

Infrastructure Capacity Planning in Digitalization of Educational Services

<https://doi.org/10.3991/ijet.v17i03.27811>

Pavel Petrov¹(✉), Mihail Radev¹, Georgi Dimitrov², Dimitrios Simeonidis¹

¹ University of Economics – Varna, Varna, Bulgaria

² University of Library Studies and Information Technologies, Sofia, Bulgaria

petrov@ue-varna.bg

Abstract—Capacity planning of the infrastructure is a difficult task performed during the virtualization of the IT infrastructure. It is the planning of the necessary hardware to which the existing systems will be transferred, and later the future systems that will work in a virtualized environment. Infrastructure capacity planning in digitalization of educational services is a process in which the IT infrastructure is needed to meet not only the current workload but also the future workload and to follow further requirements. In case the process of planning is well performed, it could more fully support the development of emerging technologies in learning. The traditional IT infrastructure, bound by the physical limitations of the devices, does not meet the requirements for flexible changes, for quick recovery after a problem. This leads to difficult management, to weaknesses in the implementation of educational processes. Virtualization could solve most of these problems for modern IT infrastructures. It provides highly efficient and reliable operation of operating systems and applications that do not depend on the computer system.

Keywords—IT infrastructure, capacity planning, virtualization, digitalization, educational services

1 Introduction

Capacity planning in the digitalization of educational services is a process in which the IT infrastructure is needed to meet not only the current workload but also the future workload and compliance with future requirements. In case the process of planning is well performed, it could more fully support the development of emerging technologies in learning.

It is well known that within one school year the IT infrastructure is loaded unevenly. The variation in the workload is related to changes in the curricula [7], expansion or contraction of the curricula, introduction of new disciplines, pandemic situations [1], etc. Due to the complexity of the IT components, the load to the whole infrastructure is difficult to predict. Therefore, capacity planning for the digitalization of educational services in universities [14] [15] should be done in terms of efficient use of resources to provide the IT services needed by the university users.

Capacity planning of the new infrastructure is among the most difficult and uncertain tasks performed during the virtualization of the IT infrastructure. It is the planning of the necessary hardware to which the existing systems will be transferred, and later the future systems that will work in a virtualized environment. It means both planning and providing resources for the current needs of the university and providing a reserve for future growth of needs [9]. The stage is extremely important for each of the types of virtualization technologies. This stage also includes the selection of virtualization software between different options, selection of a storage system, design of the network topology. It is based on the detailed inventory consideration for the applications used and the services offered by the servers in the current hardware. The process is difficult because of the future uncertainty and because of the easy movement of objects in the virtualized IT infrastructure. Hence, it is difficult to determine the loads in the virtual environment. At the same time, the future educational activity of the university, the educational situation, and future requirements must be forecasted [2] [3], provided that the future system will also be constantly changing.

In addition to the capacity of the infrastructure components, the capacity of the IT team should also be planned at this stage - on the one hand servers are consolidated and the people involved in their support will decrease, on the other hand more skilled administrators are needed to administer the additional layer in IT the infrastructure - the virtualization.

2 Stress condition in capacity planning process

To correctly determine the capacity, it is necessary to assess the potential load on the resources of the servers, the storage systems, and the network bandwidth. The assessment can be done by means of automated tests that allow to simulate the stress load of the individual components of the systems. The purpose of these tests is to report the worst-case scenario (for example, all users connect to the server at the same time to authenticate and then start the most resource-intensive application [18] at the same time) and to use it to plan digital capacity of educational services in universities [20].

The procedure we propose to follow when planning the capacity for digitalization of educational services in universities is as follows: first to monitor the metrics for productivity and load of the current IT infrastructure, then to design the new infrastructure based on good practices, manuals from the manufacturers of the virtualization solution and in addition independent from the manufacturer own calculations. The last could be realized in a test version based on leased servers or those of the company-implementer. The last step is to apply simulation stress tests to each resource and to each component of the new infrastructure and to monitor how it responds and what maximum loads it can withstand. If the capacity of a component is not planned properly, stress tests are adjusted and applied again. If a completely new IT infrastructure is built, i.e., without a previous infrastructure and its previous condition, then at the last stage of the proposed procedure historical monitoring cannot be carried out. The main groups of users will have to be determined on an expert basis, with which applications they will work,

to be able to set the parameters for computing and storage resources, as well as for the necessary capacity.

We propose to divide the stress load of the resources into the following stages: first stage – based on the analysis of the data from the monitoring of the work of the users and/or the metrics of the main elements of the IT infrastructure, so a scenario of the work of the users should be defined. The users should be divided [4] [6] into typical groups and test scenarios are developed. The next stage is the development of scripts for loading the infrastructure – to simultaneously simulate the implementation of many typical tasks and to load the infrastructure, the next two stages are related to the implementation of tests and analysis of results. During the implementation of the stress tests, different levels of load are monitored - while it is constantly increasing the indicators of system performance are observed. When results are obtained that do not meet the target performance of the IT infrastructure, it is necessary to make additional adjustments to the virtualized infrastructure and perform the tests again. The last two stages are iterative and are performed until the maximum possible results are obtained from the virtualized IT infrastructure. An option for the implementation of stress tests is to use a software tool such as Login Virtual Session Indexer (Logon VSI).

The results of these resource-intensive tests will serve as an assessment of the performance and viability of the IT infrastructure and will be the basis for planning its capacity to digitize educational services in universities. The measured efficiency indicators allow to determine the allowable load of the infrastructure and to identify the weak points - the specific resources, the lack of which leads to a decrease in productivity. The tests can be applied iteratively - after setting up and changing the configurations of the elements of the IT infrastructure. If the required resources are not properly assessed, there is a risk of failure [17] or increased losses from unplanned service outages in each virtualization project.

These are the steps for planning the digitalization of educational services in universities and determining the capacity of the elements of IT infrastructures. Capacity planning software calculators cannot give results like tests, they can be an additional tool. But the combination of resource stress tests, know-how and good practices applied by designers and implementers would lead to precise planning of the virtualized infrastructure for the expected workloads. When planning the virtualized IT infrastructure, we exclude the possibilities of hyperthreading technology due to the ambiguous opinions of the manufacturers of virtualization solutions - for example: as many processor resources (such as cache) are shared between logical processors.

In server virtualization, there is need to determine the required server hardware resource that will be able to provide services to clients. And in all cases, the provision of the necessary hardware must be done at the lowest cost, without resizing and without over insuring.

In desktop virtualization, the goal at this stage is to determine the maximum number of virtual desktops that can be supported during peak infrastructure use. Again, as with server virtualization, the requirements for processor, memory, storage data resources, and network capacity must be defined. At the server level, it is determined how many virtual desktops can be hosted on it, so that users of these desktops can perform their routine activities without adversely affecting the work of end users. Capacity planning

tests can be performed with specialized software tools, as well as with tools built into the used operating systems.

In both virtualization options, due to the ease of creating new virtualization machines and the ease of moving virtual machines and applications between hosts, a shortage of resources may occur in the future. Newly created or relocated virtual machines will change the infrastructure and the original planning for a certain number of virtual machines and the resource load caused by these machines will no longer be valid. Security problems should also be considered [11] [13].

Another problem related to working in a virtual desktop IT infrastructure is that the peak loads in the operation of the infrastructure cannot be predicted, because they depend on the loads that individual users apply, and which occur at different, undefined moments. When planning the virtualization capacity of desktops, the planning of the virtualization capacity of the storage systems is also done.

3 Resource virtualization approach and return on investment

When planning the capacity of resources for virtualization of data storage systems, the input-output operations, the capacity for storage of the images of the virtual machines and the capacity for data storage are determined [5]. Again, simulation tests and different load scenarios combined with following good practices are a solution to the problem of capacity determination in this type of virtualization technology as well. System performance indicators are obtained from monitoring software (e.g., Zabbix, Microsoft System Center Operations Manager), as well as from specialized performance monitoring software tools such as SolarWinds Storage Resource Monitor.

The design creates the architecture of the new infrastructure, following the success criteria of the first stage, the requirements for fault tolerance, the scalability of the solution and the required percentage of service availability.

Before the commence of the tests and the monitoring of the performance indicators, key performance indicators (KPIs) could be created, depending on the importance of the different metrics for the university that implements the virtualization solution [16]. The KPIs have various names in the specialized IT literature, for example - evaluation indicators, target indicators, etc. We believe that for the purposes of virtualization, performance indicators are the most accurate concept.

The possibilities at this stage are related to the three models for providing IT services to users – through its own infrastructure (on premise) [8] [19], leased (cloud) infrastructure [10] or a hybrid option [12]. Between the three models, we propose that the choice be made based on the following criteria in Table 1.

The choice depends on planning the capacity of the equipment needed to digitize educational services in universities - this is the process of calculating the total workload and calculating what hardware will need to be provided to provide the necessary platform for the virtualized IT infrastructure. Newly purchased/rented hardware must also be able to meet the peak resource needs of users in the company.

Table 1. Criteria for selecting an IT infrastructure model

Selection criteria	Own IT infrastructure (on premise)	Leased (cloud) infrastructure	Hybrid infrastructure
Security	1	3	2
Legal requirements	1	3	2
Uneven load	3	1	2
Control over critical educational processes	1	3	2
Running costs instead of capital costs	3	1	2
Limited IT team, expertise, equipment and security budget and IT resources in general	3	1	2

In order to decide on the choice of a specific virtualization solution, many factors must be analyzed. The main one in most projects is the assessment of the total cost of ownership (TCO) and return on investment of several possible virtualization solutions. The financial evaluation of the project includes an analysis of the total acquisition cost. The calculation requires the determination of all costs and revenues - purchase and installation of new servers, licenses for the new virtualization software, energy savings, cooling and physical space.

TCO in the IT field presents the total costs of a university for the acquisition and maintenance of a system. In the context of the current study, these are the total costs of building and maintaining a virtualized IT infrastructure. The purpose of this indicator is to include "hidden" costs, not just those for purchasing servers and software licenses. Thus, the calculation will include the cost of acquiring the infrastructure and they will include costs of design, purchase of servers and other hardware, training costs - for users and IT staff, for installation and configuration of software. The second group of costs are those for infrastructure maintenance and include its management, backup and recovery, hardware and software maintenance, customer support, electricity consumption, cooling costs, space occupied by equipment, time lost by users in case of system failure.

We will divide the components of the TCO into four groups - hardware costs, software costs, labor costs and infrastructure costs. We will also calculate the costs for one job.

Hardware costs include the cost of purchasing computers, servers, UPSs, switches, routers. The hardware update period is different for each university and therefore the TCO will be calculated for a period set by the specific university. The renewal period is also different for the individual hardware components, which must also be considered in the calculations.

Software costs include the cost of purchasing operating system licenses and purchasing virtualization software, considering the number of CPU cores, as in most cases the software is licensed for individual cores. This component of TCO also includes the purchase of client access licenses - for each client's right to access the server operating system.

Labor costs are the costs of paying all people involved in the process of design, installation, configuration, training of users and administrators in the IT department,

maintenance and administration, management and monitoring of the implemented virtualization solution.

Infrastructure costs include the costs of power supply and cooling of the equipment. They include the respective costs for the implemented version of the virtualization solution, calculated according to the working hours of the university. In this group we will include other costs, such as the cost of providing the necessary space for a server room with hardware conditions, costs from non-working users due to system failure.

The results for TCO will be used in the analysis and selection of a specific solution for virtualization of the IT infrastructure in the next stage of the methodology. We consider the approach to choosing a virtualization solution based only on the criterion of the lowest TCO indicator to be wrong. Thus, many other criteria are not considered, which, in combination and with different weight, will give a more accurate criterion picture for finding the most effective and efficient solution.

4 Conclusion

The choice of the specific solution from the available alternatives is made based on criteria, requirements and internal standards determined by the university. The criteria have different weights. There are more than two options for each choice of alternatives (in the case of an IT infrastructure virtualization project, it is possible to make a comparison with the traditional infrastructure option). Or we have a multi-criteria problem that requires a decision as a process of identifying and choosing between alternatives based on the predefined preferences.

The traditional IT infrastructure, bound by the physical limitations of the devices, does not meet the requirements for flexible changes, for quick recovery after a problem. This leads to difficult management, to weaknesses in the implementation of educational processes.

Virtualization is a technology that could solve most of these problems for modern IT infrastructures. It provides highly efficient and reliable operation of operating systems and applications that do not depend on the computer system. Virtualization is also a concept that allows IT resources to be allocated dynamically according to specific user needs. It covers the entire IT infrastructure - servers, applications, desktop machines, networks, storage systems.

The virtualized IT infrastructure model is changing the way universities work and traditional IT departments have the potential to become in-house service providers. The results of the implementation of virtualization solutions will be not only technical but also organizational - increased level of service to educational needs, lower costs for hardware and administrative staff, reduced energy consumption and cooling costs and as a general result - more flexible, efficient, and productive educational organizations.

5 Acknowledgment

This research is supported by NPI-45/2020 from University of Economics - Varna Science Fund.

6 References

- [1] Alasmari, T. (2021). Learning in the COVID-19 Era: Higher Education Students and Faculty's Experience with Emergency Distance Education. *International Journal of Emerging Technologies in Learning (iJET)*, 16(9), pp.40-62. <https://doi.org/10.3991/ijet.v16i09.20711>
- [2] Baglaeva, E. M., Sergeev, A. P., Shichkin, A. V., & Buevich, A. G. (2021). The extraction of the training subset for the spatial distribution modelling of the heavy metals in topsoil. *Catena*, 207, 105699. <https://doi.org/10.1016/j.catena.2021.105699>
- [3] Balabanova, I. S., Kostadinova, S. S., Markova, V. I., Sadinov, S. M., & Georgiev, G. I. (2021). Statistical techniques to determine of optimal and acceptable noise levels. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1032, No. 1, p. 012006). IOP Publishing. <https://doi.org/10.1088/1757-899X/1032/1/012006>
- [4] Dimitrov G., Bychkov O., et al. (2020). Creation of Biometric System of Identification by Facial Image. 3rd International Colloquium on Intelligent Grid Metrology (SMAGRIMET), Croatia: IEEE, pp.29-34. <https://doi.org/10.23919/SMAGRIMET48809.2020.9263995>
- [5] Kuyumdzhiyev, I. (2019). Comparing Backup and Restore Efficiency in MySQL, MS SQL Server and MongoDB. In 19 International Multidisciplinary Scientific Geoconference SGEM 2019, 19(2.1), pp.167-174. <https://doi.org/10.5593/sgem2019/2.1/S07.022>
- [6] Malkawi, R., Saifan, A. A., et al. (2020) Data Mining Tools Evaluation Based on their Quality Attributes. *International Journal of Advanced Science and Technology*, 29(3). pp.13867-13890.
- [7] Marinova, R. & Momcheva, G. (2019). Survey of Information Technology Undergraduate Degree Programs in Canada, In 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), pp.1-4. <https://doi.org/10.1109/CCECE.2019.8861715>
- [8] Omirzak, I., Razumova, Y., & Nikishina, S. (2021). New Generation Mobile Networks and Their Application in Electronic Learning. *International Journal of Interactive Mobile Technologies (IJIM)*, 15(2), pp.157-171. <https://doi.org/10.3991/ijim.v15i02.18323>
- [9] Pashev, G. (2021). An Adaptive E-learning System for Teaching Mobile Applications. *International Journal of Computer Science and Mobile Computing*, 10(9), pp.80-87. <https://doi.org/10.47760/ijcsmc.2021.v10i09.008>
- [10] Patel, E., & Kushwaha, D. S. (2019). Analysis of Workloads for Cloud Infrastructure Capacity Planning. In *Data and Communication Networks*. Springer, Singapore, pp.29-42.
- [11] Petrov P., Dimitrov G., & Ivanov S. (2018). A Comparative Study on Web Security Technologies Used in Irish and Finnish Banks. 18 International Multidisciplinary Scientific Geoconference SGEM 2018: Conference Proceedings. 18(2.1), pp.3-10. <https://doi.org/10.5593/sgem2018/2.1/S07.001>
- [12] Petrov P., Ivanov S., et al. (2021) Projects Management in Technology Start-ups for Mobile Software Development. *International Journal of Interactive Mobile Technologies (IJIM)*, 15(7), pp.194-201. <https://doi.org/10.3991/ijim.v15i07.19291>
- [13] Petrov, P., Dimitrov, P., et al. (2020). Using the Universal Two Factor Authentication Method in Web Applications by Software Emulated Device. *International Multidisciplinary Scientific GeoConference: SGEM*, 20(2.1), pp.403-410. <https://doi.org/10.5593/sgem2020/2.1/s07.052>
- [14] Petrova, S., Sergeev, A., et al. (2019). Online Public Access Catalogs in Bulgarian University Libraries an Empirical Study of Seven-Year Evolution. In 12th Annual International Conference of Education, Research and Innovation (ICERI), pp.230-236. <https://doi.org/10.21125/iceri.2019.0091>

- [15] Petrova, S., Stefanov, S., et al. (2019). Information systems used in Bulgarian university libraries as online public access catalogs. In International Multidisciplinary Scientific Geo-Conference Surveying Geology and Mining Ecology Management (SGEM), 19(2.1), pp.353-360. <https://doi.org/10.5593/sgem2019/2.1/S07.046>
- [16] Radev, M. (2017). Using the TOPSIS Method to Evaluate Projects for Virtualization. *Izvestia Journal of the Union of Scientists - Varna. Economic Sciences Series*, 6(2), pp.234-241.
- [17] Stoev, S. (2017). Integration of Risk Management Processes into the Business of IT Companies. *Izvestia Journal of the Union of Scientists - Varna. Economic Sciences Series*, 6(2), pp.225-233.
- [18] Stoev, S. (2019) Using of additional packages of components for accelerated application development. *Izvestia Journal of the Union of Scientists - Varna, Union of Scientists - Varna*, 8(2). pp.171-179. <https://doi.org/10.36997/IJUSV-ESS/2019.8.2.171>
- [19] Vasilev, J., & Stoyanova, M. (2019). Information Sharing with Upstream Partners of Supply Chains. In 19 International Multidisciplinary Scientific Geoconference SGEM 2019, 19(2.1), pp.329-336. <https://doi.org/10.5593/sgem2019/2.1/S07.043>
- [20] Wu, W., & Plakhtii, A. (2021). E-Learning Based on Cloud Computing. *International Journal of Emerging Technologies in Learning (iJET)*, 16(10), pp.4-17. <https://doi.org/10.3991/ijet.v16i10.18579>

7 Authors

Dr. Pavel Petrov is working as Associate Professor in the Department of Informatics and Computer Science at the University of Economics - Varna, Varna, Bulgaria. His research interests include distributed web systems, big data and cloud computing.

Dr. Mihail Radev is a Bulgarian computer scientist and educator. He is working in the Department of Informatics and Computer Science at the University of Economics - Varna, Varna, Bulgaria. He works also as a head of the local Cisco Academy (Email: radev@ue-varna.bg).

Dr. Georgi Dimitrov is working as Professor in the University of Library Studies and Information Technologies, Sofia, Bulgaria. He is Vice Dean of Department of Information Technology (Email: g.dimitrov@unibit.bg).

Dimitrios Simeonidis is PhD candidate in the field of Computer Science at the University of Economics - Varna, Varna, Bulgaria. He is currently working as IT Engineer in a private company in Thessaloniki, Greece (Email: dimisime@yahoo.gr).

Article submitted 2021-10-24. Resubmitted 2021-11-30. Final acceptance 2021-11-30. Final version published as submitted by the authors.