Rashid G. Alakbarov[1], Farkhad H. Pashayev[2], Mammad A. Hashimov [3]

[1, 3]Institute of Information Technology of ANAS, Baku, Azerbaijan
[2]Institute of Control Systems of ANAS, Baku, Azerbaijan
[1]rashid@iit.ab.az, [2]pasha.farhad@gmail.com, [3]mamedhashimov@gmail.com

A MODEL OF COMPUTATIONAL RESOURCES DISTRIBUTION AMONG DATA CENTER USERS

The paper defines thecondition of allocatingcomputing resources forData Center users or joining new users. In this case, the problems solved at Data Center considered equally complex, and a model of time change is developed, which is spent for the problem solution depending on the possible quantity of parallel processors to solve one problem. The relationship between the time spent on problem solution and the number of parallel processors is defined as a hyperbolic function. Consequently, the condition of allocating resource for a new user at Data Center is defined.

Keywords: Data Center, computing resource, parallel processors, hyperbolic model, condition of connection.

Introduction

Obviously, taking advantage of computing resources at the Data Centers does not only save the time, but also provides high quality problem solution. In this case, both the users, and the Data Center will get economic benefit. The benefits of users are about in timely and efficient execution of the work. Actually, this efficiency is not expressed inaccurate figures. However, experience shows that in many cases, the use of computational resources of data centers has no any alternatives to solve complex problems. Since the beginning of the use of data centers the optimization and efficiency became very vital. The perfect distribution of computing resources at data centers is based on optimization and efficiency [1, 2].

In general, practice shows that about 10 percent of the equipment resources is reserved. Reserve is primarily intended for a critical situation, i.e., it can be used at any backup failure. As a rule, these resources are not allocated for new users. Even if the new user needs a machine with higher parameters he is not allowed to be allocated reserve even temporarily.

The studies carried out by the monitoring programs used in the world practice show that only 70% of the resources allocated to users is utilized effectively. In order to achieve efficient use of remaining 30%, i.e.to distribute the computational and memory resources among users optimally provisioning technology is used [3, 4].

To ensure efficient use of resources unused thirty percent can be distributed to new users at non-peak times. In this case, there is no need to purchase new equipment if any new user applies. In addition, the system users, who need the machine of higher parameters, if necessary, can pay additional amount to achieve allocation of unused resources to solve his problem. Giving the users priorities in the scheduling and solving such issues is an important element.

The benefits of Data Centers are related to the non-still computational resources without becoming old physically and mentally. Therefore, one of the most pressing issues is to determine the condition of allocating computational resources to new users. In this case, it is important to define the time for the problem solution to satisfy the users by means of managing a number of processors connected in parallel.

Apparently, the fee paid for the solution of one problem to the Data Centre is measured by the number of processors connected in parallel and by the time spent for the problem solution. This paper considers the user applying to data centers for only one problem solution. Therefore, the problem number and the user number will match.

For the problem solution let us assume that the problems solved at the Data Center are of equal difficulty. In other words, the time spent on each issue and the number of processors connected in parallel is the same. If the condition is not covered, the problems can be split into several parts of equal (common) sizes that, each part will be imagined as a separate problem to be solved and to be of equal difficulty.

Let's denote p_i the number of processors connected in parallel to solve ith problem, the time spent to resolve the problem be T_i. Conditionally, we can imagine that the user pays money in amount of

$$v_i = p_i * T_i \qquad (1)$$

for the solution of the ith problem. Of course, the payment is directly proportional to v_i, and if we talk about real money, the sum of $p_i * T_i$ is multiplied by the coefficient appropriate to the currency. In our case, we don't need to find the coefficient or to convert it to the currency by multiplying it by the coefficient, and it does not change the essence of the problem. Therefore, we assume that the equality (1) is the amount of money paid by the user to solve ith problem. This user made the payment in the amount of

$\frac{v_i}{T_i} = \frac{p_i*T_i}{T_i} = p_i$ at the Data Centers within the common period of time.

This equation shows that, an income of Data Centers for each unit of current time is measured by the number of parallel connected processors. If i number of problem is solved at any time t at the Data Center, the income will be proportional to the quantity of $\sum_{i=1}^{i} p_i$.

Note that a number of technical limitations will occur in the process of problem solution. Let's mention two of them:
- The number of processors that can be allocated to solve each problem is limited. The number of processors used at the Data Center varies within 2,4,8,12,16,18,24. Let's denote this number by p_m;
- The total number of processors connected in parallel at the Data centers is limited. The number of processors used at the Data Center is 512. Let's denote this number by p_m;

Each user wants to resolve the problem quickly, and limits the maximum time. Let's suppose that the ith user wants to resolve the problem no later than the time T_{im}. Therefore, it is important to determine the conditions to attract new users.

The model of relation between the numbers of processors connected in parallel and the time to solve the problem.

Operational experience of the Data centers shows that there is a proper functional dependency between p_i and T_i. As an example of the process, let's use the program of finding the number of simple numerals. In the given problem, the number of simple numerals from 1 to 100,000,000 is found. As a result of experiments, the number of processors and the time to solve a problem varies as in the table 1.

The graphics in Figure 1 show that this dependence is taken in the form of graphs. Experience shows that the results are similar to other cases. The model of this dependence can be linear and hyperbolic functions, and the model with the smaller mean squared difference can be chosen [5, 6].

Table 1.

the number of processors	Time
2	70: 01: 30: 931
4	40: 59: 59: 770
8	28: 29: 24: 391
12	23: 35: 24: 241
16	16: 35: 52: 527
24	12: 56: 28: 743

a) Linear function form. In this case, let's consume the model as

$$T = kp + b \tag{2}$$

Let's find k and b which provides mean square difference

$$E = E_x = \frac{1}{I}\sum_{i=1}^{I}(kp_i + b - T_i)^2 \tag{3}$$

the minimum value by using the Table 1. to determine the coefficients k and b. These coefficients are found as the solution to the following system:

$$\begin{cases} \dfrac{\vartheta E}{\vartheta k} = 2k\sum_{i=1}^{I} p_i^2 + 2b\sum_{i=1}^{I} p_i - 2\sum_{i=1}^{I} p_i T_i = 0 \\ \dfrac{\partial E}{\partial b} = 2Ib + 2k\sum_{i=1}^{I} p_i - 2\sum_{i=1}^{I} T_i = 0 \end{cases}$$

Number of processors

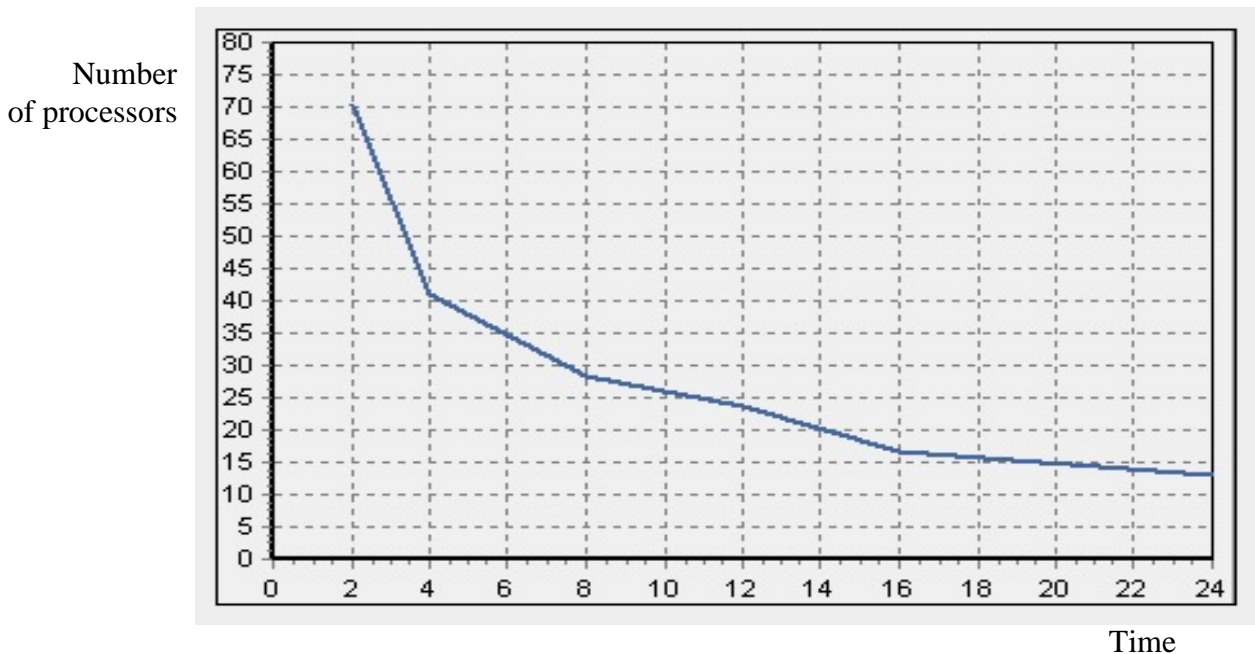

Figure 1.The graphic of dependence between the time spent for the problem solution and the number of parallel processors.

Since the table presents 6 elements we have $i = 6$. The system solution is as follows:

$$k = \frac{\sum_{i=1}^{I} p_i \sum_{i=1}^{I} T_i - I * \sum_{i=1}^{I} p_i T_i}{(\sum_{i=1}^{I} p_i)^2 - I * \sum_{i=1}^{I} (p_i)^2} \quad (4)$$

$$b = \frac{1}{I}(\sum_{i=1}^{I} T_i - k * \sum_{i=1}^{I} p_i) \quad (5)$$

By means of formulas (4), (5) and of table 1,the graph of the 2 linear functions together with the table can be set up as in Figure 2. In this case, the mean square difference is $E_x= 105.48$
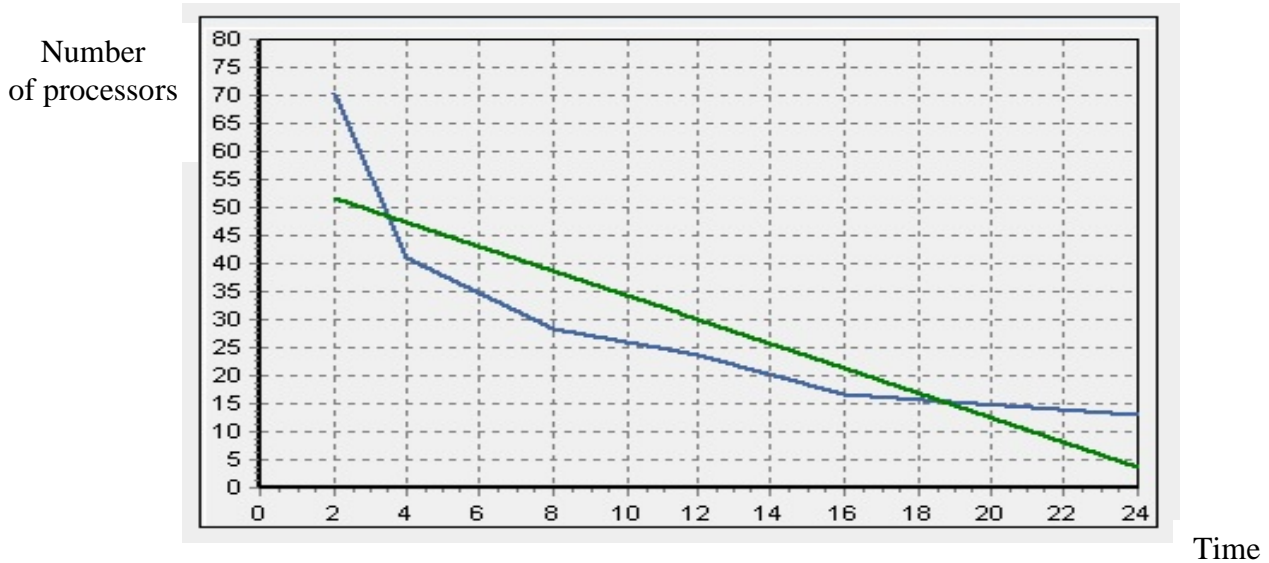
Number of processors

Time

Figure 2.Graphic description of the model for the linear function.

b) Hyperbolic function form. In this case, the model is

$$T = \frac{k}{p} \quad (6)$$

In order to determine the coefficient k, using the Table 1, let's find k, which provides minimum value mean square difference

$$E = E_h = \frac{1}{I}\sum_{i=1}^{I}(\frac{k}{p_i} - T_i)^2 \quad (7)$$

This coefficient is found by solving the following equation:

$$\frac{dE}{dk} = 2k\sum_{i=1}^{I}\frac{1}{p_i^2} - 2\sum_{i=1}^{I}\frac{T_i}{p_i} = 0$$

Thus,

$$k = \frac{\sum_{i=1}^{I}\frac{T_i}{p_i}}{\sum_{i=1}^{I}\frac{1}{p_i^2}} \quad (8).$$

Using the formulas (8) and Table 1, the graph of the hyperbolic functions (6) together with the table can be set up as in Figure 2. In this case, the mean square difference is $E_h=.57.42$ again.
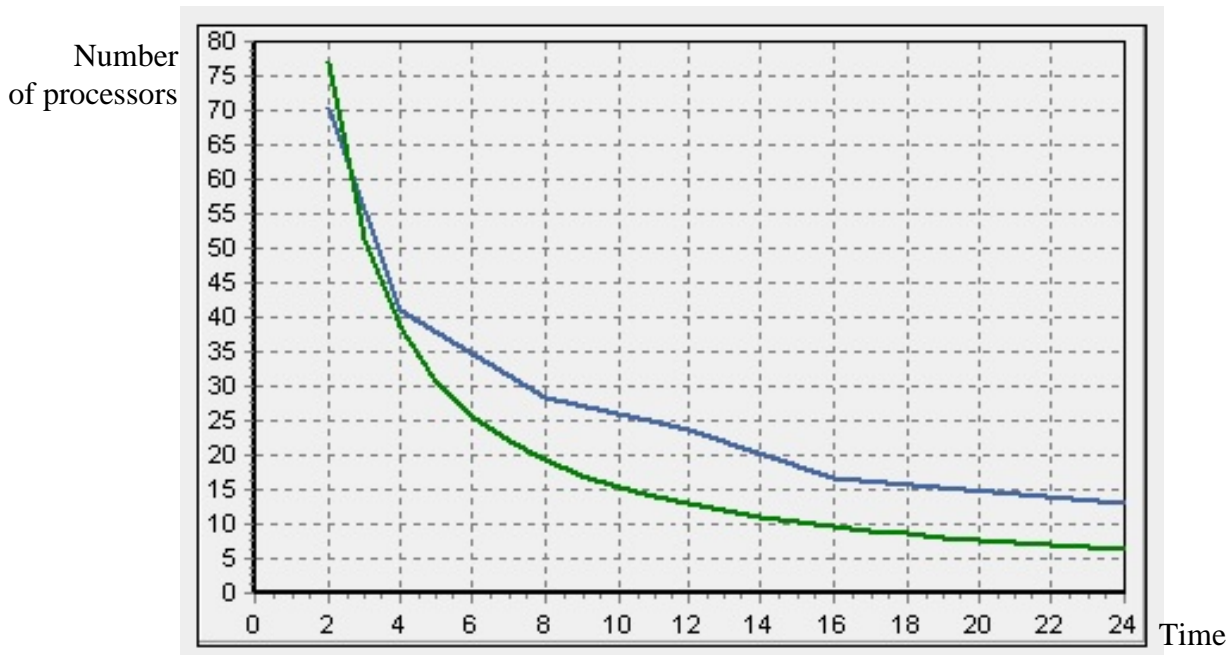

Figure 3.Graphical description of the model for the hyperbolic function.

As $E_h \ll E_x$, the hyperbolic function can be considered as a model for between the time for problem solution and the number of processors connected in parallel.

$$T = \frac{k}{p} \tag{9}$$

In practical applications, clarifying the Table 1, the coefficient k can be calculated by means of the formula (8) again.

We can come to the conclusion from the analysis of the model than, if necessary, applying the model as $p = \frac{k}{T}$, the number of processors connected in parallel can be found according to the time required to solve the problem.

Condition of distributing computational resources among new users.

It is necessary to attract new users to provide a few conditions:

- The availability of free computational resources. In other words, let the sum of the resources allocated to the i number of users be less than the number of processors connected in parallel at the Data Center:

$$\sum_{i=1}^{i} p_i < P_M$$

- Providing the requirement of the user to the time to solve the problem. Obviously, the user requires the time to solve the problem to be not more than T_m . To achieve this time,

according to the model (9), at least $p = \dfrac{k}{T_m}$ processors are required. Accordingly, the following condition must be fulfilled:

$$\frac{k}{T_m} < P_M - \sum_{i=1}^{I} p_i$$

- If we add here a limitation on the number of processors connected in parallel to solve the problem, then we can write the requirement of distributing computational resources to the new user in a form of system as follows:

$$\begin{cases} \dfrac{k}{T_m} < P_M - \sum_{i=1}^{I} p_i \\ \dfrac{k}{T_m} < p_m \end{cases}$$

Solving the system provides the requirement of distributing computational resources to the new user, and can be written as follows:

$$\frac{k}{T_m} < \min\{ p_m, P_M - \sum_{i=1}^{I} p_i \} \tag{10}.$$

Achieved results (10) can be called condition of distributing computational resources to the new user at Data Center or condition of a joining a new user the Data Center.

Conclusion

The paper studies the problem of distributing computational resources to a new user at Data Centers. The problem solution is searched within some technical problems encountered during the operation of Data Centers and under the terms of the restriction. The analytical model of dependence is developed on the basis of the experimental dependence between the number of processors connected in parallel and the time to solve a problem. Linear and hyperbolic functions are compared when developing the model. Preference is given to the model based on hyperbolic functions. Note that the adequacy of the model can be enhanced by additional experiments.

Finally, condition of distributing computational resources to a new user at Data Center was analyzed. Achieved results can be used to develop "cloud computing" services at Data Centers.

References

1. A.M.Alguliev, R.G.Alakbarov. Cloud Computing: Modern State, Problems and Prospects. Telecommunications and Radio Engineering. Volume 72, Number 3, USA. 2013, p 255-266.
2. V.V.Toporkov Models of distributed computing.-M.: FIZMATLIT, 2004.-320 p.
3. Mao, Ming; M. Humphrey (2013). "Budget constraints within scaling and Scheduling to Maximize Application Performance in the Cloud Workflows". Proceedings of the 2013 IEEE 27th International Symposium on Parallel and Distributed Processing (IPDPS2013): 67-78.
4. Jeremy LeBlanc; Adam Mendoza; Mike McNamara, NetAppAshishYajna; Rishi Manocha, Symantec. Maximize Storage Efficiency with NetApp Thin provisioning and Symantec Thin Reclamation. September 2010 | WP-7111.
5. James Stewart. Calculus early transcendentals Brooks / Cole, Cengage Learning, 2012, 1356 p.
6. HamdyA.Taha. Introduction to Operations Research. M :, "Williams", 2001, 912 p.