Журнал Белорусского государственного университета. Математика. Информатика. 2021;3:104–108
Journal of the Belarusian State University. Mathematics and Informatics. 2021;3:104–108

УДК 002.6;004.7;004.722

ПРОЕКТИРОВАНИЕ ЗАЩИЩЕННОГО ОТКАЗОУСТОЙЧИВОГО ОБЛАЧНОГО РЕПОЗИТОРИЯ ПИСЬМЕННЫХ РАБОТ ОБУЧАЮЩИХСЯ И СОТРУДНИКОВ УЧРЕЖДЕНИЙ ОБРАЗОВАНИЯ

В. П. КОЧИН[1)]***, А. В. ЖЕРЕЛО***[1)]

[1)]*Белорусский государственный университет, пр. Независимости, 4, 220030, г. Минск, Беларусь*

Показаны основные подходы к проектированию и разработке автоматизированной информационной системы защищенного облачного репозитория письменных работ обучающихся и работников учреждений образования и научных организаций (рефераты, эссе, курсовые и дипломные работы, магистерские диссертации, депонированные статьи). Описаны исследования и отработка архитектурных решений для обеспечения надежного и безопасного хранения данных с использованием облачных технологий. Рассмотрены ключевые проблемы проектирования защищенного репозитория и пути их решения. Облачный репозиторий письменных работ построен на базе распределенной файловой системы Ceph. В качестве платформы для создания облачного интерфейса использована система NextCloud, в качестве вычислительной платформы – виртуальные вычислительные ресурсы виртуальной сетевой инфраструктуры БГУ.

Ключевые слова: облачные вычисления; проектирование информационных систем; образовательные технологии; защищенное облачное хранилище; виртуализация; сетевая инфраструктура.

DESIGNING A SECURE FAIL-SAFE CLOUD REPOSITORY OF PAPERWORKS OF STUDENTS AND EMPLOYEES OF EDUCATIONAL INSTITUTIONS

V. P. KOCHYN[a]***, A. V. ZHERELO***[a]

[a]*Belarusian State University, 4 Niezaliežnasci Avenue, Minsk 220030, Belarus*
Corresponding author: V. P. Kochyn (kochyn@bsu.by)

The article discusses the main approaches to the design and development of an automated information system for a secure cloud repository of paperworks of students and employees of educational and research organisations (abstracts, essays, term papers and theses, master's theses, deposited articles), providing secure storage and secure mobile access to stored data. The research and development of architectural solutions to ensure reliable and secure data storage using cloud technologies are described. The main problems of designing a secure repository and ways to solve them are considered. The cloud repository of written works is built on the basis of the Ceph distributed file system, which uses the NextCloud system and the virtual computing resources of the virtual network infrastructure of the Belarusian State University as a platform for building a cloud interface.

Keywords: cloud computing; information systems design; educational technologies; secure cloud storage; virtualisation; network infrastructure.

Образец цитирования:
Кочин ВП, Жерело АВ. Проектирование защищенного отказоустойчивого облачного репозитория письменных работ обучающихся и сотрудников учреждений образования. *Журнал Белорусского государственного университета. Математика. Информатика.* 2021;3:104–108 (на англ.).
https://doi.org/10.33581/2520-6508-2021-3-104-108

For citation:
Kochyn VP, Zherelo AV. Designing a secure fail-safe cloud repository of paperworks of students and employees of educational institutions. *Journal of the Belarusian State University. Mathematics and Informatics.* 2021;3:104–108.
https://doi.org/10.33581/2520-6508-2021-3-104-108

Авторы:
Виктор Павлович Кочин – кандидат технических наук, доцент; начальник Центра информационных технологий.
Анатолий Владимирович Жерело – кандидат физико-математических наук, доцент; заместитель начальника Центра информационных технологий.

Authors:
Viktor P. Kochyn, PhD (engineering), docent; head of the Center for Information Technologies.
kochyn@bsu.by
Anatolii V. Zherelo, PhD (physics and mathematics), docent; deputy head of the Center for Information Technologies.
zherelo@bsu.by

Introduction

In the era of formation and development of the knowledge economy based on the production, distribution, and use of information, the education system is being transformed, responding to the challenges of the present and the future. The request for education from society, the family, and the student himself is changing. Education becomes continuous, mobile, open. At the same time, the design, development, and use of information and communication technologies are not an end in themselves but should provide a software and technical platform for the creation and application of pedagogical innovations [1; 2].

One of the characteristic features of the digital transformation of the education system is a significant increase in the generated digital data. The issue of reliable storage of a sufficiently large amount of data is very acute. Traditional data storage devices (flash drives, hard drives, optical media) do not meet modern requirements for a number of reasons. Firstly, such devices do not provide a sufficient level of data storage reliability. Secondly, it is currently necessary to have access to data from any device: home computer, work computer, phone, laptop. Thirdly, to increase the volume of stored data, the purchase of additional devices is always required. In this regard, recently there has been an increase in the popularity of using cloud data storage.

In 2018, the Belarusian State University was the first among the Belarusian universities to develop a digital transformation strategy, including updating the content, forms, and methods of teaching, changing the processes of scientific research and management by improving its information, and communication infrastructure. One of the objectives of the given strategy is to create conditions for the transition to paperless technologies which allow creating and storing paperworks of students and employees of educational institutions (abstracts, essays, term papers, theses, master's theses, deposited articles) based on cloud technologies [1; 2].

Cloud storage significantly facilitates the work of modern users by virtualising the location of the data [3; 4]. Another advantage is the ability to process data on the storage side, in which the user does not face the tasks of ensuring the reliability and fault-tolerance of the storage, as well as managing the resources that process the data [5]. Due to these advantages, network storage (OneDrive, GoogleDrive, iCloud, Yandex.Disk, DropBox, etc.) is widely used for data storage and processing. On the one hand, these services provide reliable, fault-tolerant storage, on the other hand, they have a number of disadvantages:

- these resources are commercial, which are provide only a service, hiding the details of implementation;
- data is stored on servers located outside the Republic of Belarus, which contradicts the Decree of the President of the Republic of Belarus No. 60 «On measures to improve the use of the national segment of the Internet» dated 1 February 2010;
- they are inable to integrate with existing automated systems of universities, which makes it impossible to integrate these services into the educational process.

In this regard, there is a need to create your own cloud storage. When designing a secure fault-tolerant cloud repository of students' paperworks, the following criteria must be taken into account:

- providing multiple user access to resources;
- ensuring reliability and fault-tolerance;
- hardware independence and ability to quickly scale computing resources and storage systems;
- ability to generate analytical reports and various search queries;
- integration possibility with corporate information systems;
- ability to check for anti-plagiarism on stored data.

Providing multiple user access to resources. The bottleneck of traditional storage systems is the performance of a particular disk. When multiple users access the disk at the same time, its performance is divided by all.

Ensuring reliability and fault-tolerance. The designed storage should ensure the operability of the system even in case of failure of individual hardware nodes.

Hardware independence and ability to quickly scale computing resources and storage systems. At the system designing, it is necessary to ensure the possibility of creating and scaling the system on servers of various manufacturers. This approach will allow, firstly, to use existing equipment, and, secondly, to gradually increase computing and storage resources without reference to existing solutions.

Ability to generate analytical reports and various search queries. When designing a secure fault-tolerant cloud repository, it is necessary to provide for the possibility of generating reports according to various criteria. It is also necessary to develop of a search module for stored data.

Integration possibility with corporate information systems. When designing a secure fault-tolerant cloud repository, it is necessary to provide for the possibility of integration with corporate information systems [3–5]. This is due to the following main factors:

- access to the storage must be provided based on user data from active directory;
- the ability to store and process documents and files from various information systems and services must be provided;

• access to the repository must be personalised and confirmed by the appropriate authority. For example, a user with the student role should have access to their own repository, a user with the teacher role should have access to the works of their students. The corresponding user role should be assigned automatically based on data from automated personnel and student management systems.

Ability to check for anti-plagiarism on stored data. To check uploaded students' paperworks, it is necessary to provide an interface for integration with anti-plagiarism testing systems. It will be improve the quality of term papers and theses.

Results and discussion

The information system of a secure cloud repository of students' paperworks and employees of educational and research organisations is a set of architectural solutions and software designed to ensure reliable storage of information using cloud technologies and cloud services that provide secure access to resources stored in the cloud to mobile users, regardless of the hardware and software platform used.

By organising cloud storage at the Belarusian State University, the emphasis was placed on the use of a distributed file system.

This was done due to the following disadvantages of hardware storage systems:

• the limitation of the total bandwidth of interaction with the storage system, as noted above. For example, according to the technical documentation, for the Lenovo DE6000H system, the data reading bandwidth reaches 21 Gbit/s. Obviously, with the growing number of clients, especially connecting from the outside, this bandwidth will not be enough. Using distributed file systems, the independence of requests sent to the storage is essential. Although the performance of each individual connection will be lower than in the system mentioned above, the cumulative data flow to (from) the system can be practically unlimited. This is achieved due to the possibility of increasing access points to a distributed file system resource and organising alternative access paths to data storage. The bandwidth of which is practically unlimited and in total can be significantly exceed the limit of several tens, hundreds or even thousands of gigabits per second;

• the need to localise the storage system and the systems using it within the same geographic location. Traditional data storage systems, as usual, require placement in the same data center where the consumer of their data is located. In addition, even in this case, the distance from the consumer of information to the storage system is limited. The organisation of distributed storage requires additional resources both software and hardware solutions (as an example is the idea of metro-cluster). Due to the reasons mentioned in point one, namely the separation of information transmission flows, as well as, due to some typical structure of services used in distributed file systems and virtualisation of the data access point. The proposed structure allows us to form a storage cluster more dependent on the quality of communication channels, but loosely related to management and maintenance tasks, and also allows us to increase storage space almost indefinitely.

As the basis of such a repository, it is advisable to choose one of the free distributed open-source systems.

The automatic identification system is built from the following components:

• hardware and software platform for the organisation of a fault-tolerant data storage with distributed management and data networks;

• a system for managing and monitoring the state of a fault-tolerant storage;

• IaaS platform based on data center resources for deploying microservices of the cloud storage system;

• a set of microservices images that provide a cloud storage user interface.

By choosing technologies and solutions, one of the main requirements was the use of solutions based on publicly available technologies and protocols and presented in the form of freely distributed source code.

The overall architecture of the system is shown in figure.

Fault-safe distributed storage module. The fault-safe distributed storage module was developed on the base of the Ceph distributed file system[1], as it has the following main implementation features:

• the system combines the resources of several servers by merging them into a single cluster with a single management system;

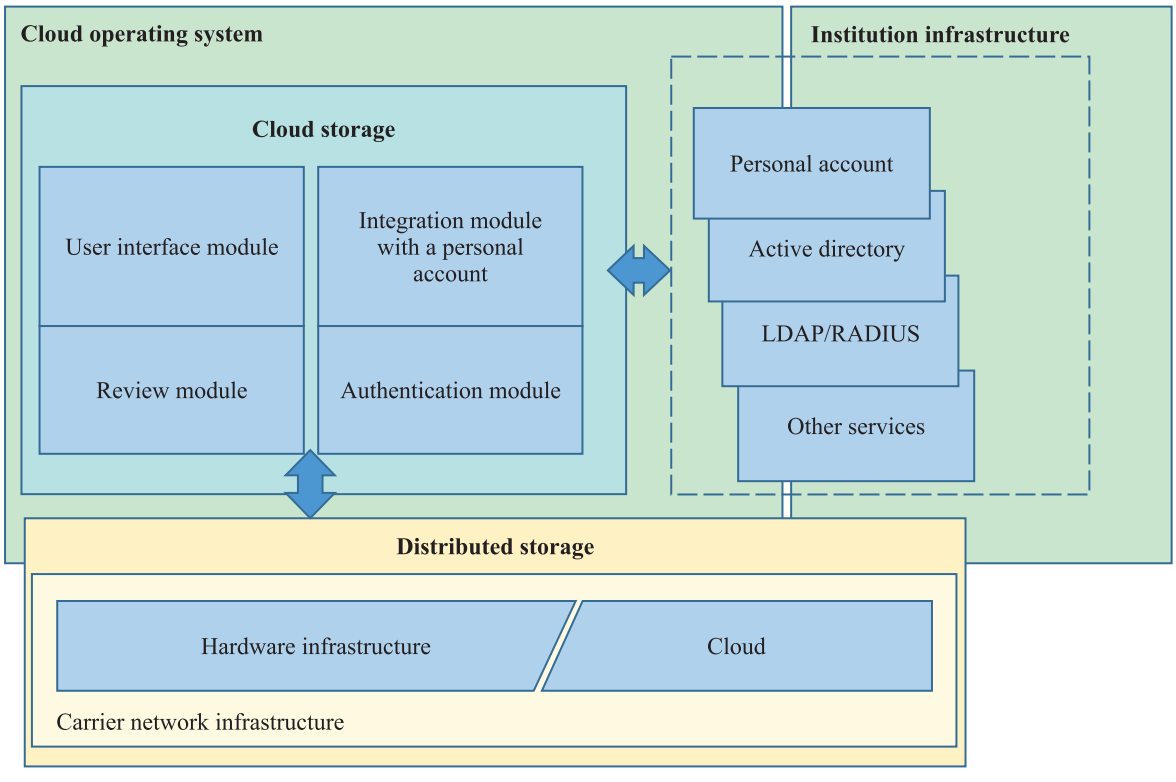
• data is stored as blocks (similar to a traditional file system), and the system continuously check the status of these blocks by monitoring and replicating these blocks on the fly, if it is necessary;

• Ceph provides a variety of interfaces for accessing stored data, which is also the subject of research to identify more productive and (or) more resilient to failures mechanisms;

• Ceph implements a web-based cluster management interface with status monitoring elements.

During the installation of Ceph cluster servers, CentOS Linux release 7.7.1908 was selected as the base operating system, kernel version 3.10.0-1062-el7.x86_64. Using the SSH service, the trust relationships necessary for Ceph deployment were set up between the cluster servers.

[1]Официальный сайт Ceph [Электронный ресурс]. URL: https://ceph.io/ (дата обращения: 10.03.2021).



General system architecture

By creating the storage module, the DHCP and NIS services were additionally deployed. The use of these services makes it possible to reduce the costs of configuring and subsequent maintenance of a fault-tolerant storage system due to the mechanism of centralised distribution of information necessary for the functioning of the system. Maintaining such as services allows, for example, to reduce the time for adding a new server to an existing storage system. Since a set of standard software is installed on the new server, and the necessary settings for integrating the server into the general system are transmitted by these services.

The DHCP server is used to distribute the network interface settings of the cluster servers. In the current configuration, two subnets are declared, corresponding to the segments described in the previous section.

In contrast, the NIS server is used to distribute configuration files which are necessary for the functioning of the cluster, in particular information about some accounts required for automated maintenance of the storage system, server names, etc.

Virtual platform for building a cloud interface. Based on the analysis of existing open solutions, the NextCloud system was chosen as the basis for the created platform that provides a cloud interface for accessing fault-tolerant storage[2]. The choice in favour of NextCloud was made based on the requirements specified in the introduction. At the moment also it is the only opensource solution in terms of functionality comparable to proprietary cloud storage. There are other cloud storages, for example, SeaFile[3], but it is not yet possible to consider them, and even to compare them, because they are in the initial phase of their development [6]. Additional complexity by choosing a platform is associated with the need to install not only the selected solution on some server, but to create an image of a virtual machine independent of the cloud platform on which such a virtual machine can be deployed in the future. In this regard, CentOS Linux GenericCloud 1907 OS was chosen as the main operating system for the virtual machine, which in turn is a cloud implementation of CentOS Linux release 7.6.1810 (Core) OS with Linux kernel version 3.10.0-957.27.2.el7.x86_64. The choice of the Linux dialect is not critical, since the system being deployed can work with any modern Linux implementation.

To ensure free migration between different cloud environments, the original image was converted to VHD format, which allows it to be run on most of the virtual environments, in particular, OpenStack and Windows Hyper-V, using which cloud solutions currently operate at the Belarusian State University.

Note that the module being created provides only interface interaction between the user and the secure storage and does not require large amounts of disk space. As a result, the time for deploying and launching a new virtual machine image is reduced, if it is necessary.

[2]Официальный сайт NextCloud [Электронный ресурс]. URL: https://nextcloud.com/ (дата обращения: 01.10.2021).
[3]Официальный сайт SeaFile [Электронный ресурс]. URL: https://seafile.com (дата обращения: 01.10.2021).

Conclusion

The developed solutions for creating a cloud repository include various levels of service provision and allow using the results obtained both in the construction of cloud repositories for educational institutions and in the design of individual modules of fault-tolerant systems. The developed secure fail-safe cloud repository of students' paperworks allows to provide the mass user access to the system and reliable storage, create a single cloud, while it can be geographically distributed.

Accordingly, the applied approaches can be used in various sectors of the economy for safe and reliable storage as well as for processing of various documents under the legislation of the Republic of Belarus.

Библиографические ссылки

1. Король АД, Воротницкий ЮИ, Кочин ВП. Дистанция в образовании: от методологии к практике. *Наука и инновации.* 2020;6:22–29.

2. Король АД, Воротницкий ЮИ, Кочин ВП. Информационно-коммуникационные технологии дистанционного и онлайн-обучения. В: Тузиков АВ, Григянец РБ, Венгеров ВН, редакторы. *Развитие информатизации и государственной системы научно-технической информации (РИНТИ-2020). Материалы XIX Международной конференции; 19 ноября 2020 г.; Минск, Беларусь.* Минск: ОИПИ НАН Беларуси; 2020. с. 22–29.

3. Кочин ВП, Воротницкий ЮИ, Жерело АВ. Виртуализация сетевой инфраструктуры учреждений образования. *Цифровая трансформация*. 2020;1:51–56.

4. Кочин ВП, Жерело АВ. Виртуализация сетевой инфраструктуры Белорусского государственного университета. *Вестник компьютерных и информационных технологий.* 2020;17(8):45–51. DOI: 10.14489/vkit.2020.08.pp.045-051.

5. Курбацкий АН, Кочин ВП, Слесаренко ОВ. Проектирование и автоматизация работы облачной кластерной системы с учетом интеграции с внешними информационными системами. *Веснік сувязі.* 2021;2:56–61.

6. Кочин ВП. *Разработать технологии аутентификации и авторизации пользователей в образовательных сетях на базе смарт-карт (отчет о научно-исследовательской работе (заключительный)).* Минск: БГУ; 2020. 68 с. № государственной регистрации 20163472.

References

1. Korol' AD, Vorotnitskii YuI, Kochyn VP. [Distance in education: from methodology to practice]. *Nauka i innovatsii.* 2020;6: 22–29. Russian.

2. Korol' AD, Vorotnitskii YuI, Kochyn VP. [Information and communication technologies of distance and online learning]. In: Tuzikov AV, Grigyanets RB, Vengerov VN, editors. *Razvitie informatizatsii i gosudarstvennoi sistemy nauchno-tekhnicheskoi informatsii (RINTI-2020). Materialy XIX Mezhdunarodnoi konferentsii; 19 noyabrya 2020 g.; Minsk, Belarus'* [Development of informatisation and the state system of scientific and technical information (RINTI-2020). Materials of the 19th International conference; 2020 November 19; Minsk, Belarus]. Minsk: Joint Institute for Informatics Problems of the National Academy of Sciences of Belarus; 2020. p. 22–29. Russian.

3. Kochyn VP, Vorotnitsky YuI, Zherelo AV. Virtualization of the network infrastructure in educational institutions. *Tsifrovaya transformatsiya.* 2020;1:51–56. Russian.

4. Kochyn VP, Zherelo AV. Virtualization of the network infrastructure of the Belarusian State University. *Vestnik komp'yuternykh i informatsionnykh tekhnologii.* 2020;17(8):45–51. Russian. DOI: 10.14489/vkit.2020.08.pp.045-051.

5. Kurbatskii AN, Kochyn VP, Slesarenko OV. Design and automation of the cloud cluster system taking into account the integration with external information systems. *Vesnik suvjazi.* 2021;2:56–61. Russian.

6. Kochyn VP. *Develop technologies for user authentication and authorization in educational networks based on smart cards (research report (final)).* Minsk: Belarusian State University; 2020. 68 p. State registration No. 20163472. Russian.

Received 15.10.2021 / revised 19.10.2021 / accepted 26.10.2021.