Modeling and optimization of processing large data arrays in information systems

I.Ya. Lvovich[1], Ya.E. Lvovich[2], A.P. Preobrazhenskiy[1], O.N. Choporov[2]

[1]*Voronezh institute of high technologies, Lenina 73a, Voronezh, Russia, 394043*
[2]*Voronezh state technical university, Moscow district 14, Voronezh, Russia, 394006*

Abstract
The development of software was carried out, which is associated with the intellectual support of designers of information systems associated with the processing of data arrays. The formation of a multi-alternative optimization model that takes into account randomized methods, which will be adequate to real information processing processes, has been carried out. This allows the selection of elements of software and hardware information systems. The conducted simulation experiments confirmed the feasibility of the proposed approaches.

Kewwords
Simulation, optimization, Petri nets, randomization, database

1. Introduction

In corporate information systems (CIS), in many cases it is necessary to solve different problems. Among them, we can single out the structural synthesis of distributed databases [1]. First of all, it is important to choose a rational number of parallel processing. In addition, priorities must be assigned in terms of fulfilling the relevant processing conditions.

The purpose of this work is to develop an integrated approach to optimize the processing of large data arrays in an information system based on the aggregation of multi-alternative methods, Petri nets, and fuzzy models.

2. Description of models and methods of processing distributed data arrays

The process of multi-alternative optimization using the transaction hierarchy was considered. The rational number of parallel processing was selected using the apparatus of Petri nets. Priorities related to the fulfillment of processing conditions were also assigned. The choice of promising options for the structure of the system was carried out within the framework of a multi-alternative optimization model [2].

The choice of a rational option was carried out in a multicriteria way within the framework of unclear requirements.

During simulation, it was necessary to determine those parameters that most affect the operation of the EIS. The organization of the simulation model of the system occurs using two hierarchical levels. There is a semantic load for the upper level. The lower level describes the behavior of the components at the upper level.

Arrays of data are received at the input. The simulation model includes three queuing systems (QS).

The model has a data block. In it, the input parameters are streams of data arrays of accounts with failures and a sign of the end of work with data. The number of data arrays is calculated for each QS.

The result is passed to the output. In the model, the values of the parameters were used (Table 1). For the QS block, the vector of numbers of data arrays for each of the data blocks is considered as an input value.

The calculation is carried out for each block. In the processing block, an imitation of entering information into the database takes place. The number of inserted records is remembered. It also computes the time after all data arrays are inserted.

Table 1
Parameters used during simulation

Data block number	Total number of datasets N_{all}	Percentage of data processed by the queuing system 1 P_{QS1}	Percentage of data processed by the queuing system 1 P_{QS2}	Percentage of data processed by the queuing system 1 P_{QS3}
1	6000	50	30	20
2	16000	70	20	10
3	19000	45	25	20

To keep track of the processes, two indicators are provided. In the first, the dependence of the number of records in the database (DB) on the analyzed day is formed. The second indicator shows the average processing time for data sets.

We considered 100 days as the model time. This corresponds to 4 working months of 25 days each. The simulation results demonstrated the filling of the database in a linear fashion. By the end of the working months, we can note the horizontal dependence in the graph. This is determined by the increased activity of the enterprise for such a period.

When analyzing 100 days of model time, the mean processing time was 3383 sec / day (that is, more than 56 minutes / day) and 5230 sec / day (more than 1 hour 27 minutes) for the model with 3 blocks that serve 6000, 16000, 19000 people and having 3 QSs. Possible options for reducing the calculation times are considered.

In order not to have a long processing time for data arrays, the database structure should be simplified. But in this case, the redundancy of information presentation will affect, the data will be presented in a non-optimal way. Then it is necessary to search for a compromise solution during the design of the database. For example, you can transfer static information from a DB to an archive.

3. Conclusion

On the basis of the created integrated approach, it is possible to establish certain dependencies among the parameters characterizing the system for processing data arrays at the enterprise. Then there are opportunities to determine the optimal ways of movement of information flows, which will provide the least load of the system.

4. References

[1] Auziņš, A. Object-Relational Database Structure Model and Structure Optimisation / A. Auziņš, J. Eiduks, A. Vasiļevska, R. Dzenis // Applied Computer Systems. – 2018. – Vol. 23(1). – P. 28-36. DOI: 10.2478/acss-2018-0004.

[2] Lvovich, I.Ya. Optimizing modeling of complex-structured objects in the problem ofimproving the efficiency of their functioning / I.Ya. Lvovich, B.N. Tishukov, A.P. Preobrazhenskiy, O.N. Choporov // IOP Conf. Ser.: Mater. Sci. Eng. – 2019. – Vol. 537. – P. 042047. DOI: 10.1088/1757-899X/537/4/042047.

[3] Moshe Zukerman Introduction to Queueing Theory and Stochastic Teletraffic Models [Electronic resource]. – Access mode: https://arxiv.org/pdf/1307.2968.pdf (13.01.2020).