

UOT 004.75

*Алекперов Р.Г.<sup>1</sup>, Пашаев Ф.Г.<sup>2</sup>, Гашимов М.А.<sup>3</sup>*<sup>1,3</sup>Институт Информационных Технологий НАНА, Баку, Азербайджан<sup>2</sup>Институт Систем Управления НАНА, Баку, Азербайджан<sup>1</sup>rashid@iit.ab.az, <sup>2</sup>pasha.farhad@gmail.com, <sup>3</sup>mamedhashimov@gmail.com

### МОДЕЛЬ РАСПРЕДЕЛЕНИЯ МЕЖДУ ПОЛЬЗОВАТЕЛЯМИ ВЫЧИСЛИТЕЛЬНЫХ РЕСУРСОВ В ДАТА-ЦЕНТРЕ

*В статье определены условия подключения новых пользователей или распределения вычислительных ресурсов между пользователями в Дата-центре. При этом вопросы, решаемые в Дата-центре, были приняты равносильными, и построена модель изменения времени, потраченного на решение задачи в зависимости от количества параллельных процессоров, выделенных для решения одной задачи. Первоначальный вид модели определен визуально из графиков зависимости между временем, потраченным на решение задачи, и количеством соответствующих параллельных процессоров. После сравнения результатов отклонений выходных значений моделей и фактических данных связь между временем, потраченным на решение задачи, и количеством параллельных процессоров определена как гиперболическая функция. В итоге определено условие выделения вычислительной мощности для нового пользователя в Дата-центре.*

*Ключевые слова:* Дата-центр, вычислительная мощность, параллельные процессоры, гиперболическая модель, условие подключения.

#### Введение

Известно, что использование вычислительных ресурсов в дата-центрах наряду с экономией времени пользователей также обеспечивает качественное решение их задач. В этом случае и пользователи, и Дата-центр получают экономическую выгоду. Выгода пользователей заключается в своевременном и качественном выполнении работы. Эта выгода иногда не выражается точными числами. Но опыт показывает, что во многих случаях для решения некоторых задач использование вычислительных мощностей дата-центров считается неальтернативным путем. С момента использования дата-центров оптимизация и эффективность были выдвинуты на передний план. В основе оптимизации и эффективности лежит более совершенное распределение вычислительных ресурсов центра обработки данных [1, 2].

Известно, что примерно 10% общих ресурсов оборудования хранится в качестве резерва. Резерв в первую очередь предназначен для критической ситуации, следовательно, может быть использован при любых расстройствах. Как правило, новым пользователям не выделяется место в резерве. В случае, если одному из новых пользователей понадобится машина с более высокими параметрами, даже временно ему не выделяется место в резерве.

Из исследований, проведенных с помощью мониторинговых программ, используемых в мировой практике, установлено, что эффективно используются только 70% ресурсов, выделяемых для пользователей. С целью эффективного использования остальных 30% для оптимального распределения вычислительных ресурсов и ресурсов памяти между пользователями применяется технология резервирования [3, 4].

Для обеспечения эффективности использования ресурсов эти неиспользуемые 30% ресурсов могут быть выделены для новых пользователей в непиковые часы. В этом случае, если обратится новый пользователь, то не будет надобности приобрести новое оборудование. Кроме того, если одному из пользователей системы будет нужна машина с более высокими параметрами, тогда он может добиться решения проблемы

распределением тех ресурсов, которые не используются при оплате дополнительной суммы. Предоставление пользователям приоритетов в планировании и решении таких вопросов является важным фактором.

Пользы, полученные дата-центрами, связаны с тем, что вычислительные мощности не остаются в нерабочем состоянии, не подвергаются физическому и моральному старению без приноса прибыли. Поэтому одним из наиболее актуальных вопросов является определение условий распределения вычислительной мощности между новыми пользователями. Одним из основных вопросов, решение которых необходимо, является установление срока решения задач путем управления количеством процессоров, которые можно подключить параллельно в выгодном для пользователей порядке.

Известно, что денежные средства, выплачиваемые пользователем в Дата-центр для решения одной задачи, измеряются количеством параллельно подключаемых процессоров и временем, потраченным на решение задачи. В данной статье мы рассмотрим вопрос обращения одного пользователя в Дата-центр для решения только одной задачи. Поэтому серийные номера задачи и пользователя будут одинаковыми.

Для решения поставленной задачи мы предположим, что задачи, решаемые в Дата-центре, являются равносильными. Другими словами, время, потраченное на решение каждой задачи, и количество процессоров, параллельно подключенных в этом промежутке времени, являются равными. Если это условие не будет удовлетворено, то эти задачи можно разбить на части, каждая из которых будет равносильной и рассматриваться и решаться как отдельная задача.

Для решения  $i$ -ой задачи обозначим количество параллельно подключаемых процессоров через  $p_i$ , время, потраченное на решение задачи через  $T_i$ . Тогда условно можно полагать, что для решения  $i$ -ой задачи пользователь будет платить деньги в сумме

$$v_i = p_i * T_i . \quad (1)$$

Конечно, уплаченные деньги будут прямо пропорциональны  $v_i$ , и если речь пойдет о настоящих деньгах, тогда произведение  $p_i * T_i$  будет умножаться на соответствующий коэффициент денежной единицы. В нашем случае знание этого коэффициента, умножение на этот коэффициент и переход на истинные деньги не имеют значения и не меняет сущность вопроса. Поэтому мы принимаем, что равенство (1) является деньгами, оплаченными пользователем для решения  $i$ -ой задачи. Этот пользователь за единицу времени в дата-центры будет выплачивать деньги в сумме

$$\frac{v_i}{T_i} = \frac{p_i * T_i}{T_i} = p_i .$$

Это равенство показывает, что в течение каждой текущей единицы времени прибыль дата-центров измеряется количеством параллельно подключенных процессоров. Если в любой момент  $t$  в Дата-центре выполняется  $I$  количество задач, тогда прибыль будет пропорциональна количеству  $\sum_{i=1}^I p_i$ .

Надо отметить, что при организации вычислительного процесса для решения задачи выявляется ряд технических ограничений. Отметим два из этих ограничений:

- ограничено количество процессоров, которые могут быть выделены на решение каждой из задач. Количество используемых процессоров в Дата-центре варьируется в интервале 2, 4, 8, 12, 16, 18, 24. Это количество будет обозначаться через  $p_m$ ;

- ограничено общее количество процессоров, параллельно подключаемых в Дата-центрах. Количество процессоров, используемых нами в Дата-центре, равняется 512. Будем обозначать это число  $p_M$ .

Каждый пользователь в свою очередь желает незамедлительного решения своей задачи и ставит ограничение на время выполнения. Предположим, что  $i$ -ый пользователь хочет получить решение задачи не позднее  $T_{im}$  времени. Поэтому определение условий привлечения нового пользователя имеет большое значение.

### Синтез модели связи между количеством параллельно подключаемых процессоров и временем решения задачи

Опыт эксплуатации дата-центров показывает, что имеется определенная функциональная зависимость между  $p_i$  и  $T_i$ . В качестве примера мы будем пользоваться программой поиска количества простых чисел. В данной задаче проводится операция поиска количества простых чисел среди чисел от 1 до 100 миллионов. В результате экспериментов было установлено, что количество процессоров при решении одной задачи и время, потраченное на решение, меняются, как указано в таблице 1.

Если изобразить эту зависимость в виде графика, то получим график, указанный на рисунке 1. Эксперименты показывают, что в других случаях результаты тоже аналогичны. Построив математическую модель этой зависимости в виде линейной и гиперболической функций, мы можем получить модель, имеющую небольшую среднеквадратическую разницу с фактическими данными [5, 6].

Таблица 1

Зависимость потраченного времени от количества процессоров

Количество процессоров	Потраченное время
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

#### а) Случай линейной функции.

В этом случае возьмем модель в виде

$$T = kp + b \quad (2)$$

Для определения коэффициентов  $b$  будем использовать таблицу 1

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

тем самым найя  $k$  и  $b$ , которые дают минимальную оценку среднеквадратической разнице. Эти коэффициенты определяются решением ниже приведенной системы:

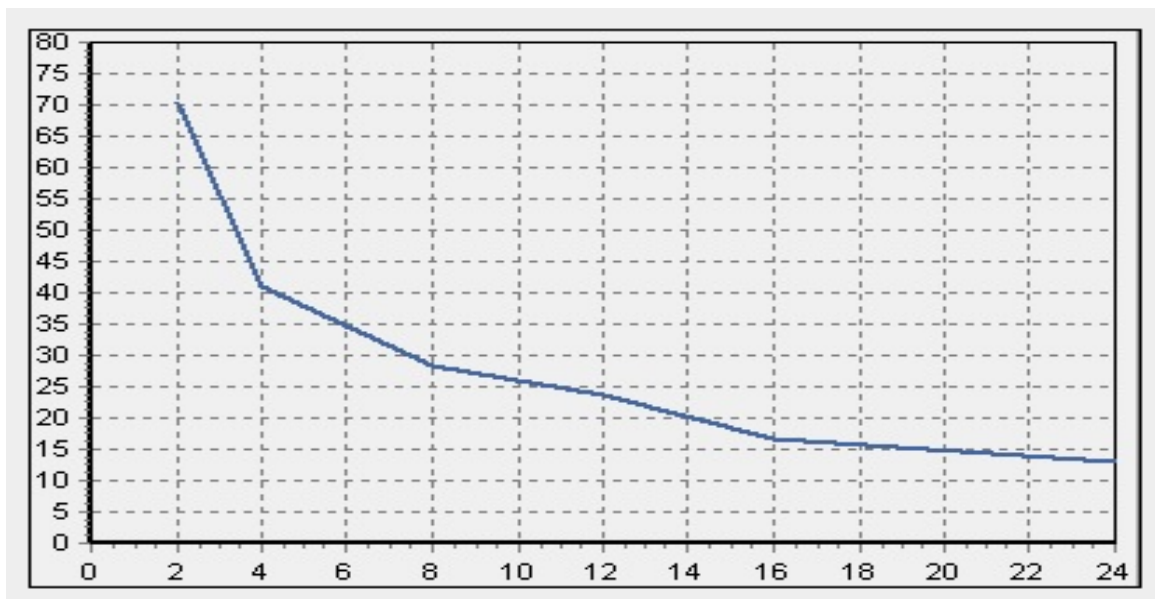


Рис.1. График зависимости между временем, потраченным на решение задачи, и параллельными процессорами

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

Вследствие того, что в таблице имеются 6 элементов, то в нашем случае  $I=6$ . Решение системы заключается в следующем:

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

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

С помощью формул (4), (5) и таблицы 1 график двух линейных функций вместе с таблицей может быть построен как на рисунке 2. Тогда повторно вычисленная среднеквадратическая разность будет  $E_x = 105.48$ .

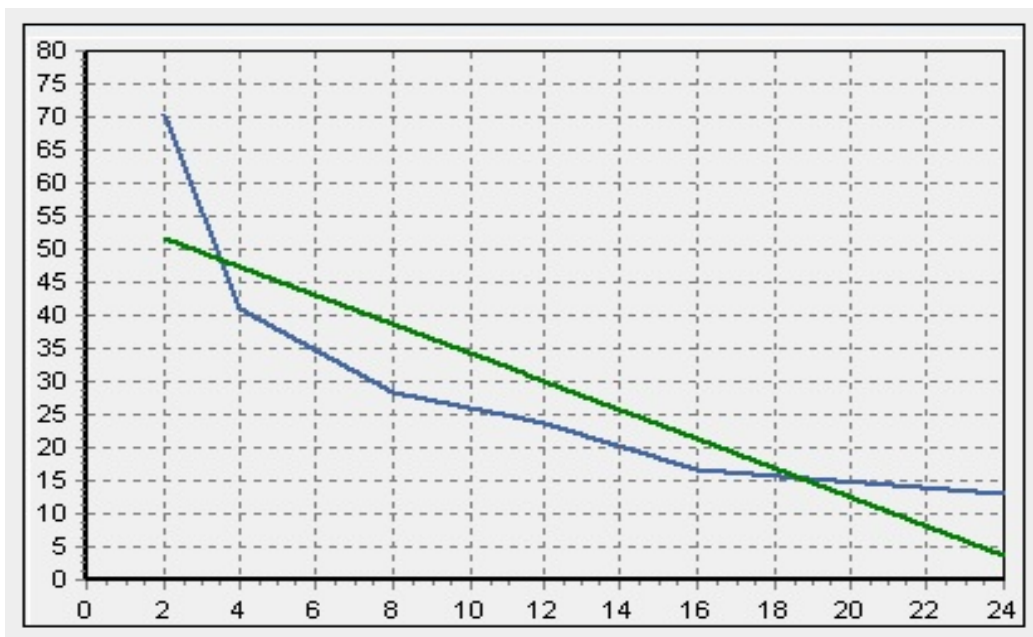


Рис.2. Графическое изображение модели для случая линейной функции

*b) Случай гиперболической функции.*

В этом случае возьмем модель в виде

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

опять, для определения коэффициента  $k$ , используя таблицу 1

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

найдем  $k$ , дающий минимальную оценку среднеквадратической разнице. Этот коэффициент можно найти путем решения ниже приведенного уравнения:

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

Следовательно,

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

Применив формулы (8) и таблицу 1, график гиперболической функции (6) вместе с таблицей можно построить как на рисунке 3.

Тогда повторно вычисленная среднеквадратическая разница будет  $E_h=57.42$ .

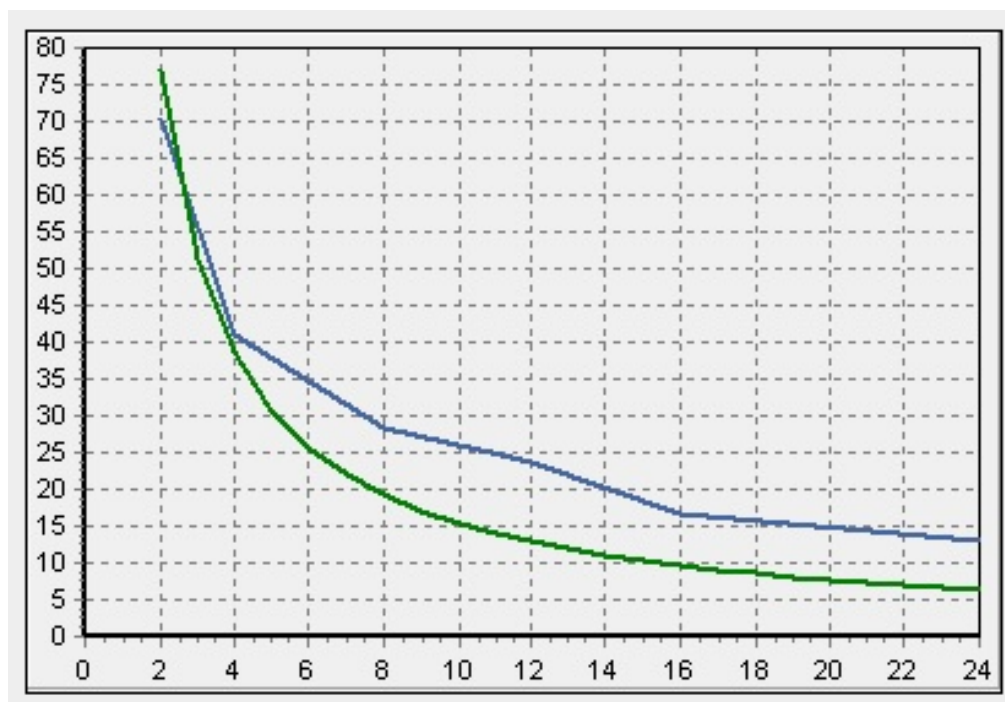


Рис.3. Графическое описание модели для случая гиперболической функции

Ввиду того, что  $E_n \ll E_x$ , гиперболическая функция может быть принята как модель зависимости времени решения задачи и количества параллельно подключаемых процессоров.

$$T = \frac{k}{p} . \quad (9)$$

При практическом применении путем уточнения таблицы 1 можно заново провести вычисление коэффициента  $k$  формулой (8).

Анализируя модель, можно прийти к выводу, что при необходимости, применив модель как  $p = \frac{k}{T}$ , можно определить количество процессоров, которые будут подключаться параллельно в зависимости от времени, требуемого для решения задачи.

#### Условия распределения вычислительных ресурсов между новыми пользователями

Для привлечения новых пользователей необходимо удовлетворение нескольких условий:

- наличие свободной вычислительной мощности. То есть сумма мощностей, выделенных для  $I$  количества пользователей в текущее время, должна быть меньше количества процессоров, которые могут параллельно подключаться к дата-центру.

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

- удовлетворение поставленных пользователем условий относительно времени решения задачи. Известно, что пользователь поставил требование, чтобы время решения задачи было больше, чем  $T_m$ . Для того чтобы ограничиться этим промежутком времени

по построенной нами модели (9), требуется, по меньшей мере,  $p = \frac{k}{T_m}$  количество процессоров. Следовательно, ниже приведенное условие должно быть удовлетворено:

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

▪ если сюда включить ограничение, поставленное на количество параллельно подключаемых процессоров для решения одной задачи, то условие выделения новому пользователю вычислительной мощности можно оформить в виде следующей системы:

$$\left\{ \begin{array}{l} \frac{k}{T_m} < P_M - \sum_{i=1}^I p_i \\ \frac{k}{T_m} < p_m \end{array} \right. .$$

Решение системы ставит условие выделения новому пользователю вычислительной мощности. Систему можно оформить следующим образом:

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

Полученное заключительное условие (10) можно назвать условием выделения вычислительной мощности новому пользователю в Дата-центре и условием подключения нового пользователя к дата-центру.

### Заключение

В статье рассмотрена проблема выделения новому пользователю вычислительной мощности в дата-центрах. Поиск решения данного вопроса был осуществлен в пределах некоторых технических проблем и условий ограничений, возникших в ходе эксплуатации дата-центров. Для решения одной задачи на основе экспериментальной зависимости между количеством параллельно подключаемых процессоров и временем решения задачи осуществлен синтез аналитической модели этой зависимости. Отдано предпочтение модели, построенной на основе гиперболической функции. Было отмечено, что в результате дополнительных экспериментов может быть повышена адекватность модели.

В заключение проанализированы условия выделения вычислительной мощности для нового пользователя в Дата-центре и определено итоговое условие.

Результаты, полученные в статье, можно использовать в создании услуг cloud computing в дата-центрах.

### Литература

1. Alguliev R.M., Alakbarov R.Q. Cloud Computing: Modern State, Problems and Prospects. Telecommunications and Radio Engineering, USA, 2013, vol.72, no.3, pp.255–266.
2. Топорков В.В. Модели распределенных вычислений, М.:ФИЗМАТЛИТ, 2004, 320 с.
3. Ming Mao, Humphrey M. Scaling and Scheduling to Maximize Application Performance within Budget Constraints in Cloud Workflows // Proceedings of the 2013 IEEE 27th International Symposium on Parallel and Distributed Processing (IPDPS 2013), 2013, pp.67–78.
4. LeBlanc Jeremy, Mendoza Adam, McNamara Mike, Yajnik NetApp Ashish; Manocha Rishi. Maximize Storage Efficiency with NetApp Thin Provisioning and Symantec Thin Reclamation. Symantec September 2010WP-7111.

5. Stewart James. Calculus early transcendentals Brooks / Cole, Cengage Learning 2012, 1356 p.
6. Хэмди А.Таха. Введение в исследование операций. М.: «Вильямс», 2001, 912 с.

#### UOT 004.75

**Ələkbərov Rəşid Q.<sup>1</sup>, Paşayev Fəhrad H.<sup>2</sup>, Həşimov Məmməd A.<sup>3</sup>**

<sup>1,3</sup>AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

<sup>2</sup>AMEA İdarəetmə Sistemləri İnstitutu, Bakı, Azərbaycan

<sup>1</sup>[rashid@iit.ab.az](mailto:rashid@iit.ab.az), <sup>2</sup>[pasha.farhad@gmail.com](mailto:pasha.farhad@gmail.com), <sup>3</sup>[m.hashimov@iit.ab.az](mailto:m.hashimov@iit.ab.az)

#### **Verilənlərin emalı mərkəzlərində istifadəçilər arasında hesablama resurslarının paylanması modeli**

Məqalədə Data Mərkəzdə istifadəçilərə hesablama resurslarının ayrılması və ya yeni istifadəçinin qoşulması şərti müəyyənləşdirilmişdir. Bu zaman Data Mərkəzdə həll edilən məsələlər eynigüclü qəbul edilmiş və bir məsələnin həlli üçün ayrılacaq paralel prosessorların sayından asılı olaraq məsələnin həllinə sərf edilən zamanın dəyişməsi modeli qurulmuşdur. Modellərin ilkin görünüşü məsələnin həllinə sərf edilmiş vaxt ilə müvafiq paralel prosessorların miqdarı arasında asılılıq qrafiklərindən vizual olaraq müəyyən edilmişdir. Modellərin və faktiki məlumatların çıxış qiymətlərinin çatışmazlıq nəticələrinin müqayisəsindən sonra məsələlərin həllinə sərf edilən zamanla paralel prosessorların sayı arasındakı əlaqə hiperbolik funksiya kimi təyin edilmişdir. Nəticədə Data Mərkəzdə yeni istifadəçiyə hesablama gücünün ayrılması şərti müəyyən edilmişdir.

*Açar sözlər:* Data Mərkəz, hesablama gücü, paralel prosessorlar, hiperbolik model, qoşulma şərti.

**Rashid Q. Alakbarov<sup>1</sup>, Farkhad H. Pashayev<sup>2</sup>, Mammad A. Hashimov<sup>3</sup>**

<sup>1,3</sup>Institute of Information Technology of ANAS, Baku, Azerbaijan

<sup>2</sup>Institute of Control Systems of ANAS, Baku, Azerbaijan

<sup>1</sup>[rashid@iit.ab.az](mailto:rashid@iit.ab.az), <sup>2</sup>[pasha.farhad@gmail.com](mailto:pasha.farhad@gmail.com), <sup>3</sup>[m.hashimov@iit.ab.az](mailto:m.hashimov@iit.ab.az)

#### **Model of Computational Resources Distribution Among Data Center Users**

The paper defines the condition of allocating computing resources for Data Center users or joining new users. In this case, the problems solved at Data Center considered equal, and a model of time change spent on the problem solution subjected to the possible quantity of parallel processors to solve one problem is developed. The initial view of the models is determined visually from the graphs of the relationship between the time spent on the problem solution and the amount of the respective parallel processors. After comparing the results of deviations of output values of models and actual data, 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.