

AI-Powered Green Cloud and Data Center

¹Savita Rohilla and ²Dr. Anju

¹Research scholar, Department computer science and engineering, Om Sterling Global University, Hisar

²Assistant Professor, Department of computer science and engineering, Om Sterling Global university, Hisar

Corresponding author email: dranju@osgu.ac.in

Abstract: As the scale of cloud computing expands, its impact on energy and the environment is becoming more and more prominent. According to statistics, data center's energy consumption has accounted for 50% of operating costs of the data centers. The rising energy consumption not only needs energy in large quantity but also imposes heavy pressure on the environment. The high energy consumption of cloud data center has become an issue, people pay close attention to, in the information technology field. It is also a problem to be solved urgently. At present, high energy consumption is caused by two reasons. First, a resource scheduling mechanism with the priority of completion time causes low server use ratio, and it is a pervasive phenomenon that small tasks take high consumption. Second, the current refrigerating system of the data center is based on peak value strategy, which causes excessive cooling supply, increases operation cost, and leads to huge waste of energy. In this paper, considering the reason of high energy consumption of the data center, a new framework of green cloud data center is put forward. Using the techniques relevant to artificial intelligence, we put forward a scheduling control engine and an intelligent refrigerating engine aiming at reduction of energy consumption. In addition, we build a green cloud data center platform, realize the scheduling control engine, and verify the feasibility of the framework. It indicates that the framework can realize a cloud platform with low power consumption and a high energy efficient data center operation.

[Rohilla, S., Vijaypal and Anju. **AI-Powered Green Cloud and Data Center**. *The International Journal of Interpretation, Observation and Analysis*, 2026; Volume 2, Issue 1 (April-June): Page Number:35-41. ISSN 2349-0713, Peer-reviewed (online/offline), Refereed, Indexed and International Journal (Since 2013), Global Impact Factor: 6.489

Key Words :- Cloud computing, Resource scheduling mechanism, Green Cloud, Waste of energy, Power consumption, Data Center, Dynamic voltage, Scheduling control, Clock frequency and Neural network

Introduction

The construction of cloud computing centers and cloud platforms is accelerating in the whole world. However, with the increase of cloud computing- oriented data centers, the energy consumption that supports data center becomes larger and larger. Energy consumption is a non negligible part of cloud computing operation costs. Besides, high energy consumption also puts tremendous pressure on the environment and energy. Statistically, the power consumption of cloud computing has been more than 1% total power consumption of the global. Energy consumption of data centers includes: continuous power supply to servers, memorizers and other basic equipment, the energy consumption for cooling infrastructure and the unavoidable small energy consumption in operation. Idle servers and high energy consumption of small tasks are ubiquitous in data center. It can be seen that the problem of energy optimization has become a hot topic. Based on the background, the concept of green cloud data center is put forward, the inevitable trend of the development of cloud data centers. Green data center optimizes the IT equipment, refrigeration equipment, illuminating

system and electric system in machine room, maximizes energy efficiency and minimizes the influence on environment. Cloud computing has become development direction of the new generation of data center. Green cloud data center has become the cutting edge of energy-saving research. Up to now, some relevant technical researches have been applied to the energy consumption optimization strategy for cloud computing. Dynamic voltage and frequency scaling adjusts voltage of power supply and clock frequency dynamically based on the current energy consumption of CPU. Virtualization technology is used in cloud platform to create several virtual machines in host, which is for decreasing the use of hardware resources and improving resource utilization rate.

At present, research on energy consumption in data centers is mainly divided into two aspects: the research on resource allocation and scheduling for raising resource utilization rate, the optimization of power supply and refrigerating system of data center for lowering the operation and maintenance cost of data center. The former research is based on traditional scheduling for improvement and optimization. Scheduling control problem is a

common research. In many application scenarios, resource scheduling and task allocation issues need to be considered. For example, literature puts forward a dynamic scheduling mechanism based on game theory. In a dynamic programming-based solution has been proposed to schedule the computation offloading from vehicles to clouds. In the iterative energy minimum algorithm is used to make scheduling decisions for mobile agent paths. Literature puts forward a new dynamic resource scheduling mechanism in IoT factories. In, the resource scheduling is carried out by using the method of dual-target mixing. Literature uses particle swarm optimization and K-means algorithm to obtain the best service combination. The quantity of researches on optimization of power supply system and heat-removal system is small. Besides, power supply system is relevant to the quantity of equipment operated. The main researches are relevant to the optimization of refrigerating system. The control strategy of the cooling system in the data center is an important direction of energy optimization. The literature addresses the technical and economic issues associated with refrigeration system in data centers through the use of absorption cooling machines. Heuristic optimization algorithm is used. In cloud data center base, electric equipment, refrigerating system and power distribution constitute a mutually related system. Involving complex dynamics principle, traditional method is incapable. At present, it is a hotspot to solve complex problems by introducing AI technology. The author of literature used deep learning to explore the occurrence regularity of

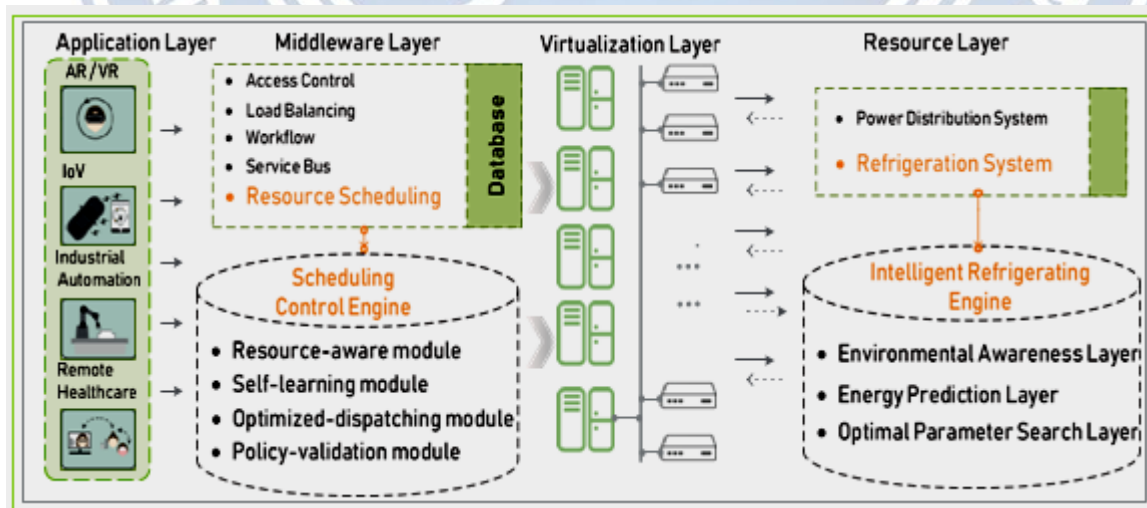
chronic diseases and predicted risks of disease based on convolution neural network. A new cache mode proposed in proposes an intelligent content distribution mechanism in high-density networks.

Architecture of AI-Enable Green Cloud

Starting from request workflow of users, traditional cloud computing is comprised of application layer, middleware layer, virtual layer and resource layer. Based on the original architecture of cloud computing, we put forward AI-enable green cloud, including scheduling control engine and intelligent refrigerating engine. Fig shows the architecture of AI-enable green cloud. Existing advanced communications and networking technologies such as can be deployed for connecting massive mobile terminals and clouds in the system. The two engines are deployed at the middleware layer and resource layer of cloud computing respectively. The functional modules of scheduling control engine and intelligent refrigerating engine are described in details.

Scheduling Control Engine

Resource scheduling is the key issue of data center. It directly influences the efficacy and cost of cloud system. To cope with the uncertain request of users and burst of networking flow in cloud computing, the current resource scheduling makes decision aiming at high efficiency of completion or maximizing resource utilization, which incurs low efficiency of energy use and low resource utilization rate. Energy is wasted seriously. Physical servers in large quantity for storing and processing data are not used reasonably.



The architecture of ai-anable green cloud

resource perception module.

Because the response to users request is in real time, in-service learning of prediction model is not realistic. Therefore, learning of prediction model is in self-learning module. Resource perception module records the current resource information and predicted load information as data. The actual load information can be got within short time as label. The data group containing data and corresponding label is a sample. Further, with the increase of sample data, experience base can be build. The prediction model is constantly trained using the samples to make the model have more and more experience. In case of sparse request task, the model in scheduling optimization module can be updated. Namely, better model is sent to scheduling module to replace the existing model.

Traditional scheduling algorithms include First Come First Served, priority of short-time operation etc. Shift in turn and priority are the resource planning methods targeting time. Scheduling optimization module allocates resources to data center based on completion time and multi-target scheduling scheme for energy consumption. Besides, the task scheduling of cloud platform is complex and frequent. Heuristic algorithm can offer a feasible solution to living example of combination optimization problem to be solved at reasonable cost. However, owing to the complex and dynamic environment of cloud platform, it is very slow to search the optimal combination. Reinforced learning is used to learn each scheduling. Accumulate experience constantly and use reinforced learning to initialize combination optimization. In this way, the heuristic algorithm can accelerate its search and make a wise decision.

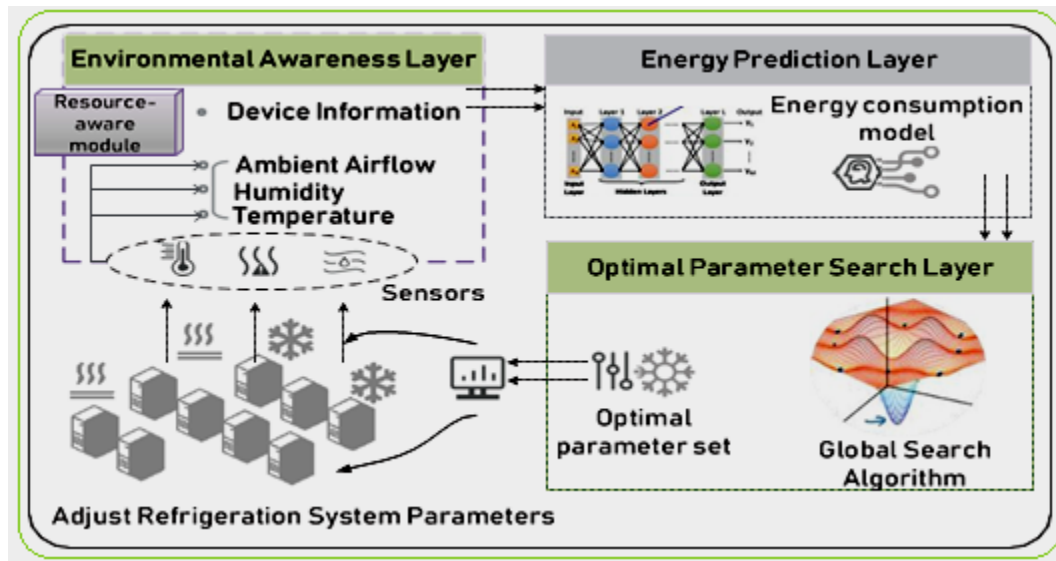
Strategy verification module is the last step before scheduling strategy is executed. Before scheduling optimization module is trained to be strong enough, as for the whole data center, it is extremely dangerous to directly use it. Therefore, it shall be verified before executing scheduling scheme. The verification module uses basic limitation conditions to verify the scheduling strategy and guarantee the reliability of scheduling engine. For example, the total computing power consumed for the verification module to check existing tasks and pre-assigned tasks of computing node shall not be more than the current usable computing power of the resource node. Before scheduling strategy controls resources, verify it to guarantee reliable work of data center. If the

scheduling optimization model is not strong enough, the traditional scheduling algorithm shall be used firstly until the scheduling optimization model becomes strong enough.

Intelligent Refrigerating Engine

Data center centrally processes, stores, and calculates data through the cluster server. These nodes generate a lot of heat when performing computing tasks. If the data center does not dissipate heat in time, the high-temperature environment of the equipment room will reduce the computing power of the physical equipment. As a high-density equipment cluster, cloud data center produces heat at high level. To guarantee the timely service completion of cloud computing, refrigerating system is crucial for the whole data center. In the whole energy consumption of data center, the energy consumption of refrigerating system accounts for 1/3 of the total energy consumption. Refrigeration is related to the heat dissipation of server and the environment of machine room. Energy consumption of data center cannot be further optimized by only depending on hardware optimization and traditional cooling control strategy. Considering the mutual relevance and complex coupling of the refrigeration control system in cloud data center, the intelligent refrigerating engine in this paper uses deep learning to explore key elements of energy consumption, predict energy consumption and build intelligent cloud refrigeration engine. The detailed structure is shown in Figure and every functional layer is described in details below.

Before intelligent refrigeration of the intelligent refrigerating engine for data center, it needs to ascertain the surrounding environmental information. Environment perception layer gets environmental information from data center in real time, including temperature, moisture, airflow, the size of machine room etc. At the same time, it gets the status of resources and equipment from the resource perception module in scheduling engine. Then all information is integrated as the input data at optimal parameter search layer. The size and structure of machine room is fixed and can be set statically during deployment. As for temperature, moisture and other dynamic information can be got after corresponding sensors are installed in machine room. Environment perception layer gets the dynamic data from sensor and resource scheduling module in real time, integrates them and transmits them to energy consumption prediction layer.



The architecture of Cloud refrigeration engine.

Energy consumption prediction layer uses the information of environment and resource status to predict the current energy consumption of data center. Traditional algorithm takes long time to observe data and analyze the degree of relevance between data and the energy consumption of data center. Perhaps the acute insight of experts is needed. Deep learning is used in this paper for energy consumption modeling without model learning. Concretely, we firstly use the data of data center to train deep neural network model. Above relevant data is same as the information got from resource perception module. The purpose of model is to get the relation between energy consumption and environment information. Energy consumption of the next time slot can be predicted based on the current environment. For example, Power Usage Effectiveness is used as the evaluation indicator of energy consumption efficiency. Energy consumption layer is the model of PUE prediction according to the current environment. Previously, PUE model uses deep learning to learn and analyze large quantity of data, explore the key relations between data and energy consumption, extract the important features of influencing energy consumption and build the mapping relation between environmental data and energy consumption indicator PUE.

The parameter optimization issue in refrigerating system is a nonlinear issue of global optimization and combined optimization. Intelligent optimization algorithm can solve the issue in short time. Therefore, this paper uses intelligent optimization algorithm to solve the optimal

parameter group in refrigerating system. In optimal parameter search, the data at environment perception layer is the input, energy consumption prediction model is the objective function of optimization algorithm, and the natural process is simulated to get the optimal parameter group. Then the optimal parameter group as control command is sent to refrigeration control system. The control system adjusts refrigeration equipment using commands. It shall be noted that refrigeration engine is different from scheduling engine and it is not motivated by computing task. Therefore, the engine uses time polling mechanism for intelligent refrigeration control. When timer is activated, environment perception layer gets the current data and predicted information from data center and resource perception layer. Then energy consumption prediction model is regarded as the objective function at the optimal parameter layer. Optimization search method is used for global optimization and getting group of optimal parameters. At last, the group of optimal parameters is transferred to control system. The control system adjusts the parameters of refrigeration equipment correspondingly.

Resource Scheduling of Cloud Data Center

In this section, we mainly introduce the LSTM-based, Long short-term memory-based predictive model and RL-based, Reinforcement learning-based, decision model related to the Scheduling control engine.

• LATM- Based Predictive Model

For global resource scheduling of cloud data center, it is necessary to estimate the load of resource node in the next time

slot. The prediction of the load of data center can be simplified to be the prediction of every resource node. It is to predict the load of single resource node in the next time slot. Therefore, this paper puts forward the prediction model of cloud data center. It can be seen as a time recurrent neural network model. RNN has strong ability of deep semantic expression and exploring the timing sequence information in data. RNN has poor prediction effect as for large change gradient of resource load. To maintain the long-term memory of RNN, we use LSTM to add some structures to RNN and avoid the dependence of prediction model on abnormal data. At the moment t , the prediction model based on LSTM :-

$$ht = \sigma(Whxt + Uhht-1 + bh)$$

$$ot = \sigma_y(Wyht + by)$$

xt is the input unit at moment t and ot is the output unit at moment t . The current status ht of RNN is determined by status of the last moment $ht-1$ and current input xt . The output unit ot is determined by the current status ht . Wh is the weight from input layer to hidden layer. U is the weight of self-circulation at hidden layer. bh means the deviations, σ_h and σ_y are activation functions. Time series data is the input data of RNN and the output data of network output layer is the prediction of series at the next moment.

• RL-BASED DECISION MODEL

Discrete Particle Swarm Optimization (DPSO) algorithm is used in this paper to search optimal solution of resource scheduling. It is assumed that there are M resource nodes and N tasks to be allocated. Resource scheduling is to allocate N tasks to M resource nodes. This minimizes the time of completing tasks and the consumption of energy. The distribution matrix and velocity matrix can be expressed as follows :-

$$X_i = [(x_{i1}, \dots, x_{iM}), \dots, (x_{iM}, \dots, x_{iM})]$$

$$V_i = [(v_{i1}, \dots, v_{iM}), \dots, (v_{iM}, \dots, v_{iM})]$$

x_{ij} is the allocation of the i^{th} task to the j^{th} server and $x_{ij} \in \{0, 1\}$. v_{ij} is the probability of the value 0 or 1 of the particle x_{ij} in the position. To make the speed of particle v_{ij} as probability, $\text{Sig}(v_{ij})$ function is used to map v_{ij} to the interval $[0, 1]$.

Conclusion

AI technology is introduced into cloud data center and an architecture of AI-enable green cloud is put forward in this paper. Besides, this paper introduces scheduling control engine and intelligent refrigeration engine. In one aspect, resources are scheduled reasonably to raise resource utilization rate, decrease the use of physical machine and decrease energy consumption. In another aspect, our work realizes the intelligent adjustment of refrigerating system considering environment and resources, which decrease the operation and maintenance cost of cloud data center. Besides, the models and algorithms relevant to scheduling control engine are introduced in details. Test bed is established and the theoretical algorithm model is verified. In experiment result, our architecture can effectively decrease energy consumption of data center. This paper introduces AI technology in cloud data center to make the cloud data center be a green cloud data center platform with 'low consumption of energy, green energy supply, intelligent energy use and high efficiency of energy conversion. In the future, we will enhance the consolidation of mobile networking with AI regarding edge computing, caching and offloading.

References

- W. Xiang, N. Wang, and Y. Zhou (2016), "An energy-efficient routing algorithm for software-defined wireless sensor networks," IEEE Sensors J., Vol. 16, No. 20, pp. 393–408.
- L. Zhou, D. Wu, J. Chen, and Z. Dong (2018), "Greening the smart cities : Energy- efficient massive content delivery via D2D communications," IEEE Trans. Ind. Informat., Vol. 14, No. 04, pp. 162–171.
- K. Bilal, S. U. Khan, and A. Y. Zomaya (2013), "Green data center networks: Challenges and opportunities," in Proc. Int. Conf. Frontiers Inf. Technol., pp. 229–234.
- L. Hou (2016), "Internet of Things cloud: Architecture and implementation," IEEE Commun. Mag., Vol. 54, No. 12, pp. 32–39.
- M. Chen, J. Yang, X. Zhu, X. Wang, M. Liu, and J. Song (2017), "Smart home 2.0 : Innovative smart home system powered by botanical IoT and emotion detection," Mobile Netw. Appl., Vol. 22, No. 6, pp. 259–272.
- M. Chen, J. Zhou, G. Tao, J. Yang, and L. Hu (2018), "Wearable affective robot" IEEE Access, Vol. 06, pp. 166–175.
- Z. Zhou, H. Zhang, X. Du, P. Li, and X. Yu (2013), "Prometheus: Privacy-aware data retrieval on hybrid cloud," in Proc. IEEE

- INFOCOM, Turin, Italy, pp. 264–277.
- Z. Guan, J. Li, L. Wu, Y. Zhang, J. Wu, and X. Du (2017), **“Achieving efficient and secure data acquisition for cloud-supported Internet of Things in smart grid,”** IEEE Internet Things J., vol. 4, no. 6, pp. 193–214.
 - Y. Xiao, X. Du, J. Zhang, F. Hu, and S. Guizani (2007), **“Internet protocol television (IPTV): The killer application for the next-generation Internet,”** IEEE Commun. Mag., Vol. 45, No. 11, pp. 126–134.
 - Y. Kessaci (2011), **“Parallel evolutionary algorithms for energy aware scheduling,”** in Intelligent Decision Systems in Large-Scale Distributed Environments. Berlin, Germany, Springer, pp. 75–100.
 - G. Wei, A. V. Vasilakos, Y. Zheng, and N. Xiong (2010), **“A game-theoretic method of fair resource allocation for cloud computing services,”** J. Super- comput., Vol. 54, No. 2, pp. 252–269.
 - M. Chen, L. T. Yang, T. Kwon, L. Zhou, and M. Jo (2011), **“Itinerary planning for energy-efficient agent communications in wireless sensor networks,”** IEEE Trans. Veh. Technol., Vol. 60, No. 07, pp. 329–341.
 - Q. Deng, D. Meisner, and A. Bhattacharjee (2012), **“CoScale : Coordinating CPU and memory system DVFS in server systems,”** in Proc. IEEE or ACM Int. Symp. Micro architecture, pp. 143–154.
 - Di Costanzo, M. D. De Assuncao, and R. Buyya (2009), **“Harnessing cloud technologies for a virtualized distributed computing infrastructure,”** IEEE Internet Computer, Vol. 13, No. 05, pp. 24–33.
 - Y. Li and M. Chen, (2015), **“Software-defined network function virtualization: A survey,”** IEEE Access, Vol. 03, pp. 254–257.