

The 8th International Conference "Distributed Computing and Grid-technologies in Science and Education" (GRID 2018)



Contribution ID: 204

Type: Sectional reports

Essential aspects of IT training technology for processing, storage and intellectual analysis of the big data using the virtual computer lab

Monday, 10 September 2018 15:00 (15 minutes)

This paper discusses issues surrounding the training of specialists in the field of storage, processing and intellectual analysis of big data using virtual computer lab and its main architectural components.

Summary

I. INTRODUCTION

Many new professions in the field of the virtual computer lab entered the labor market, corresponding to the needs of the Russian economy. These professions are: a designer of artificial intelligence systems, the big data analyst, specialist on the information promotion, specialist on the machine learning, analyst of the robotized process, digital information manager, developer of the informational system based on blockchain, and even such rare professions as a planner of the interaction with the artificial intelligence and cognitive copywriter. A new training methodology of the skilled IT specialists should be created as the new professions have entered the labor market. The professionals should have enough knowledge and skills to meet the current realities and needs of the business. They should get a job and the social position in the leading sector companies, market leaders of the high-tech goods and the services of new generation. These people will create and leap the economy of the future forward. Nowadays we face the primary task of supplying the software tools and technologies to provide the effective work in the classroom and at home, cause a true amazement and sincere admiration for the progress in math and computer science, making the students more self-confident and giving the reasons for the future actions. We should create sustainable development not only at the national level, but also in the globalization and international partnership. The import substitution policy is a short-term strategy, and for a long-term one it is necessary to consider that the education and the development of IT is out of politics as the innovation requires a deep understanding of all the existing achievements of humankind to date.

Accepting the modern challenges, we began creating not only the technical environment but also the space for the knowledge sharing. We draw an analogy of the physical laws and thermodynamics with the laws of functioning of sociotechnical systems. We create the special conditions for the cooperation between the students and teachers by using of our flagship project "Virtual Computer Lab" (VCL), which has been constantly improved for more than 12 years. It gives an opportunity for the students to create and expand own created multi-component corporate information systems both by themselves and in a team, form the teaching aids in a team, making both the freshmen and graduates take part in the process. It significantly increases the cognitive perception of teaching material.

II. BACKGROUND

The training of the «consumers» should be cut off in the process of the IT specialists' education, and we should spare no effort to training of the «creative doers». For this purpose, it is important not only to study the ways of creating the information systems from the scratch, paying attention to the configuring and adjustment of the equipment, connection and integration of all the necessary parts of the system without any help, and only after that to accomplish issue-oriented tasks. The practical study of the modern peer-to-peer protocols is very important nowadays. The students get acquainted with the approaches to the improvement of the existing systems, not changing end-user presentation level. They are taught how to setup modern systems with the horizontal scaling for the data storage, distributed map-reduce analytics, OLAP analysis based on

the materialized views, which accelerate the data output in the business intelligence system without decreasing of the reliability and increased cost as compared to in-memory solutions. In addition to the training of programmers in the development of the mobile solutions and cloud services, it is necessary to focus attention on the training of the programmers in the neural network solutions using open source frameworks such as TensorFlow, Keras, OpenAI. The training should involve the use of the modern processor designs including AVX, FMA, SSE4x instructions and technologies of the distributed computing MPI, CUDA, OpenCL for the effective solving the tasks of the cognitive informatics and machine learning.

The expert of the future is an expert, which has not only the fundamental scientific knowledge, but he is a promising engineer with an outstanding potential and is able to compose and make the capable computing solutions suitable for the project. Only the skilled professionals of this level can create the right conditions for the science development and its practical applications at an increasing rate.

All above-mentioned problems can be solved in the virtual computer lab, which has become not only the innovative tool for the training of the high skilled IT specialists, but also a demanded space for the technical cooperation between a final-year student and a potential employer. It gives an opportunity to show your qualification in real time, and to present the employer's problem in the virtual format and try to solve it together, attracting the young minds and sometimes people with different ways of thinking, for example, the history of the neural network expansion and the idea of calculation of the back propagation errors, using the gradient descent method and so on.

That is why the priority of the university is to create the most favorable conditions for the forming of the professional competence in IT, which will help the graduates to solve a wide range of the tasks, happening during all the stages of the corporate information systems development, including the design itself. It is evident that to form the professional competence the students should do the following in order master a lot of literature, do many practical tasks and make research works on the modern information systems, their deployment, maintenance and effective appliance for solving the problem-oriented tasks and so on.

The following problems had to be solved for the effective target training of IT specialists: a lack of class hours for solving the necessary and sufficient practical tasks of the complex information systems studying; it's impossible to get the work experience in the complex information systems by use of the personal computer with an average capacity as such systems demand different requirements of the hardware in comparison with the home, office and portable computers; one sometimes has problems during the setup and maintenance of the information systems, these tasks cannot be solved without the work experience in such systems; the price of some products licenses is extremely high for a user, in most cases, one needs a license only for the educational process.

The main way to solve these problems has been to create a virtual computer lab that is able to solve the problem of insufficient computing and software resources and to provide an adequate level of technological and methodological support; to teach how to use modern technologies to work with distributed information systems; to organize group work with educational materials by involving users in the process of improving these materials and allowing them to communicate freely with each other on the basis of self-organizational principles.

III. BRIEF CONCEPT OF USING VIRTUAL COMPUTER LAB

The Virtual Computing Lab provides a set of software and hardware-based virtualization, containerization and management tools that enable the flexible and on-demand provision and use of computing resources, knowledge management system, theoretical materials and practical cookbooks in the form of cloud services for carrying out research projects, scientific computational calculations and tasks related to the development of complex corporate and other distributed information systems. The service also provides dedicated virtual servers for innovative projects that are carried out by students and staff at the Institute of System Analysis and Control of State Dubna University.

One main distinguishing trait of the Virtual Computer Lab is its self-organizing principles, which make it possible to transition students from a rigid system of group security policies to a new system where each student can develop a sense of personal responsibility, respect for colleagues, and tolerance, which should provide a solid foundation for strengthening and developing basic civilizational values in the education environment. Thus, today the need has arisen to incorporate technologies into the educational process that will contribute to global integration in the foreseeable future.

It is not arbitrary that education that is conducted through high-availability distributed information systems is a priority, because these types of software solutions have become an integral part of modern business. That's why the task of designing and deploying failover clusters forms the topic of several special courses, which are designed to satisfy the demand for these skills by modern companies. When designing corporate information systems and ensuring the availability of critical applications that are independent of a hardware and software environment, it is critically important to ensure the successful implementation of many key business processes. Downtime, including for scheduled maintenance, leads to additional costs and the loss of customers, and the long outages are simply unacceptable for modern high-tech enterprises.

IV. FUNDAMENTAL COMPONENTS OF VIRTUAL COMPUTER LAB

Modern blade servers are the hardware components that support virtual computer labs. They are high-performance, high-capacity, but compact and allow the space in the server room to be used more efficiently. The software platform of the Virtual Computing Lab is implemented based on the VMware vSphere Software,

which consists of vSphere ESXi hypervisors with some hand-made enhancements and optimizations for some specific hardware that handle all the computing work of the virtual machines as well as vCenter Server central management servers.

The vCenter Server consists of the following key components:

vCenter Single Sign-On. This component is critical to the whole environment, since it provides secure authentication services for many vSphere components. Single Sign-On creates an internal secure domain in which the various components and solutions that are included in the vSphere ecosystem are registered during the installation or upgrade process, and subsequently they will be assigned basic infrastructural resources. Within the VCL architecture this component is responsible not only for internal authentication services, but it is also used to authenticate users from the university's internal domain who have Microsoft Active Directory accounts at the university.

vCenter Server. The vCenter Server component is a central component that is used to manage the vSphere environment. This module provides management and monitoring interfaces for several vSphere nodes, and it also enables the use of such technologies as VMware vSphere vMotion and VMware vSphere High Availability.

vCenter Inventory Service. Approximately ninety percent of vSphere Web Client requests to the server are just requests to read the current configuration of the system and its state. The Inventory Service is a component that caches most of the information about the current state of the environment to respond to vSphere Web Client requests to reduce the load on vCenter basic processes.

vSphere Server for Web Client (vSphere Web Client). vSphere Web Client is the main interface that is used to centrally manage the environment. It can be divided into two parts: the first server part, which serves requests from the second part, which is the end user's Adobe Flex compatible browser with support for NPAPI-plugins. It is worth noting that the VCL may also be managed using the vCenter Server Desktop Client that is installed on the end user's computer.

vCenter Server Database. The database is one of the key modules in the vCenter Server stack architecture. Almost every request sent to the vCenter Server entails communicating with the database. This database is the main storage location for vCenter Server parameters, and it is also a repository of statistical data. Saved statistical data make it possible to optimize system performance during subsequent analysis.

The NVidia Tesla, Volta, Pascal, Maxwell GPUs could be used for 3D virtualization and VMware Horizon Suite is used for remote VDI connections as well as for creating images of virtual servers and workstations that are separated into layers using VMware ThinApp and for managing these images. This solution is very important for machine learning due to significant increasing of neural networks training speed.

A centralized management portal as well as a knowledge management system were created to improve productivity of work in the Virtual Computer Laboratory. The need to create such a system was conditioned by the fact that students are able to improve productivity of remote learning by themselves, so it is important to create a social network between all participants as well as to create an environment that allows pupils the opportunity to independently engage in such processes as the identification, acquisition, presentation, and use (distribution) of knowledge without the direct involvement of the instructor.

Methods of use (propagation) are directly related to storage methods and, consequently, the technological tools that may be used for the transmission of formal knowledge include knowledge bases with various search functionality; blogs, wikis, and social networks; "Wiki Textbooks" that allow all participants to collaboratively create and update educational content and exchange practical problems (including from real companies); as well as user blogs, forums, and group chat systems.

The new practice with containers is different compared to VMware case and effectively complements it for a wide range of practical tasks. For the underlying operating system kernel can be used for all containers. On the one hand, it introduces restrictions on the use of other operating systems while, on the other hand, it improves payload on the north of a similar configuration. This can be achieved due to the specifics of the containerization architecture, which we will examine on the example of Docker.

Docker uses a client-server architecture in which the Docker-client interacts with the Docker daemon, enabling the operations of creating and launching containers on the server and providing them to students. In general, a containerization system can be represented in the form of three key components: images, registries, and containers. Images represent read-only templates that contain an operating system based on the same kernel version as the host system with necessary pre-configured and adapted software. These images are created, modified if necessary, and then used for deployment of individual solitary containers. The images are stored in the registry, which is a tool for their storage and distribution. The registry content corresponds the curriculum and laboratory plans prepared by the teaching staff.

The containers per se are, in fact, like catalogues (directories) of an operating system, where all the changes made by the user and the system software while work are stored. Each container installed from an image provides the capacity for fast creation, start, stop, move, and delete. It also works as a safe sandbox for running applications, allowing the student to carry out any experiments without compromising the base operating system, while maintaining the highest level of performance. Current evolution of VCL lead to development of design templates for both corporate IT deployment and students learning project.

V. CONCLUSION

It should also be emphasized that the virtual computer lab has helped us provide an optimal and sustainable

technological, educational-organizational, scientific-methodological, and regulatory-administrative environment for supporting innovative approaches to computer education. It promotes the integration of the scientific and educational potential of Dubna State University and the formation of industry and academic research partnerships with leading companies that are potential employers of graduates of the Institute of System Analysis and Control.

The results that the Institute of System Analysis and Control has achieved in improving the educational process represent strategic foundations for overcoming perhaps one of the most acute problems in modern education: the fact that it tends to respond to changes in the external environment weakly and slowly.

Primary authors: Mrs CHEREMISINA, Evgenia (Dubna International University of Nature, Society and Man. Institute of system analysis and management); Mr BELOV, Mikhail (Dubna State Univeristy); TOKAREVA, Nadezhda (Dubna Univeristy); KRYUKOV, Yury (Alekseevich)

Presenter: Mr BELOV, Mikhail (Dubna State Univeristy)

Session Classification: Scientific, industry and business applications in distributed computing systems, education

Track Classification: 4. Scientific, industry and business applications in distributed computing systems