UDC 004.9 **doi: 10.26906/SUNZ.2023.1.101**

A. Kovalenko, R. Miroshnychenko, A. Martyntsov

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

DISTRIBUTED COMPUTING SYSTEMS BASED ON THE USE OF GRID TECHNOLOGIES

Abstract. The article describes the use of Grid technology in telecommunication systems and the creation of Grid telecommunication systems. Network service providers are focusing on huge consolidated data centers. Particular attention began to be paid to improving the methods of scheduling data processing tasks. The shortcomings of modern approaches to planning are considered. These include the inability to provide the maximum total priority of tasks performed at individual planning stages in distributed computing systems. The developed planning procedure is described. It overcomes this shortcoming. But at the same time, to reduce the total processing time of tasks in the system in comparison with the planning procedure based on solving the problem of the least coverage. The desire for the maximum sum of priorities of the selected tasks is the main criterion for selecting tasks from the queue, which we will characterize by the importance coefficient. The coefficient of preservation of importance and the coefficient of acceleration of the operation of the Grid-system segment are described. The article also proposes a solution to improve the efficiency and quality of task servicing. It is proposed to expand the functionality of distributed telecommunication systems based on the use of computer clusters using Grid technologies.

Keywords: Grid systems, Grid technologies, telecommunication systems.

Introduction

Modern computing systems represent a set of communication channels uniting them. They also represent a distributed environment for information processing and computing. The complication of various applications leads to an increase in the workload that are performed in distributed environments. On the one hand, there is a need to improve the performance of organization systems for more efficient use of resources and organization in a distributed computing environment. On the other hand, providing users with a wide range of distributed computing systems and various services. This additionally imposes requirements on the organization of distributed information processing and computing. Thus, modern distributed computing systems represent a distributed computing environment. Which is intended both for information processing and organization of calculations in a distributed environment. The term distributed computing or metacomputing was introduced in 1987 by Larry Smarr and Charles Catlett. This approach is based on the integration of computers through a data transmission network of specialized system software into a distributed computing system. The purpose of such a system is to provide equal access for users to its resources.

Resources - a set of software and hardware for the execution of processes. Resource examples: processor, communication medium, application software, storage system, etc. A process is a stream of instructions for a computer processor with a single address space, processor register values, stack, open files, global variables, etc.

The processors on the nodes of a distributed computing system often manage individual resources. We will consider the generalized concept of a virtual node (VN). Thus, VN is one computational element (processor core), considered in conjunction with comparisons of resources to it. If several processors use the same resources characterized by volume (RAM, permanent memory, etc.), then we will assume that the volume of these resources is evenly divided among all virtual nodes corresponding to these processors. Figure 1 schematically shows the ratio of VN and resources in one computing node.

The phrase "Other resources" in Fig. 1 denotes additional capabilities of computing nodes. For example, the presence of graphics or signal processors, specialized electronic boards for accelerating calculations.

Main part

A distributed computing system may include computing systems of different architectures. A description of the features of modern computing architectures is available in [1]. Features of the architecture of modern high-performance systems are described in.

The most common implementations of distributed computing systems are systems built using Grid technology [1, 2]. The definition of the term Grid is currently not consistent. We will adhere to the definition of Ian Foster and Carl Kesselman [1]: "Grid is a consistent, open and standardized environment that provides flexible, secure and coordinated resource allocation within a virtual organization." A virtual organization is understood as an association of users and real organizations, which is a subset of resources in the Grid. The purpose of introducing such a union is the need for separation of powers in the Grid between users.

In most cases, computational clusters act as Grid nodes. They are located remotely and differ in their characteristics. In the general case, a computing cluster is a parallel scalable computing system and includes a set of high-performance computers (cluster processor nodes) connected by communication networks and controlled by a single task manager [3, 4].

When integrating processor nodes into a single inter-node cluster, communications are carried out by passing messages between task processes.

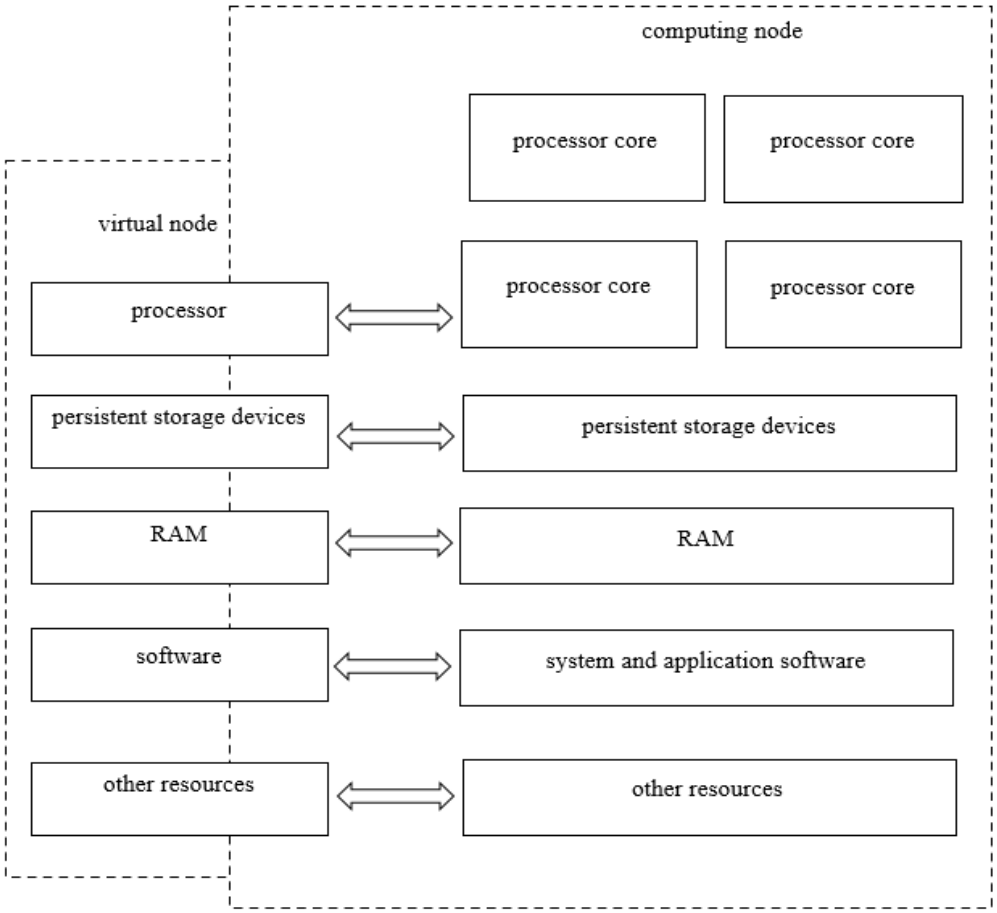

Fig. 1. Relationship between virtual and physical nodes of a distributed computing system

In addition to computing clusters, other types of nodes can be included in the Grid. For example, visualization workstations or network storages. The term Grid is based on the electrical power grid metaphor, which provides users with electricity on demand without the need for knowledge of where and how it is generated.

Access to resources in the Grid is usually carried out through services [5] present on each of the nodes. Services must be standardized within the Grid. This allows uniform access to all resources. Absolutely regardless of its type and local policies on the site. The Open Grid Services Architecture standard for building Grid services unifies the creation, naming, and interaction of services.

The Globus Toolkit has become the de facto standard for building a Grid based on the Open Grid Services Architecture. This software package includes task management services. For monitoring and coordinating the remote execution of tasks. Also, data collection services based on the use of the LDAP protocol. Grid Security Infrastructure Security Services. Its components are based on X.509 digital certificate technology and data management services. They allow users to access, transfer and manage distributed data.

In a computing system, applications can be divided into three classes: distributed, parallel, and non-parallel.

Distributed Grid applications (multi-site Grid applications) - a class of applications that can be run on multiple Grid nodes. In other words, the processes of one distributed application can be launched on several computing clusters simultaneously.

Distributed applications are practically not used in the Grid due to the lack of support from the Grid toolkits. The main difficulty here is that computing clusters are controlled by their local resource management systems. Thus, the control of all processes of a task distributed among several clusters becomes more complicated.

There is another reason for the low prevalence of this class of applications. The difference between the speed characteristics of data transfer between nodes of the same cluster and between several clusters can be very different. Accounting for this property of data transmission networks complicates the distributed application method.

Also, the heterogeneity of the architecture of processors in the composition of Grid sites can become an obstacle to the development of the sphere of distributed applications. When running a program, it is useful to compile the source code for a given processor architecture after scheduling and running the job on the Grid node. Such compilation in some cases can significantly speed up the application by optimizing the binary code for a given processor architecture. Obviously, in this case, the distributed application must be compiled and run in several different copies. This makes it difficult to design such applications.

In the vast majority of cases, it is possible to decompose a distributed algorithm into several parallel

or non-parallel ones. After the completion of the calculations of many such tasks corresponding to one distributed task, the received data is collected and processed separately.

Parallel applications belong to a different class. Unlike distributed applications of this class, they are designed to run on a homogeneous computing cluster or a multiprocessor system with shared memory.

A program with the number of processes will be called parallel N > 1, concurrently executed on N homogeneous processors. The processes of a parallel program can potentially exchange data with each other as over a data transmission network. It is possible to use the mechanisms of interaction of processes within the framework of the same operating system (Inter-Process Communication). All processes of a parallel program must be scheduled to run simultaneously. This type of planning is called gang-scheduling.

In the general case, a parallel task is a user instruction to run one parallel program with certain resource requirements. It gives her all the data files necessary for the calculation. In scheduling parallel tasks for Grid, it is important that the processes of one program do not go beyond one VN group - the Grid node. An example of a parallel task is the execution of an MPI application on the nodes of a computing cluster.

A program that includes only one computational process will be called non-parallel. A non-parallel task would be a user guide to execute one or more non-parallel programs between which there is no interaction. Often such a task involves running multiple copies of the same program with different input parameters.

In non-parallel applications, a control program is often used. It collects the results of the calculations and issues a new portion of the input data. This type of architecture is called Supplier-consumer [1]. Isolated computing processes often make it easier to develop and maintain applications. Examples of algorithms with isolated processes are Map-Reduce [1, 2] and a family of k-means algorithms [6], which are used in data analysis methods, mathematical statistics, and other fields of knowledge.

Conclusions

Therefore, the principal direction in the development of modern distributed computing systems is the development of appropriate methods of planning models. Therefore, increasing the efficiency and quality of servicing tasks, characterized by the coefficient of importance, the coefficient of acceleration of the tasks. Also disclosed is the expansion of the functionality of distributed telecommunication systems based on the use of computer clusters using Grid technologies. The goal is achieved by developing a method for scheduling the execution of tasks.

They are based on planning the execution of tasks, solving the problem of non-linear Boolean programming, developing a solution method based on the rank approach. This approach improves planning efficiency.

REFERENCES

1. Foster I. The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration. – Argonne National Laboratory, Argonne, Ill., 2002– 57 p.
2. Kovalenko, A. and Kuchuk H. (2018), “Methods for synthesis of informational and technical structures of critical application object’s control system”, *Advanced Information Systems*, 2018, Vol. 2, No. 1, pp. 22–27, DOI: https://doi.org/10.20998/2522-9052.2018.1.04
3. Yakovleva, O., & Nikolaieva, K. (2020). Research of descriptor based image normalization and comparative analysis of SURF, AKAZE descriptors. *Advanced Information Systems*, *4*(4), 89–101. https://doi.org/10.20998/2522-9052.2020.4.13
4. Ruban, I.V., Martovytskyi, V.O., Kovalenko, A.A. and Lukova-Chuiko, N.V. (2019), “Identification in Informative Systems on the Basis of Users' Behaviour”, Proceedings of the International Conference on Advanced Optoelectronics and Lasers, CAOL 2019-September,9019446, pp. 574-577, DOI: https://doi.org/10.1109/CAOL46282.2019.9019446
5. Rukkas, K., & Zholtkevych, G. (2020). Probabilistic model for estimation of cap-guarantees for distributed datastore. *Advanced Information Systems*, *4*(2), 47–50. https://doi.org/10.20998/2522-9052.2020.2.09
6. Foster I., Kesselman C., Nick J.M. and Tuecke S. Grid Services for Distributed Systems Integration // IEEE Computer.– 2002.– № 6.– P. 37-46.

Received (Надійшла) 24.12.2023
Accepted for publication (Прийнята до друку) 22.02.2023

Розподілені обчислювальні системи на основі використанням Grid технологій

А.А.. Коваленко, Р. О. Мирошниченко, А. Ф. Мартинцов

Анотація. У статті обґрунтовано використання та впровадження технології Grid у телекомунікаційні системи та створення телекомунікаційних Grid систем. Постачальники мережних сервісів, спираючись на великі консолідовані центри обміну даних, особливу увагу стали приділяти вдосконаленню методів планування завдань обробки даних. Розглянуто недоліки сучасних підходів планування, до них можна віднести відсутність можливості забезпечити максимальний сумарний пріоритет виконуваних завдань на окремих етапах планування в розподілених обчислювальних системах. Описана розроблена процедура планування, що дозволяє подолати зазначений недолік, при цьому зменшити сумарний час обробки завдань в системі в порівнянні з процедурою планування на основі рішення завдання про найменше покриття. Прагнення до максимальної суми пріоритетів обраних завдань є головним критерієм при виборі завдань з черги, який будемо характеризувати коефіцієнтом важливості, коефіцієнтом збереження важливості та коефіцієнтом прискорення роботи сегмента Grid - системи. Також у статті запропоновано рішення щодо підвищенні ефективності та якості обслуговування завдань, а також розширення функціональних можливостей розподілених телекомунікаційних систем на основі використання комп'ютерних кластерів із застосуванням Grid технологій.

Ключові слова: Grid системи, Grid технології, телекомунікаційні системи.