The optimal placement of information resources on the nodes of a distributed information processing system based on a two-tier and three-tier client-server architecture

Scoba A N, Mikhaylov V K, Ayesh Ahmed Nafea Ayesh, Skorik N S
Platov South Russia State Polytechnic University (NPI), Novocherassk, Russia

Sec-kret@mail.ru

Abstract. This paper presents the results of numerical experiments. The models the optimal allocation of information resources on the nodes of the distributed information processing system (IPS) were compared. They are implemented on the basis of a two-tier and three-tier client-server architecture. The article presents a conceptual model of the functioning of a distributed IPS based on a three-tier client-server architecture. It also shows a mathematical model for determining the stationary probabilities of network states and the average system response time to user requests.

1. Introduction

The problem the optimum location of information resources on LAN nodes is one of the most important tasks. Development and implementation of a distributed information processing system (IPS) are impossible without its solution. This was noted earlier in the article [1]. The task includes the selection of a network topology, an efficiency criterion, the construction of a mathematical model, the development of an optimization algorithm and its software implementation. Publications [2-4] are devoted to this issue. Analysis publications have shown that at present only certain aspects of this problem can be considered. These works have drawbacks: the developed models are separated from the real network structure; the question determining the elements of the matrix, transition probabilities and service rates at network nodes remains open when developing analytical models of data transmission networks; the results of experimental studies and optimization algorithms are missing.

This article presents a conceptual model of the functioning of a distributed information processing system based on a three-tier client-server architecture. There had been obtained the mathematical model for this model. It determines the stationary probabilities of network states using the apparatus of exponential queuing networks. The problem of optimal placement of information resources on the nodes of the computer network is solved using a heuristic algorithm. This placement is made according to the criterion of the minimum average system response time to user requests. The results of numerical experiments are given in comparison with the model of a distributed information processing system. The system is based on a two-tier client-server architecture. Based on the analysis of the results, it is possible to more rationally organize the computational process in the system.

2. Conceptual model

This architecture differs from the two-tier client-server architecture in the physical separation of programs. The database server (responsible for data storage) is separated from the application server (processes data) [5]. This allows optimizing the load on the network and computing equipment. The architecture is implemented: the database server (MySQL server); application server (web server); any browser serves as a client.

The article [5] describes the work of distributed IPS on the basis of a LAN three-tier client-server architecture. It is built as follows: DBMS and DB as a set of files are placed on the hard disk of dedicated computers (database server), business analysis software (business applications) is hosted on application servers; There are many client computers, each of which has a thin client installed — a
client application that implements the user interface. The user has the ability to run an application on each of client computers. This application accesses the business analysis software through the user interface. The software corresponds to a business application hosted on an application server. This business application analyzes user requirements and generates SQL queries to the database located on database servers. Information on SQL queries is copied to the application server. Thereafter, it is returned to the user's client application using business analysis software. This software uses the user interface to display the result of the query on the computer screen.

The conceptual model of the functioning a distributed IPS on the basis of a three-tier client-server architecture differs from a two-tier client-server architecture (a conceptual model is presented in the article [6]). It is modified: devices simulating the operation of database servers are added – $DB_1, ..., DB_s, ..., DB_n$; buffer memory of database servers – $BDB_1, ..., BDB_s, ..., BDB_n$; devices simulating the work of application servers – $SP_1, ..., SP_s, ..., SP_n$; buffer memory application servers – $BSP_1, ..., BSP_s, ..., BSP_n$. A conceptual model of the functioning of a distributed IPS based on a three-tier client-server architecture is presented in figure 1.

Assumptions were made to formalize the formulation of the problem [7]: the network topology is a star; network architecture — three-tier, client-server; request type — one-time; subject area — information service; process type — parallel; information service mode — reading; the way to ensure data integrity is blocking at the level of the entire database. The following model assumptions should be considered:

1) The number of application servers coincides with the number of database servers, i.e. $|SP| = |DB|$.

2) The case is considered when initialization of a business application $r_j$ ($j = 1, s$) requires relationships $R_k$ ($k = 1, d$) such that $r_j \in SP_i \Leftrightarrow R_k \in DB_j$ ($i = 1, n$). It is required to find such a DDB location across database servers so that the average response time of a distributed IPS to requests from thin clients would be minimal.

3. Mathematical model
The characteristics of the distributed IPS that are of interest are determined by the stationary probabilities of the network states. Let $P(i)$ be the stationary probability that the network is in a state...
\( \hat{i} \), where \( \hat{i} = i_{11}, \ldots, i_{1l}, \ldots, i_{n1}, \ldots, i_{nN}, \ldots, i_{nN} \). For a three-tier architecture - \( N = 3n + 1 \), and for

a two-tier architecture - \( N = 2n + 1 \). The process of changing the states of such a network is described by a homogeneous regular Markov process. The equation of the global balance for a stationary mode, the functioning of such a network, according to [1,6], will have the form:

\[
\sum_{k=1}^{N} \sum_{r=1}^{n} P(i) \mu_{kr} = \sum_{l=1}^{N} \sum_{k=1}^{n} P(i_{lr} - 1) \mu_{lr} P_{lk}(r),
\]

where \( \mu_{kr} (r=1,n, k=1,N) \) is service intensity in the \( k \)-th message center of the \( r \)-th user; \( \mu_{lr} (r=1,n, l=1,N) \) is service intensity in the \( l \)-th message center of the \( r \)-th user; \( P_{lk}(r) \) is the probability that the message of the \( r \)-th user after servicing in the \( l \)-th center falls into the \( k \)-th center; \( 1_{lr}, 1_{kr} \) - vector, in the \( k \)-th and \( l \)-th coordinate of which, respectively, on the \( r \)-th place is 1, and all other values are zero.

Calculation of the average response time of the system to the requests of "thin clients" is made by the formula:

\[
\bar{T} = \frac{1}{\sum_{s=1}^{n} \lambda_s} \sum_{s=1}^{n} \lambda_s \bar{T}_s,
\]

where \( \lambda_s (s=\overline{1,n}) \) - the intensity of query formation by the \( s \)-th «thin client»; \( \bar{T}_s (s=\overline{1,n}) \) - the average response time of the system to the request of the \( s \)-th «thin client». The quantity \( \bar{T}_s \) is defined as \( \bar{T}_s = \frac{N_s}{\lambda_s} \), where \( N_s (s=\overline{1,n}) \) is average number of requests of the \( s \)-th «thin client»; \( \lambda_s (s=\overline{1,n}) \) is average intensity of query formation by the \( s \)-th «thin client». The quantities \( N_s \) and \( \lambda_s \) define, as: \( N_s = 1 - P_s(d) \); \( \lambda_s = \lambda_s P_s(d) \), where \( P_s(d) \) - the probability that the \( s \)-th «thin client» is in the active state (forms a query). It was shown in [6] that the calculation of the quantity \( \bar{T} \) by formula (1) reduces to the calculation of the normalizing constant \( G(N_1, \ldots, N_n) \), for which Buzen’s recursive method can be used [8,9].

4. Results of numerical experiments

The paper [10] presents a heuristic algorithm for solving the problem of optimal placement of information resources on the nodes of a distributed information processing system by the criterion of the minimum average response time of the system to user requests. The algorithm is developed by the authors of this paper [11].

Present algorithm and the calculation of the value \( \bar{T} \) are programmed in C#, on the platform NET Framework 4.6.1, and a computer based on Intel processor, with a clock speed of 1.8 GHz, the amount of RAM 8 GB, with numerical values of the source data: read speed in the \( s \)-th node \( \{VSP_s, VDB_s\} \in [60000;100000] \) KB/s, \( s=\overline{1,n} \); speed of writing to the memory of the \( s \)-th node \( \{DSP_s, DDB_s\} \in [1 \cdot 10^7;3 \cdot 10^7] \) KB/s, \( s=\overline{1,n} \); processor performance of the \( s \)-th node \( \{PSP_s, PDB_s\} \in [2.5 \cdot 10^9;3.5 \cdot 10^9] \) operations/s, \( s=\overline{1,n} \); the speed of data transmission over the communication channel \( \theta \in [1000;10000] \) KB/s; constant transmission delay on the channel.
\[ \theta_0 = 3 \cdot 10^{-4} \text{c}; \text{ constant processing delay in the node } \alpha_1 = \alpha_2 = 3 \cdot 10^{-6}; \text{ volume } j\text{-th relations } V_j \in [50000;150000]\text{KB}, \ j = 1,d; \text{ volume of information read } b_{lj} \in [1;1000]\text{KB} \left( l = 1,q, j = 1,d \right), \ j = 1,d \text{ on the } l\text{-th request for reading from the } j\text{-th relation, on } l\text{-th request for reading from the } j\text{-th relation; } \bar{b}_{lj} \in [1;500]\text{KB} \left( l = 1,q, j = 1,d \right) \text{ – the amount of information received after processing by the } l\text{-th request for reading from the } j\text{-th relationship.}

Using the developed software, a number of numerical experiments were carried out, the results of which are shown in figures 2-7 [12].

In figures 2-3 are graphs of the dependence of \( \bar{T} \) on the dimension of the solved problems of the optimal allocation of the distributed databases to the distributed nodes IPS: \( \bar{n}_i = (n \times q \times d), i = 1,6, \) where \( \bar{n}_1 = (5 \times 10 \times 10), \bar{n}_2 = (7 \times 30 \times 20), \bar{n}_3 = (9 \times 50 \times 30), \bar{n}_4 = (11 \times 70 \times 40), \bar{n}_5 = (13 \times 90 \times 50), \bar{n}_6 = (15 \times 100 \times 60). \)

**Figure 2.** Dependence \( \bar{T} \) on the size of the tasks to be solved, taking into account the locks on the basis of client-server: a - two-tier architecture; b - three-tier architecture.

**Figure 3.** Dependence \( \bar{T} \) on the size of the solved problems without taking into account the locks on the basis of client-server: a - two-tier architecture; b - three-tier architecture.

Figures 4-5 show graphs of the dependence of the value of \( \bar{T} \) on the reading speed at network nodes (KB/s), for different task sizes \( \bar{n}_i \) and for a fixed data rate on the channel \( \theta = 5000 \text{ (KB/s)} \) [13].
Figure 4. Dependence $\bar{T}$ on the read speed in the nodes, taking into account the locks on the basis of client-server: a - two-tier architecture; b - three-tier architecture.

Figure 5. Dependence $\bar{T}$ on the speed of reading in network nodes without taking into account locks on the basis of client-server: a - two-tier architecture; b - three-tier architecture.

Figures 6-7 show graphs of the dependence of the value $\bar{T}$ on the transmission rate of data over a channel $\theta$ for different dimensions of tasks $\bar{n}_i$.

Figure 6. Dependence of $\bar{T}$ on the data transfer rate on the channel $\theta$ taking into account locks on the basis of client-server: a - two-tier architecture; b - three-tier architecture.
Figure 7. Dependence of $T$ on the data rate on the channel $\theta$ without taking into account locks on the basis of client-server: a - two-tier architecture; b - three-tier architecture.

5. Conclusion

Obtained modeling results analysis showed that the presence in the model of blocking at the level of the whole database for small dimensions of allocating information resources over the distributed IPS nodes problem does not have a significant effect on the reactivity of the entire system as a whole; however, the influence of the blocking time on its reactivity becomes more significant and this value must be taken into account when constructing real models of computing systems [14].

The model of the optimal allocation of information resources by nodes of distributed IPS on the basis of the three-tier client-server architecture developed by the authors, taking into account the effect of locks, shows better results than the model based on a two-tier architecture and can be used to plan more rational organization of the computational process and to provide the required indicators of system reactivity.

References

[1] Scoba A N and Panfilov A N 2017 Izv. Universities. Electromechanics 2 77-84
[2] Boyko V V and Savinkov V M 1989 Designing Databases of Information Systems (Moscow: Finance and Statistics) 351
[3] Kuznetsov N A, Kulba V V and Kosyachenko S A 2002 Methods of Analysis and Synthesis of Modular Information-Control Systems (Moscow: FIZMATLIT) 880
[4] Ahmad I, Karlapalem K and Kwook Y K 2002 Distributed and Parallel Databases. 1 5-32
[5] Karpov A E 2007 Architecture of Distributed Software Systems (Moscow: MAKSS Press) 130
[6] Lizunkov V, Politsinskaya E, Malushko E, Kindaev A, Minin M 2018 Population of the world and regions as the principal energy consumer International journal of energy economics and policy 8 (3) 250-257
[7] Fedotov N G, Moiseev A V, Syemov A A, Lizunkov V G, Kindaev A Y 2017 New Methods of Three-Dimensional Images Recognition Based on Stochastic Geometry and Functional Analysis IOP Conference Series: Materials Science and Engineering 177 (1) 012047
[8] Skoba A N and Sostina E V 2015 Engineering Bulletin of the Don 2 [Electronic resource] Access mode: http://ivdon.ru/en/magazine/archive/n2y2015/2882 (access date: 17.09.2018)
[9] Chernomorov G A, Vorobyoov S P and Skoba A N 1997 Engineering, Economics, Culture: Yubil. Sat. Sci. Tr. Prof.-Teacher. Composition of the NSTU. (Rostov-on-Don: Hephaestus) 178-186
[10] Ergunova O T, Lizunkov V G, Malushko E Yu, Marchuk V I, Ignatenko A Yu 2017 Forming the system of strategic innovation management at the high-tech engineering enterprises IOP Conference Series: Materials Science and Engineering 177 (1) 012046
[11] Zhohikashvili V A and Vishnevsky V M 1988 Network Queuing. Theory and Application to Computer Networks (Moscow: Radio and Communication) 192
[12] Vishnevsky V M 2003 *Theoretical Bases of Designing Computer Networks* (Moscow: Technosphere) 512

[13] Ferreira R, Lizunkov V G, Politsinskaya E V 2017 Formation of entrepreneurial competencies of university graduates in conditions of transition to the universities of the third generation *Novosibirsk State Pedagogical University Bulletin* 7 (6) 195-211

[14] Skoba A N, Ayesh Ahmed Nafey Ayesh and Mikhailov V K 2018 *Engineering Bulletin of the Don* [Electronic resource] Access mode: http://ivdon.ru/en/magazine/archive/n1y2018/4774 (access date: 19.09.2018)