple of storage engines, along with InnoDB, CSV, and NDB. MySQL is additionally succesful of replicating statistics and partitioning tables for higher overall performance and durability. MySQL customers are not required to research new commands; they can get right of entry to their records the use of popular SQL commands.

EXISTING SYSTEM

The load stability trouble as job mission optimization and more often than not center of attention on creating optimization algorithms for quickly convergence. However, they expect the load of VM/VMH is static and do now not reflect onconsideration on the fee of migration and the load of VMHs after balancing, ensuing in confined effectiveness in actual environments. Therefore, load balancing by means of the migration of VMs can be viewed as a combinatorial problem, whereby the fantastic set of VMs are chosen from VMHs and migrated to suited hosts stage.

PROPOSED SYSTEM

In this study, two genetic-based techniques are built-in and presented. First, overall performance fashions of digital machines (VMs) are extracted from their growing parameters and corresponding overall performance measured in a cloud computing environment. The gene expression programming (GEP) is utilized for producing symbolic regression fashions that describe the overall performance of VMs and are used for predicting masses of VMHs after load-balance. Secondly, with the VMH masses estimated through GEP, the genetic algorithm considers the modern-day and the future hundreds of VMHs and decides an most fulfilling VM-VMH mission for migrating VMs and performing load-balance. GA offers an intuitive and speedy approach of identifying VM-VMH assignments for load stability and flexibly works with more than a few goal features for distinctive administration purposes. The genetic-

based methods, GA and GEP, can be without problems applied via the cloud administrator. The overall performance and usability of the proposed strategies are evaluated and proved to be legitimate in a actual cloud surroundings.

IMPLEMENTATION

Implementation is the stage of the undertaking when the theoretical layout is became out into a working system. Thus it can be regarded to be the most quintessential stage in attaining a profitable new gadget and in giving the user, self assurance that the new gadget will work and be effective. The implementation stage entails cautious planning, investigation of the current machine and it’s constraints on implementation, designing of techniques to acquire changeover and contrast of changeover strategies.

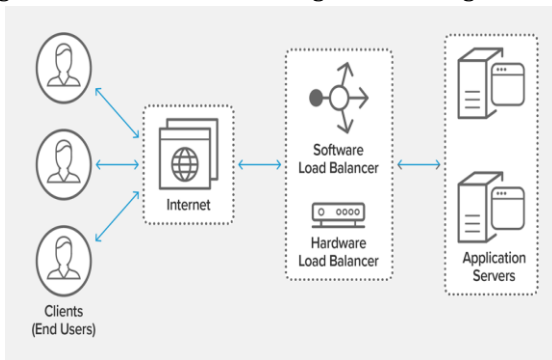


Fig 2: Implementation

MODULES:

- Data Owner
- Data User
- Semi-Trusted Server
- Trusted Authority
- Secret Sharing On Lazy Re-Encryption
- Data Access Control

DATA OWNER:

Data proprietor will have to register originally to get get admission to the profile. Data Owner will add the file to the cloud server in the encrypted format. Random encryption key technology is taking place whilst importing the file to the cloud. Encrypted file will be saved on the cloud.

DATA USER:

Data User will originally ask for the key to the Authority to confirm and decrypt the file in the cloud. Data User can get

admission to the file primarily based on the key obtained from mail id. As per the key acquired the customer can confirm and decrypt the information from the cloud.

SEMI-TRUSTED SERVER:

LDSS is designed below the identical assumptions proposed in zero that the CSP is truthful however curious, which potential that the CSP will faithfully execute the operations requested via users, however it will peek on what customers have saved in the cloud. The CSP will faithfully shop users’ data, undertake an preliminary get admission to control, replace information in accordance to users’ requests. However, CSP might also do malicious movements such as collusion with customers to get the information in simple text. In LDSS, proxy encryption server and proxy decryption server are brought to aid customers to encrypt and decrypt facts so that user-side overhead can be minimized. In essence, proxy servers are additionally machines in the cloud. Thus, we think about that they are straightforward however curious simply as the CSP.

TRUSTED AUTHORITY:

A depended on authority (TA) is introduced. It is accountable of producing public and personal keys, and distributing attribute keys to users. With this mechanism, customers can share and get admission to information except being conscious of the encryption and decryption operations. We anticipate TA is totally credible, and a relied on channel exists between the TA and each and every user. The reality that a relied on channel exists doesn’t imply that the information can be shared thru the relied on channel, for the records can be in a giant amount. TA is solely used to switch keys (in a small amount) securely between users. In addition, it’s requested that TA is on-line all the time due to the fact records customers may additionally get admission to facts at any time and want TA to replace attribute keys.

SECRET SHARING ON LAZY RE-ENCRYPTION:

Cipher textual content get right of entry to control, information wishes to be re-encrypted when some users’ get right of entry to privileges to the facts are revoked. However, typical re-encryption brings heavy computational overhead, and the accessed plaintext records may additionally already be saved on these facts users. Therefore, this paper adopts the lazy re-encryption approach proposed. With lazy re-encryption, when a user’s get entry to privilege is revoked, facts is now not re-encrypted till the statistics proprietor updates the facts the file of the get right of entry to manage

coverage that includes these attributes will be marked. Later, when the records proprietor updates this file, it first assessments the mark to see if it has been marked as revoked. If that is the case, this file will be re-encrypted.

DATA ACCESS CONTROL:

The protection of get entry to manipulate coverage is that no members should be aware of the unique content material of the get entry to manipulate coverage besides records proprietors primarily based VM. LDSS introduces attribute description subject so that get admission to manage coverage is described by means of the corresponding attribute description bit. ESP and the Cloud can solely get the relationships between one of a kind attribute description bits, however now not the unique content material of get right of entry to manipulate strategy, therefore defending the get admission to manage method.

INPUT AND OUTPUT DESIGN

INPUT DESIGN

The enter graph is the hyperlink between the records machine and the user. It consists of the growing specification and approaches for records coaching and these steps are quintessential to put transaction facts in to a usable shape for processing can be accomplished by way of inspecting the laptop to study statistics from a written or printed file or it can show up by way of having humans keying the records immediately into the system. The graph of enter focuses on controlling the quantity of enter required, controlling the errors, heading off delay, keeping off more steps and retaining the procedure simple. The enter is designed in such a way so that it gives safety and ease of use with conserving the privacy. Input Design considered the following things:

- ✓ What data should be given as input?
- ✓ How the data should be arranged or coded?
- ✓ The dialog to guide the operating personnel in providing input.
- ✓ Methods for preparing input validations and steps to follow when error occur.

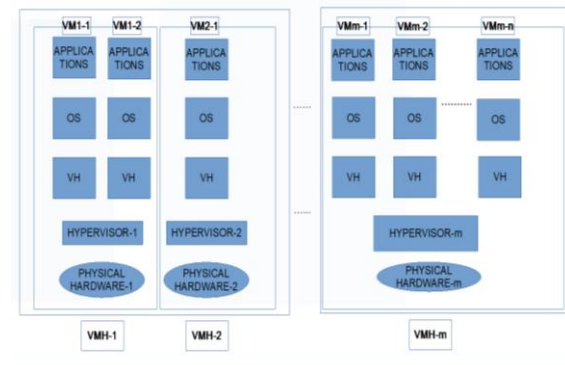


Fig 3: Input Design

OBJECTIVES

1. Input Design is the system of changing a user-oriented description of the enter into a computer-based system. This format is necessary to keep away from mistakes in the facts enter technique and exhibit the right course to the administration for getting right facts from the computerized system.
2. It is accomplished by way of growing hassle-free displays for the facts entry to deal with massive extent of data. The purpose of designing enter is to make facts entry simpler and to be free from errors. The statistics entry display screen is designed in such a way that all the statistics manipulates can be performed. It additionally gives report viewing facilities.
3. When the information is entered it will take a look at for its validity. Data can be entered with the assist of screens. Appropriate messages are furnished as when wished so that the person will now not be in maize of instant. Thus the goal of enter sketch is to create an enter design that is handy to follow

OUTPUT DESIGN

1. A fine output is one, which meets the necessities of the stop consumer and offers the facts clearly. In any machine consequences of processing are communicated to the customers and to different device via outputs. In output format it is decided how the data is to be displaced for on the spot want and additionally the tough reproduction output. It is the most essential and direct supply statistics to the user. Efficient and sensible output diagram improves the system's relationship to assist person decision-making.

2. Designing pc output have to proceed in an organized, properly thinking out manner; the proper output need to be developed whilst making sure that every output aspect is designed so that human beings will locate the machine can use effortlessly and effectively. When evaluation diagram laptop output, they ought to Identify the particular output that is wished to meet the requirements.
3. Select techniques for providing information.
4. Create document, report, or different codecs that incorporate records produced via the system.

balancing and is competitive and promising. Although the proposed method performs well for loadbalance, it can be improved in some aspects and discussed.



Fig 5: Data Owner New Register

The output form of an information system should accomplish one or more of the following objectives.

- ❖ Convey information about past activities, current status or projections of the
- ❖ Future.
- ❖ Signal important events, opportunities, problems, or warnings.
- ❖ Trigger an action.
- ❖ Confirm an action.



Fig 6: Login Page

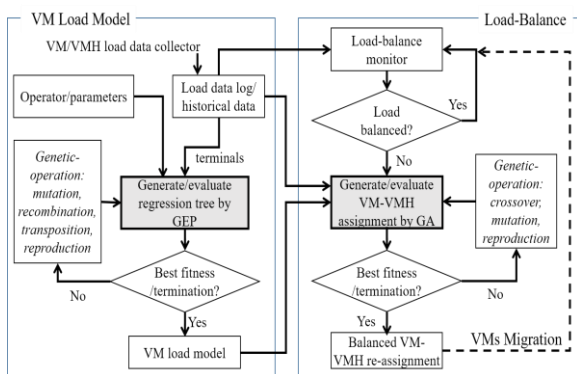


Fig 4: Data Flow Diagram

RESULTS AND WORKING:

A load balancing mechanism based on evolutionary computing. Loads of VMs and the associated resource parameters are measured and used for constructing symbolic regression models of VM using GEP. An optimal combination of VM-VMH assignment is decided by GA, which predicts VMH loads by GEP models and suggests the VMs be migrated for load-balancing. Experiments are conducted in a small-scale but real cloud environment. The proposed method demonstrates its effectiveness and efficiency on load

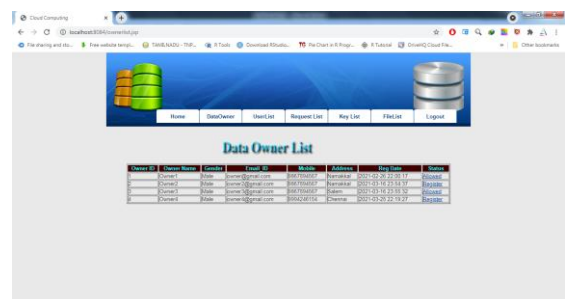


Fig 7: Data Owner List

REFERENCES

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," Communications of the ACM, vol. 53, no. 4, pp. 50– 58, Apr. 2010.

[2] P. Mell and T. Grance, "The NIST definition of cloud computing," National Institute of Standards and Technology (NIST), Gaithersburg, MD, Tech. Rep. 800-145, September 2011.

[3] T. Anderson, L. Peterson, S. Shenker, and J. Turner, "Overcoming the internet impasse through virtualization," *Computer*, vol. 38, no. 4, pp. 34–41, 2005.

[4] J. E. Smith and R. Nair, "The architecture of virtual machines," *Computer*, vol. 38, no. 5, pp. 32–38, May 2005.

[5] E. J. Ghomi, A. M. Rahmani, and N. N. Qader, "Load-balancing algorithms in cloud computing: A survey," *Journal of Network and Computer Applications*, vol. 88, pp. 50–71, 2017.

[6] R. Yadav and W. Zhang, "MeReg: Managing energy-SLA tradeoff for green mobile cloud computing," *Wireless Communications and Mobile Computing*, vol. 2017, pp. 1–11, 12 2017.

[7] M. Ala'Anzy and M. Othman, "Load balancing and server consolidation in cloud computing environments: A meta-study," *IEEE Access*, vol. 7, pp. 141 868–141 887, 2019.

[8] J. Li, G. Luo, N. Cheng, Q. Yuan, Z. Wu, S. Gao, and Z. Liu, "An End-to-End load balancer based on deep learning for vehicular network traffic control," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 953–966, Feb 2019.

[9] H. Yao, X. Yuan, P. Zhang, J. Wang, C. Jiang, and M. Guizani, "Machine learning aided load balance routing scheme considering queue utilization," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7987–7999, Aug 2019.

[10] S. H. Doong, C. C. Lai, S. J. Lee, C. S. Ouyang, and C. H. Wu, "Performance modeling of virtual machines hosted on Xen," in *Proceedings of the 2012 International Conference on Cloud Computing and Virtualization*, 2010.

[11] D. Ardagna, S. Casolari, M. Colajanni, and B. Panicucci, "Dual timescale distributed capacity allocation and load redirect algorithms for cloud systems." *Journal of Parallel and Distributed Computing*, vol. 72, no. 6, pp. 796–808, 2012.