UDC 004

Telenyk Sergii, Zharikov Eduard

OPERATOR FORM TO FORMULATE, ANALYZE AND SOLVE THE CLOUD DATA CENTER IT INFRASTRUCTURE MANAGEMENT TASKS

The cloud data center IT infrastructure management tasks belongs to the management task class of nonlinear non-stationary discrete objects with variable structure. The quality of IT services and the performance of cloud data centers are directly dependent on the performance of the infrastructure layer. The non-stationarity, alteration of the structure and parameters of the management object does not allow the use of traditional management methods to ensure the provision of cloud services with specified quality. In order to manage the IT infrastructure resources of the cloud data center, the special operator form based on the definition, implementation and evaluation of management strategies in the defined mode of the data center is proposed in the article. Management strategy is determined by assessing the state of IT infrastructure using metrics. The implementation of the defined strategy is accomplished through planning and subsequent applying of basic management methods. The evaluation of the strategy's suitability is determined by the special operator based on the criterion of costs for supporting the operation of the physical infrastructure in order to further refine the choice of management strategy. Management strategies and strategies selection model for cloud data center resources management have been developed.

I. Introduction

The operation of cloud data centers is characterized by many circumstances, which are due to sectoral, geographical, international and political features of the implementation of information processes and the provision of information services. These circumstances affect the set and quality of information and cloud services [1] and also significantly affect the functionality of the IT infrastructure resource management system.

The analysis of scientific works of domestic and foreign authors showed that the IT infrastructure of cloud data centers is in the stage of intensive development and modernization [2, 3, 4, 5]. Its development requires the development of new and improvement of existing methods and technologies of IT infrastructure management for improving the efficiency of IT infrastructure management systems (ITIMS) to provide the specified quality of service delivery. The complexity of organizing the processes of interaction of elements of IT infrastructure does not allow to create their adequate models without additional restrictions, tools and industry standards. The system properties of the IT infrastructure elements have not been properly investigated.

Analysis of the basic concepts and approaches to managing the IT infrastructure of cloud datacenter allowed the authors to draw the following conclusions:

– a modern datacenter that implements the concept of cloud computing is a nonlinear non-stationary discrete object with a variable structure characterized by complexity, hierarchy, multidimensionality and multidimensionality;

– existing approaches and methods for managing the IT infrastructure of a cloud datacenter often do not take into account the interconnections and interactions at all levels of the hierarchy of information processes, do not adapt to the intensity of use and release of resources, but focus only on solving a specific task at a certain level (infrastructure, platform or software);

– insufficient attention is given to the development of IT infrastructure management methods and models using computational resource consumption and load forecasts for infrastructure elements, and the forecasting methods and models used are focused on a specific combination of workloads and do not adapt to current data center operating conditions;

– insufficient attention is paid to the development of integrated IT infrastructure management methods that simultaneously address virtualization aspects, software-defined IT infrastructure performance, features of computing resources, data and network storage, and the impact of hyper-converged architectures.

The findings of the analysis determine the need to develop information technology management IT infrastructure of the data center cloud services provider using the latest approaches, models and methods based on methods of integrated and hierarchical management, artificial intelligence, models and methods of forecasting load and infrastructure consumption of resources and infrastructure consumption.

II. Generalized scheme for managing the IT infrastructure of cloud datacenter

Tasks set determine the development of management information technology (MIT) cloud IT center IT infrastructure based on a hierarchical adaptive decomposition-compensation approach, which, on the one hand, allows to achieve the specified quality of cloud services in the hierarchical structure of information processes, and on the other - to adapt to changes in forecasting Consumption of IT infrastructure computing resources.

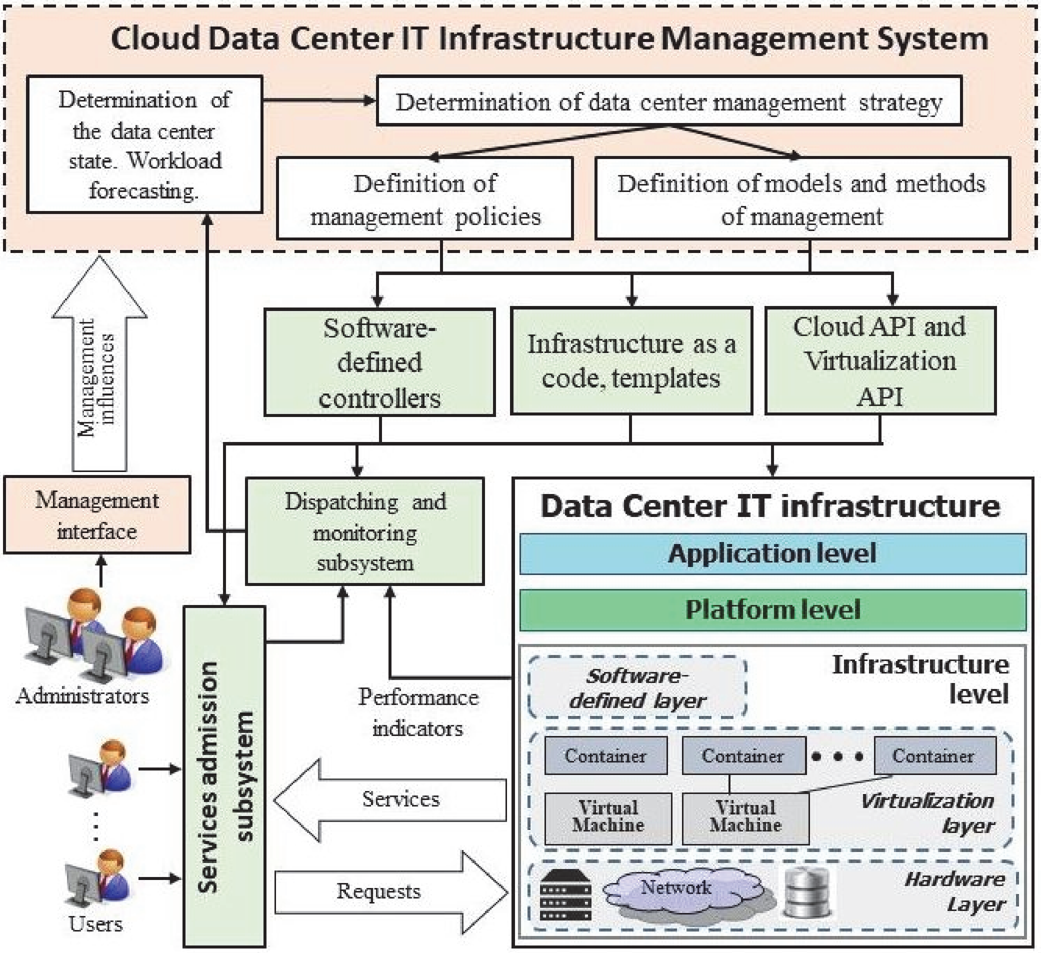

Figure 1. Generalized scheme of cloud datacenter IT infrastructure management system

A generalized scheme of cloud datacenter IT infrastructure management system is shown in Fig. 1. Cloud datacenter is a hierarchical software and hardware system consisting of three levels that correspond to the three service models of cloud services. The MIT includes: the scheduling, monitoring and forecasting subsystem; a subsystem for defining management strategies and selecting management methods; subsystem of generation of control influences affecting the software and hardware of the data center.

The ITIMS is implemented in the microservice architecture and includes: the subsystem of dispatching, monitoring and forecasting; a subsystem for defining management strategies and selecting management methods; subsystem of generation of control influences affecting the software and hardware of the data center.

Performance metrics include monitoring data for (physical and virtual) servers, storage, network devices, system and application software on the basis of which metrics are calculated; qualitative indicators of work of services; indicators of electricity consumption. The generated control effects are fed through APIs to actuators such as on/off devices, hypervisors, software-defined controllers, load balancers, fault tolerance and reliability controllers. The task of the subsystem of definition of strategies and methods of management is to develop, on the basis of the received information and values of metrics, decisions and signals of management, related to the optimization of the functional modes of the IT infrastructure objects and the allocation of IT infrastructure resources between these objects in the conditions of variables loads.

III. Formulation of the task of managing the IT infrastructure of cloud data center

The IaaS service model enables the physical server (FS) hardware to be used more efficiently by virtualizing its local resources. The physical server provides resources such as CPU time, memory, local storage, and network access. In doing so, the client is given a portion of the FS resources in the form of a virtual machine (VM) or container. To implement modern cloud services, the client deploys one or more VMs of the required configuration that is defined by the cloud service provider. Each VM hypervisor provides a share of FS resources. In terms of managing cloud data center resources, the resources of all FSs are pooled and provided to virtual machines for use. This creates a complex set of tasks related to managing the cloud datacenter. One of the options for reallocating pool resources between VMs is to implement a virtual machine consolidation process or virtual machine placement. VM consolidation is the deployment of VMs to FSs based on virtualization technologies to achieve certain performance indicators for data center resources as a whole while delivering specified productivity. In addition, management impacts are generated for virtual machines, physical servers, network devices, repositories, applications, and other objects at each level of the data center IT infrastructure.

The task of managing the load and resources of the cloud IT center IT infrastructure is characterized by the following factors: large dimension, ie the number of VMs is hundreds of thousands, the number of FSs is thousands; change in the structure and dimension of the matrix of distribution of VM on the FS with time;

 ISSN 1560-8956

dynamics of loads and inability to determine their distribution, ie the use of resources changes at every step under the influence of customer requests; nonlinear dependence of SLA quality indicators on the resources used. Thus, the task assigned belongs to the class of problems of control of nonlinear non-stationary discrete objects with variable structures. One of the main tasks in managing the resources of a cloud datacenter IT infrastructure is to deploy and relocate VMs in such a way that they deploy fewer physical servers, reduce the number of VM migrations, and provide the required quality of IT services delivery at all levels of the cloud computing service model. This VM distribution / redistribution process is a function of many factors, variables, constraints, and criteria and is a complex of real-time load management optimization tasks for which a fixed set of models and methods cannot be fixed and made.

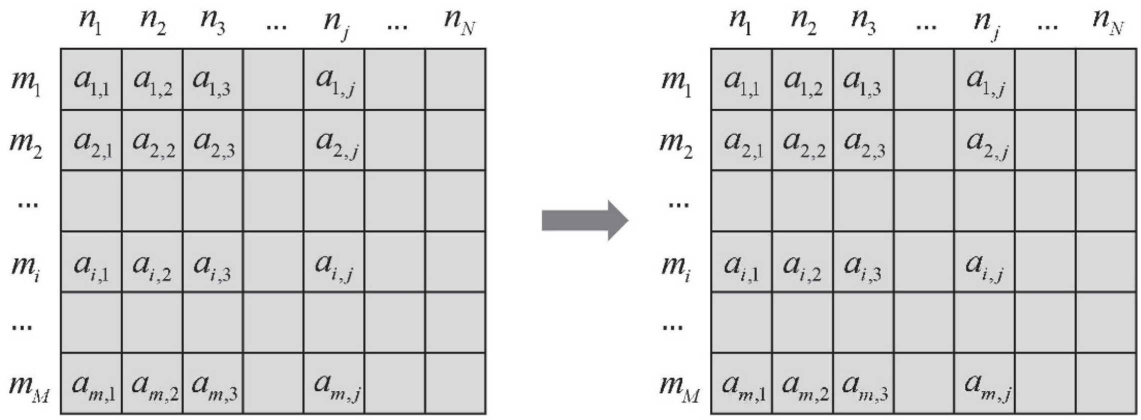

Figure 2. VM placement matrix on management step *t* (on the left) and management step *t*+1 (on the right), where *mi* – PM ID, *nj* – VM ID.

Based on the cloud data center dynamics model proposed in [6], for each control step *t* we define a matrix $\mathbf{A}: M \times N$ (Fig. 2), the elements of which are integer variables $a_{ij}^{h}(t) \in \{0,1\}$, $h = \overline{1,H}$, which indicate whether the *j*-th VM of type *h* is running on the *i*-th FS ($a_{ij}^{h}(t) = 1$), or not in use ($a_{ij}^{h}(t) = 0$). The number of changes in values in the columns of the matrix **A** from 0 to 1 is equal to the number of completed VM migrations between FSs. That is, for every *nj* the equation holds $\sum_{i=1}^{M} a_{ij} = 1$. This takes into account the types of FS and VM: $h_{ij}(t)$ – the type *j*-th VM running on the *i*-th FS; *H* – the number of VM types, $g = \overline{1,G}$ – the type of FS, *G* – the number of FS types. Using the *k*-th resource of *i*-th FS by the *j*-th VM we denote as $r_{ij}^{k}(t) \in [0,1]$ $i = \overline{1,M}$, $j = \overline{1,N}$, $k \in K$, where *K* is a set of resource types, *M* is a variable number of FSs, *N* is a variable number of VMs. VM ceases to exist if for every k $r_{ij}^{k}(t)=0$. The use of the *k*-th resource on the *i*-th FS is denoted as $R_i^k(t)\in[0,1]$, $i=\overline{1,M}$.

We define the variables that denote the input of an ITIMS as: *q*(*t*) – number of new VMs to be deployed; *r*(*t*) – the number of VMs that are stopped. We define the variables that denote the management effects on an ITIMS: *p*(*t*) – the number of VMs that migrate; *k*(*t*) – the number of on/off FSs. Let's define disturbing effects $\delta(t+1),\varepsilon(t+1)$ – the number of FS and VMs that failed, $\delta,\varepsilon\geq 0$. Then the next state of the system is determined by the equation $\mathbf{A}(t+1)=F\left[\mathbf{A}(t),\hat{N}(t+1),U(t),Z(t)\right]$, where *F*[•] is a management functional that is implemented by the chosen management strategy; $U(t)=\{k(t),p(t)\}$ – management effects; $\hat{N}(t+1)=N(t)+\hat{q}(t)-\hat{r}(t)-\varepsilon(t+1)$ – the number of VMs in the next management step, *Z*(*t*) – plan for VM migration between FSs. The amount of FS in the next step is determined by the equation $M(t+1)=M(t)+k(t)-\delta(t+1)$. Criteria defined by the chosen management strategy: minimization of the number of active FS, *M*(*t*); minimization of VM migration, *p*(*t*); minimization of the resource full load cases number; minimization of the VM deployment delay, etc. Limitations that need to be taken into account when implementing different strategies: the number of input/output migrations to the FS, the limited use of each resource within the FS, the limited speed of network interaction, the limited speed of data processing in the Storage Systems, etc.

The task of the cloud data center IT infrastructure load and resources management is characterized by the following features: *large size*, *N* is determinated by hundreds of thousands of VMs, *M* is determinated by thousands of FS; changes in the structure of the matrix **A**, while *M* and *N* change with time; *the dynamics of loads and the inability to determine their distribution*, while $r_{ij}^{k}(t)$ і $R_i^k(t)$ change at every step under the influence of customer requests; *nonlinear dependence of SLAs on the resources used*. This problem is solved in the publications of the authors [6,7,8,9].

Substantial reduction in uncertainty in the cloud datacenter IT infrastructure management provides forecasting. The rationale for predicting the load on IT infrastructure resources involves finding a compromise between the desire to release as many FSs as possible or to keep them active in order to reduce the penalties for SLA violations resulting from resource scarcity. Accurate forecasting helps reduce energy consumption and penalties by making good management decisions about deploying new VMs and migrating existing ones.

Since it is not possible to select the parameters and coefficients of the resource consumption forecasting model under any load combination, it remains at each management step to create a new model using the new data obtained from the monitoring subsystem. The amount of monitoring data required to obtain the forecast model is determined by the selection of the most appropriate forecasting method based on recent data [9].

The absence of a strict criterion for the application of one or another method of load forecasting causes the use of the combined method with adaptation of model parameters. The combined method consists in the application of several (alternative) forecasting methods with adaptation of the training sample size and subsequent selection of the forecast based on a defined criterion. A set of alternative forecasting methods is determined by previous studies of their application to test samples from real traces, for example from [10, 11]. To determine the control at the current step, the forecast obtained by the method and size of the training sample, which gave the minimum average absolute error of the forecast in percentages in the previous step, is selected. For a variant of the combining method, not one pair of "method-window" is taken into account, but several pairs from the previous one, as well as from several previous control steps.

IV Operator form to formulate, analyze and solve the cloud data center IT infrastructure management tasks

So, the state of a cloud datacenter IT infrastructure as nonlinear non-stationary discrete object with variable structure is proposed to define in three dimensions: the dimension of resources (surplus or lack); the dimension of loads (variable or constant); the dimension of loads dynamics (with a trend of increasing or decreasing). The belonging the current state to specific dimension values is determined by a set of metrics based on monitoring data. It influences the definition of the IT infrastructure resource management strategy.

Let's look at the fundamental provisions of the mentioned operator form adopted for the development of MIT of cloud data center IT infrastructure. In this case, depending on the current status of the cloud datacenter, defined management strategy and load forecasting results the set of criteria and constraints change under the influence of technological features of the cloud data center and the need to provide services with the specified quality indicators.

Thus, it is necessary each time to build new models of cloud data center and to apply different managemen algorithms and methods in the form of implementation schemes in order to adapt to the current operating conditions.

The need to develop a operator form is influenced by a number of factors that address the challenges of managing IT infrastructure. It should be noted that this idea came from the experience of effectively applying the powerful concepts of object-oriented programming and theory of automatic control. Indeed, the concepts of object-oriented programming such as class and object, determining the behavior of objects by class belonging (functions or methods), class hierarchy, polymorphism and overloading of methods, inheritance of characteristics and behavior lead to program structuring, building a system of components that should become building blocks at every level of IT infrastructure management.

We introduce the factors that influence the operator form. First, it is the presence of several levels of IT infrastructure management, due to its complexity and multifunctionality. It seems appropriate to choose management decisions by pre-selecting a management strategy. And since there are several strategies and the results can only be evaluated after planning, the operator form is very convenient. Second, it is multicriteria when the criterion itself is the subject of choice.

Given the above, the objective function of the data center IT infrastructure management as a nonlinear non-stationary discrete object with variable structure in the operator form can be presented as follows:

$$\min\left[\mathbf{C}_1\mathbf{F}_i\left[P,V\right]+\mathbf{C}_2\mathbf{M}_i\left[P,V\right]+\mathbf{C}_3\mathbf{L}_i\left[P,V\right]\right] \tag{1}$$

where $\mathbf{F}_i\left[P,V\right]=(b_1^F,b_2^F,...,b_m^F)^T$ – operator *i*-th strategy, which determines the FS of the set *P* with coherent VMs of the set *V*, $b^F\in\{0,1\}$;

$\mathbf{M}_i\left[P,V\right]=(b_1^M,b_2^M,...,b_m^M)^T$ – operator *i*-th strategy, which determines distribution noncoherent VMs of the set *V* between FS of the set *P*, $b^M\in\{0,1\}$;

$\mathbf{L}_i\left[P,V\right]=(b_1^L,b_2^L,...,b_m^L)^T$ – operator *i*-th strategy, which determines the FS of the set *P* that have been put to sleep, $b^C\in\{0,1\}$;

$\mathbf{C}_1=(c_1^1,c_2^1,...,c_m^1)$ – the vector of the coefficients of costs to support the work of the FS;

$\mathbf{C}_2=(c_1^2,c_2^2,...,c_m^2)$ – the vector of the coefficients of costs to support the work of the FS, allocated for servicing noncoherent VMs;

$\mathbf{C}_3 = (c_1^3, c_2^3, ..., c_m^3)$ – the vector of the coefficients of costs to support the work of the FS in the off state.

The result of the action of the operators **F** and **M**, which receive at the input vectors of characteristics of FS and BM, is the matrix of distribution of VM between FS **A**(t+1). The result of the action of the operator **L**, which receives at the input the metrics of the cloud data center state, is the migration plan VM $Z(t)$ and the management effects $U(t)$. The result of the objective management function calculation is the value of operating costs for maintaining the cloud datacenter.

Let's introduce the concept of an operator's implementation scheme, which depends on the historical data collected by the systems in the management process, on the values of the metrics and on the results of forecasting. In this case, the operator scheme provides mechanisms that are then implemented according to strategies. Management models, set of constraints and criteria are determined in the management process, depending on historical data and forecasting results.

The operational approach to the cloud datacenter IT infrastructure management is generally implemented consistently at three levels: strategic level, planning levels and management levels. The level of strategies defines the following management strategies: lack of resources at constant load; excess resources at constant load; lack of resources and trend to reduce load; excess resources and trend to reduce load; lack of resources and trend to increase load; excess resources and a trend to increase load. The management strategy is defined through three stages: assessing the state of the IT infrastructure using metrics, receiving tasks and requests from the user, and forecasting workload.

Each management strategy is implemented by solving the following planning tasks: calculation of metrics that reflect the trends of load and composition of IT services provided by the provider at the current time; load forecasting on three levels of cloud computing service model; determination of the optimization methods to be applied at the current time, using the operator approach taking into account the specifics of the data center provider, the composition of the optimization methods and the current conditions of the data center.

The result of the work of the planning level is a subset of management tasks, the implementation of which leads to the achievement of the goal of a specific management strategy. The identified tasks are implemented using software-defined techniques and technologies at three levels of the cloud computing service model.

A subset of management tasks for a particular strategy is implemented by the implementation schemes and basic methods developed in [6, 7, 8, 9]. The application of a particular method of resource management or combinations thereof, depending on the state of resources and load of data center, as well as the defined strategy is shown in table 1.

Table 1. **Choice of strategy and scheme of implementation of resource management depending on the state of data center**

Preconditions	Strategies	Schemes of strategy implementation	Basic methods
The average viability of the virtual machine; physical server imbalance indicator; ratio of required resources to the average amount of available resources; the threshold of available resources; data center capacity metric.	Management with a lack of resources at constant load, ***S*1**	Uniform consolidation of VM	Modified annealing method.
	Management with excess resources at constant load, ***S*2**	Management of migration and deployment of VMs in a stable mode. A two-step method of resources management.	A modified ray search method.
	Management with a lack of resources and a trend to reduce load, ***S*3**	Integrated Resource Management (IRM).	Integrated resource management method.
	Management with excess resources and a trend to reduce load, ***S*4**	Integrated Resource Management	Integrated resource management method. Data center performance management method.
	Management with a lack of resources and a trend to increase load, ***S*5**	Dynamic consolidation and deployment of VMs.	Modified method of learning with reinforcement.
	Management with excess resources and a trend to increase load, ***S*6**	Dynamic consolidation and deployment of VMs with performance management.	Modified method of learning with reinforcement. Data center performance management method.

Operator implementation schemes depend on historical data, metric values, and forecasting results and and embody a specific management strategy. The parameters of the management model, the set of the constraints and the criteria are determined in the management process. Management strategies S are determined by the set of models $\Omega(D,O,G)$ and management methods with the set of variables D, the set of constraints O, and the set of criteria G, which are determined in the course of system operation.

 ISSN 1560-8956

Each management strategy $S_i = \Omega(D_i, O_i, G_i), D_i \in D, O_i \in O, G_i \in G$ implements a specific operator of the objective function (1). Each subset consists of an arbitrary number of elements, $D_i = \{d_{0,i}, d_{1,i}, \ldots\}$, $O_i = \{o_{0,i}, o_{1,i}, \ldots\}$, $G_i = \{g_{0,i}, g_{1,i}, \ldots\}$, $D_i \neq \varnothing, O_i \neq \varnothing, G_i \neq \varnothing \;\; \forall i$.

Thus, the implementation scheme of the operators of the objective function (1) is determined by the system

$$\begin{cases} \mathbf{F}_i\left[P,V\right]:\left\langle \sum_{i=1}^{N(t)} res_i^{VM}, \sum_{j=1}^{M(t)} Res_j^{PM}, \Omega(D_i,O_i,G_i) \right\rangle \rightarrow \mathbf{A}(t+1) \\ \mathbf{M}_i\left[P,V\right]:\left\langle K, \Omega(D_i,O_i,G_i) \right\rangle \rightarrow M(t+1) \\ \mathbf{L}_i\left[P,V\right]:\left\langle K, \Omega(D_i,O_i,G_i) \right\rangle \rightarrow Z(t) \end{cases}, \quad (2)$$

where **F***i*, **M***i*, **L***i* – operators of the objective function (1), res_i^{VM} – the resources that the *i*-th VM consumes, Res_i^{PM} – resources consumed by the j-th FS, *K* – a set of metrics to evaluate the state of a cloud data center, $\Omega(D_i, O_i, G_i), D_i \in D, O_i \in O, G_i \in G$ – selected model in the process of cloud data center resource management. The specific values are selected according to the configuration of the cloud data center, the values of the table 1 and the requirements for implementing each control method.

V Models for determining management strategies

In order to automatically determine the management strategy, it is necessary to develop a model and a method of fuzzy inference, as well as a "game with nature" model, which will allow to determine the management strategy depending on the metrics of the cloud data center state assessment.

Let's consider a model of fuzzy inference in which the control strategy is represented by the output variable *X*, and the metrics of the cloud data center state are represented by *n* input variables of the set of metrics *K*. In this case, the input variables are: *K*1 is the average VM viability factor value, *K*2 is the FS imbalance indicator, *K*3 is the ratio of required resources to the average amount of available resources, *K*4 is the free resource threshold and *K*5 is the metric of the data center capacity. It is worth noting that the metrics *K*1, *K*2 *K*3 are proposed for the first time. The output variable takes values corresponding to the names of the strategies: *S*1 - with a lack of resources at constant load, *S*2 - with excess resources at a constant load, *S*3 - with a lack of resources and a trend to reduce load, *S*4 - with an excess of resources and a trend to reduce the load, *S*5 - with a lack of resources and a trend to increase the load, *S*6 - with an excess of resources and a trend to increase the load. Using the rules of the form "If $K1 \in A1 \wedge K2 \in A2 \wedge \ldots \wedge Kn \in An$, then $X \in Si$", where *An* is the set of values of the input variable *Kn*, a fuzzy Mamdani output is implemented to determine the data center management strategy for the current management step. The approach presented in [12] is used to construct membership functions. This approach has been modified by developing a new technique for selecting peaks, which is to combine several adjacent intervals to determine the peak. To do this, we use historical data of system operation.

Let,s consider the "game with nature" model, where the first player is the cloud data center management strategy module (MSM player) and the second player is "nature" in the form of a load on cloud data center resources (the data center player). However, "nature" is not an ally or adversary of the first player. The first step in developing this model is to build a payoff matrix.

The MSM player has six strategies *Si*, *i*=1,…,6 (table 1), the data center player can be in one of n states *Aj*, *j*=1,...,n. The states of a data center player are determined by combinations of the values of the following metrics: *K*1 is the average value of the VM viability factor that characterizes the load level; *K*2 is the straight line of load trend coefficient calculated using historical data of the instantaneous viability factor of VM; *K*3 is a data center capacity metric that characterizes excess or lack of resources.

The values of the state variables can be as follows: if $0.95 \le K_1 \le 1.05$ the load is constant, otherwise variable; if $K_2 \le -0.1$, then the load decreases, if $K_2 \ge 0.1$, then the load increases, otherwise the load is considered constant; $-0.1 < K_2 < 0.1$; if $K_3 \le \delta$, where δ is an expertly determined resource utilization threshold, then there is an excess of resources, otherwise there is a lack of resources in the data center. The gain of applying the strategy *Si* in the state *Aj* is denoted by *qij*. The coding of strategies depending on the values of the metrics *K* is shown in table 2.

Taking into consideration the above, we have to agree that the payoff matrix Q takes the following form (table 3). The initial values of the gains *qij* are selected by using the Delphi method based on expert judgment.

ISSN 1560-8956

Table 2. **The coding of the data center player states in the "game with nature"**

Availability of resources	**Load**	**Trend of load change**	**Data center player states**
$K_3 > \delta$	$0.95 \le K_1 \le 1.05$	$-0.1 < K_2 < 0.1$	*A*1
$K_3 \le \delta$	$0.95 \le K_1 \le 1.05$	$-0.1 < K_2 < 0.1$	*A*2
$K_3 > \delta$	$0 < K_1 < 0.95, K_1 > 1.05$	$K_2 \le -0.1$	*A*3
$K_3 \le \delta$	$0 < K_1 < 0.95, K_1 > 1.05$	$K_2 \le -0.1$	*A*4
$K_3 > \delta$	$0 < K_1 < 0.95, K_1 > 1.05$	$K_2 \ge 0.1$	*A*5
$K_3 \le \delta$	$0 < K_1 < 0.95, K_1 > 1.05$	$K_2 \ge 0.1$	*A*6

Table 3. **The payoff matrix of the "game with nature"**

The data center resource management strategy implementation scheme	**The MSM player strategy**	**The data center player state**					
		***A*1**	***A*2**	***A*3**	***A*4**	***A*5**	***A*6**
Uniform consolidation of VM method	*S*1	*q*11	*q*12	*q*13	*q*14	*q*15	*q*16
Management of VMs migration and deployment in a stable mode method. A two-step method of resources management.	*S*2	*q*21	*q*22	*q*23	*q*24	*q*25	*q*26
Integrated Resource Management	*S*3	*q*31	*q*32	*q*33	*q*34	*q*35	*q*36
Integrated Resource Management	*S*4	*q*41	*q*42	*q*43	*q*44	*q*45	*q*46
Dynamic consolidation and deployment of VMs method. Method of performance management.	*S*5	*q*51	*q*52	*q*53	*q*54	*q*55	*q*56
Dynamic consolidation and deployment of VMs method. Method of performance management.	*S*6	*q*61	*q*62	*q*63	*q*64	*q*65	*q*66

The data center player's gains *qij* in the management process are determined based on historical data by the calculation of the objective function (1). As the criterion for the choice of optimal solutions, Wald's criterion is used. When applying this criterion, a strategy is chosen, which guarantees gain no less, than the lower price of the game. That is, the best result (the position of extreme pessimism) is chosen from all the bad results.

VI. Conclusions

The modern cloud datacenter is a complex management object characterized by hierarchy, multidimensionality and variability. However, insufficient attention is paid to the development of methods and models for integrated management of IT infrastructure using load forecasts. The information technology for managing the cloud datacenter IT infrastructure is proposed based on the operator approach, which is implemented consistently at three levels - strategies, planning and management. The peculiarities of the cloud data center IT infrastructure load and resources management tasks are defined: large size of the optimization problem; changing the structure of the system; dynamics of loads and inability to determine their distribution; nonlinear dependence of SLA quality indicators on the resources used. The tasks and stages of the cloud data center IT infrastructure management have been developed taking into account the hierarchy of subsystems and management strategies. Management strategies and strategies selection model for managing cloud data center load and resources using the "game with nature" model have been developed.

REFERENCES

1. P. Mell and T. Grance, “The NIST definition of cloud computing,” Gaithersburg, USA, Special Publication 800-145, vol. 800, no. 145, 2011, p. 7.

2. A. Kupin, I. Muzyka and R. Ivchenko, "Information Technologies of Processing Big Industrial Data and Decision-Making Methods," 2018 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), Kharkiv, Ukraine, 2018, pp. 303-307.

3. S. L. Kryvyi, Y. V. Boyko, S. D. Pogorilyy, O. F. Boretskyi and M. M. Glybovets, “Design of Grid Structures on the Basis of Transition Systems with the Substantiation of the Correctness of Their Operation,” Cybernetics and Systems Analysis, vol. 53, no. 1, pp. 105–114, January 2017.

4. R. da Rosa Righi, M. Lehmann, M. M. Gomes, J. C. Nobre, C. A. da Costa, S. J. Rigo, M. Lena, R. F. Mohr and L. R. B. de Oliveira, “A Survey on Global Management View: Toward Combining System Monitoring, Resource Management, and Load Prediction,” Journal of Grid Computing, pp. 1–30, 2019.

5. M. A. Khan, A. Paplinski, A. M. Khan, M. Murshed and R. Buyya, “Dynamic virtual machine consolidation algorithms for energy-efficient cloud resource management: a review,” Sustainable Cloud and Energy Services, Springer, pp. 135–165, 2018.

 ISSN 1560-8956

6. Zharikov E. Cloud datacenter integrated resource management based on virtual machines // Mathematical Machines and Systems, 2018. – № 2. – P. 21-32.

7. Telenyk S., Zharikov E. and Rolik O., "Modeling of the Data Center Resource Management Using Reinforcement Learning," 2018 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), Kharkiv, Ukraine, 2018, pp. 289-296.

8. Zharikov E., Rolik O., Telenyk S. A Decomposition Approach to Hierarchical Management of Cloud Data Center Services //2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT). – IEEE, 2018. – Т. 1. – С. 42-47.

9. Zharikov, E., Telenyk, S. & Bidyuk, P. Adaptive Workload Forecasting in Cloud Data Centers. J Grid Computing, 2019. doi:10.1007/s10723-019-09501-2

10. GWA-T-12 Bitbrains [Online] Available from: http://gwa.ewi.tudelft.nl/datasets/gwa-t-12-bitbrains [Accessed June 01, 2019].

11. C. Reiss, J. Wilkes, and J. L. Hellerstein, “Google cluster-usage traces: format+schema,” Google Inc., Mountain View, CA, USA, Technical Report, 2011.

12. Telenyk S., Bidyuk P., Zharikov E. and Yasochka M., "Assessment of cloud service provider quality metrics," 2017 International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo), Odessa, 2017, pp. 1-5. doi: 10.1109/UkrMiCo.2017.8095422"