

ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ҰЛТТЫҚ ҒЫЛЫМ АКАДЕМИЯСЫНЫҢ

Х А Б А Р Л А Р Ы

ИЗВЕСТИЯ

НАЦИОНАЛЬНОЙ АКАДЕМИИ НАУК
РЕСПУБЛИКИ КАЗАХСТАН

NEWS

OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN

**ФИЗИКА-МАТЕМАТИКА
СЕРИЯСЫ**



СЕРИЯ

ФИЗИКО-МАТЕМАТИЧЕСКАЯ



**PHYSICO-MATHEMATICAL
SERIES**

3 (313)

МАМЫР – МАУСЫМ 2017 Ж.

МАЙ – ИЮНЬ 2017 г.

MAY – JUNE 2017

1963 ЖЫЛДЫҢ ҚАҢТАР АЙЫНАН ШЫҒА БАСТАҒАН
ИЗДАЕТСЯ С ЯНВАРЯ 1963 ГОДА
PUBLISHED SINCE JANUARY 1963

ЖЫЛЫНА 6 РЕТ ШЫҒАДЫ
ВЫХОДИТ 6 РАЗ В ГОД
PUBLISHED 6 TIMES A YEAR

АЛМАТЫ, ҚР ҰҒА
АЛМАТЫ, НАН РК
ALMATY, NAS RK

Б а с р е д а к т о р ы
ф.-м.ғ.д., проф., ҚР ҰҒА академигі **Ғ.М. Мұтанов**

Р е д а к ц и я а л қ а с ы:

Жұмаділдаев А.С. проф., академик (Қазақстан)
Кальменов Т.Ш. проф., академик (Қазақстан)
Жантаев Ж.Ш. проф., корр.-мүшесі (Қазақстан)
Өмірбаев У.У. проф. корр.-мүшесі (Қазақстан)
Жүсіпов М.А. проф. (Қазақстан)
Жұмабаев Д.С. проф. (Қазақстан)
Асанова А.Т. проф. (Қазақстан)
Бошқаев К.А. PhD докторы (Қазақстан)
Сұраған Д. PhD докторы (Қазақстан)
Quevedo Hernando проф. (Мексика),
Джунушалиев В.Д. проф. (Қырғыстан)
Вишневский И.Н. проф., академик (Украина)
Ковалев А.М. проф., академик (Украина)
Михалевич А.А. проф., академик (Белорус)
Пашаев А. проф., академик (Әзірбайжан)
Такибаев Н.Ж. проф., академик (Қазақстан), бас ред. орынбасары
Тигиняну И. проф., академик (Молдова)

«ҚР ҰҒА Хабарлары. Физика-математикалық сериясы».

ISSN 2518-1726 (Online), ISSN 1991-346X (Print)

Меншіктенуші: «Қазақстан Республикасының Ұлттық ғылым академиясы» РҚБ (Алматы қ.)
Қазақстан республикасының Мәдениет пен ақпарат министрлігінің Ақпарат және мұрағат комитетінде
01.06.2006 ж. берілген №5543-Ж мерзімдік басылым тіркеуіне қойылу туралы куәлік

Мерзімділігі: жылына 6 рет.
Тиражы: 300 дана.

Редакцияның мекенжайы: 050010, Алматы қ., Шевченко көш., 28, 219 бөл., 220, тел.: 272-13-19, 272-13-18,
www.nauka-nanrk.kz / physics-mathematics.kz

© Қазақстан Республикасының Ұлттық ғылым академиясы, 2017

Типографияның мекенжайы: «Аруна» ЖК, Алматы қ., Муратбаева көш., 75.

Главный редактор
д.ф.-м.н., проф. академик НАН РК **Г.М. Мутанов**

Редакционная коллегия:

Джумадильдаев А.С. проф., академик (Казахстан)
Кальменов Т.Ш. проф., академик (Казахстан)
Жантаев Ж.Ш. проф., чл.-корр. (Казахстан)
Умирбаев У.У. проф. чл.-корр. (Казахстан)
Жусупов М.А. проф. (Казахстан)
Джумабаев Д.С. проф. (Казахстан)
Асанова А.Т. проф. (Казахстан)
Бошкаев К.А. доктор PhD (Казахстан)
Сураган Д. доктор PhD (Казахстан)
Quevedo Hernando проф. (Мексика),
Джунушалиев В.Д. проф. (Кыргызстан)
Вишневский И.Н. проф., академик (Украина)
Ковалев А.М. проф., академик (Украина)
Михалевич А.А. проф., академик (Беларусь)
Пашаев А. проф., академик (Азербайджан)
Такибаев Н.Ж. проф., академик (Казахстан), зам. гл. ред.
Тигиняну И. проф., академик (Молдова)

«Известия НАН РК. Серия физико-математическая».

ISSN 2518-1726 (Online), ISSN 1991-346X (Print)

Собственник: РОО «Национальная академия наук Республики Казахстан» (г. Алматы)

Свидетельство о постановке на учет периодического печатного издания в Комитете информации и архивов
Министерства культуры и информации Республики Казахстан №5543-Ж, выданное 01.06.2006 г.

Периодичность: 6 раз в год.

Тираж: 300 экземпляров.

Адрес редакции: 050010, г. Алматы, ул. Шевченко, 28, ком. 219, 220, тел.: 272-13-19, 272-13-18,
www.nauka-nanrk.kz / physics-mathematics.kz

© Национальная академия наук Республики Казахстан, 2017

Адрес типографии: ИП «Аруна», г. Алматы, ул. Муратбаева, 75.

E d i t o r i n c h i e f
doctor of physics and mathematics, professor, academician of NAS RK **G.M. Mutanov**

E d i t o r i a l b o a r d:

Dzhumadildayev A.S. prof., academician (Kazakhstan)
Kalmenov T.Sh. prof., academician (Kazakhstan)
Zhantayev Zh.Sh. prof., corr. member. (Kazakhstan)
Umirbayev U.U. prof. corr. member. (Kazakhstan)
Zhusupov M.A. prof. (Kazakhstan)
Dzhumabayev D.S. prof. (Kazakhstan)
Asanova A.T. prof. (Kazakhstan)
Boshkayev K.A. PhD (Kazakhstan)
Suragan D. PhD (Kazakhstan)
Quevedo Hernando prof. (Mexico),
Dzhunushaliyev V.D. prof. (Kyrgyzstan)
Vishnevskiy I.N. prof., academician (Ukraine)
Kovalev A.M. prof., academician (Ukraine)
Mikhalevich A.A. prof., academician (Belarus)
Pashayev A. prof., academician (Azerbaijan)
Takibayev N.Zh. prof., academician (Kazakhstan), deputy editor in chief.
Tiginyanu I. prof., academician (Moldova)

News of the National Academy of Sciences of the Republic of Kazakhstan. Physical-mathematical series.

ISSN 2518-1726 (Online), ISSN 1991-346X (Print)

Owner: RPA "National Academy of Sciences of the Republic of Kazakhstan" (Almaty)

The certificate of registration of a periodic printed publication in the Committee of information and archives of the Ministry of culture and information of the Republic of Kazakhstan N 5543-Ж, issued 01.06.2006

Periodicity: 6 times a year

Circulation: 300 copies

Editorial address: 28, Shevchenko str., of. 219, 220, Almaty, 050010, tel. 272-13-19, 272-13-18,
[www.nauka-nanrk.kz / physics-mathematics.kz](http://www.nauka-nanrk.kz/physics-mathematics.kz)

© National Academy of Sciences of the Republic of Kazakhstan, 2017

Address of printing house: ST "Aruna", 75, Muratbayev str, Almaty

NEWS

OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF KAZAKHSTAN

PHYSICO-MATHEMATICAL SERIES

ISSN 1991-346X

Volume 2, Number 312 (2017), 167 – 172

UDC 004.421.2.519.178

A.S. Shomanov¹, D.Zh. Akhmed-Zaki¹, E.N. Amirgaliyev¹, M.E. Mansurova¹¹al-Farabi Kazakh National University, Almaty, Kazakhstan.
e-mail: adai.shomanov@gmail.com**About the problem of key distribution in Mapreduce model**

Abstract. An important result in the field of processing of large amounts of data is the invention of the Mapreduce model. This model is based on the separation of data processing between parallel running processes. Data processing is performed using two types of functions: map and reduce. Map defines the transformation of input text by parallel processes on the basis of which a set of key / value pairs is generated. Reduce is operate on specific key and a list of all values associated with this key by performing a set of operations on this list of values, result of which is a pair with a key value and a certain aggregate value obtained as a result of these operations. The working environment of Mapreduce is a cluster consisting of a set of compute nodes. The nodes in the cluster must be connected by the communication network. Each node is scheduled to launch map and reduce processes.

Thus, an important task is to plan the distribution of keys among reduce processes in order to minimize data transfer operations over the network. Importance is justified by the fact that data transfer operations over the network greatly increase the processing time in case of incorrectly planned key distribution process. This task is NP-complete. The NP-completeness of the problem limits its exact solution, even for small parameters of the input data. For its effective solution in the current article, an approach based on the genetic algorithm is proposed. This approach can not guarantee an exact solution, but gives rather good approximated results. Another advantage of this approach is the ability to solve problems with large parameters of input data.

Keywords: Mapreduce, optimization, parallel processing, Generalized assignment problem.

УДК 004.421.2.519.178

А.С. Шоманов¹, Д.Ж. Ахмед-Заки¹, Е.Н. Амиргалиев¹, М.Е. Мансурова¹¹Казахский Национальный Университет им. аль-Фараби**О ЗАДАЧЕ ОПТИМИЗАЦИИ РАСПРЕДЕЛЕНИЯ
КЛЮЧЕЙ В MAPREDUCE МОДЕЛИ**

Аннотация. Важным результатом в области обработки больших объемов данных является изобретение Mapreduce модели. В основе данной модели лежит разделение обработки данных между параллельными процессами. Обработка данных производится с использованием двух типов функций: map и reduce. Map задает преобразование входного текста параллельными процессами на основе которого генерируется набор пар ключ/значение. Reduce на основе определенного ключа и списка всех значений, связанных с этим ключом, производит набор операций над списком значений, результатом которого является пара со значением ключа и определенного агрегированного значения, полученного как результат этих операций. Средой работы Mapreduce является кластер, состоящий из набора вычислительных узлов. Узлы в кластере должны быть соединены коммуникационной сетью. На каждом узле производится запуск процессов map и reduce.

Таким образом, важной задачей является планирование распределения ключей по процессам reduce с целью минимизации операций передачи данных по сети. Важность обоснована тем, что

операции передачи данных по сети в значительной степени увеличивают время обработки данных в случае неправильно спланированного процесса распределения ключей. Данная задача является NP-полной. NP-полнота задачи ограничивает ее точное решение даже для малых параметров входных данных. Для ее эффективного решения в текущей статье предлагается подход на основе генетического алгоритма. Данный подход не может гарантировать точного решения, но дает достаточно хорошие аппроксимированные результаты. Другим преимуществом данного подхода является возможность решения задач с большими параметрами входных данных.

Ключевые слова: Mapreduce, оптимизация, параллельная обработка данных, Обобщенная задача о назначениях.

В основе технологии Mapreduce [1-5] лежит идея параллельной обработки данных с применением двух основных функций: map и reduce. Функция map позволяет на основе входного набора данных произвести преобразование, в котором формируются в результате пары ключ/значение. Далее происходит группировка полученных пар ключ/значение, так что каждая преобразование reduce оперирует только на множестве значений, имеющих один и тот же ключ. Для параллелизации процесса обработки на основе Mapreduce, данные разбиваются между параллельно работающими map процессами на этапе map, а затем на этапе reduce создаются параллельные процессы для выполнения функции reduce по каждому ключу. Таким образом, при обработке больших массивов данных достигается высокий уровень масштабирования. Основной цикл работы Mapreduce состоит из 4 этапов :

- **Init.** Задается описание функций map и reduce, входные и выходные директории и другие параметры.
- **Map.** Каждый процесс map сканирует данные, переданные ему в качестве входного параметра. В ходе обработки данных функцией map формируется список пар ключ/значение согласно функции map, заданной пользователем.
- **Shuffle.** Происходит распределение пар ключ/значение по reduce процессам таким образом, что каждый reduce процесс обрабатывает только один, предназначенный для него, уникальный ключ.
- **Reduce.** Каждый reduce процесс выполняет операции на наборе пар ключ/значение согласно функции reduce, заданной пользователем.

Одной из основных проблем, связанных с обработкой больших объемов данных на основе технологии Mapreduce является оптимизация задачи по назначению ключей различным параллельным reduce процессам. Задача оптимизации включает в себя балансировку нагрузки между этими процессами, сокращения объема передаваемых данных, а также уменьшения количества операций передачи и обмена данными по сетевым каналам.

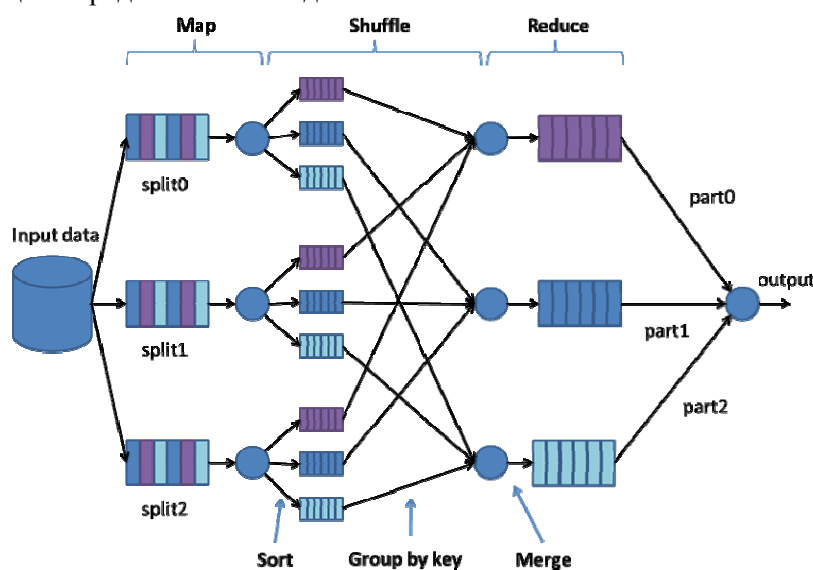


Рисунок 1 - Принцип работы Mapreduce

Из рисунка 1 видно, что данные с одинаковыми ключами на этапе распределения ключей (Shuffle) передаются одному процессу, ответственному за обработку определенного ключа. Так как объем данных может быть огромным, то соответственно, важно правильно определить, какому процессу нужно назначить определенный ключ для минимизации операций передачи данных по сети.

Для задачи распределения вычислительной нагрузки между параллельно работающими процессами и оптимального балансирования объема данных нами был применен эвристический подход на основе генетического алгоритма. Генетические алгоритмы [6-9] используются для решения многих задач в области комбинаторной и многокритериальной оптимизации. Основную часть задач в данных областях составляют NP-полные задачи, для которых не существует полиномиального алгоритма для их решения. Таким образом, многие важные задачи из данных областей остаются без эффективных подходов, которые способны были бы за разумное время решить данную задачу даже для самых малых параметров входных данных. Таким образом, единственными методами способными дать более или менее подходящее решение для больших размерностей данных задач являются различные классы эвристических алгоритмов.

Формулировка задачи оптимизации состоит из следующего множества условий:

$$\min \sum_{i=0}^{threads-1} \sum_{j=1}^{keys} x_{ij} \times cost_{ij} \quad (1)$$

$$x_{ij} \in \{0,1\} \quad (1)$$

$$\min \left(\max_{i,j=0..threads-1} |load_i - load_j| \right) \quad (2)$$

$$load_i = \sum_{t=0}^{threads-1} \sum_{j=1}^{keys} x_{ij} \times size_{tj} \quad (3)$$

Задача оптимизации, приведенная выше, является модификацией “Обобщенной задачи о назначениях” [10-13]. Данная задача относится к классу NP-полных задач. Различные генетические алгоритмы для решения данной задачи были описаны ранее [14,15].

Для того чтобы найти стоимость назначения ключа j потоку i нам требуется создать матрицу, в которой каждому элементу данной матрицы $cost_{ij}$ присваивается значение стоимости назначения ключа j потоку i . Данная стоимость задается на основе количества элементов определенного ключа, которые требуется передать определенному потоку. В формуле (3) дается определение функционала балансировки нагрузки. Функционал балансировки нагрузки рассчитывается как минимальное значение максимальной разницы общей нагрузки $load$ между любыми парами различных потоков. Значение общей нагрузки $load$ для каждого потока рассчитывается согласно формуле (4). Формула (2) описывает область определения для переменной x_{ij} . Значение x_{ij} задается равным 1, если поток i был назначен для обработки ключа под номером j , в противном случае, данное значение равно 0 и, соответственно, поток i не был назначен для обработки ключа j . Для модифицированной “Обобщенной задачи о назначениях” применимо к процедуре распределения ключей в Mapreduce модели, требуется оптимизировать значение обоих функционалов для получения приемлемого решения.

Для того чтобы применить к вышеописанной задаче подход с применением генетического алгоритма нужно определить как можно представить задачу на языке генетического алгоритма. Решение (особь) для задачи можно представить в виде вектора значений, где i -ому элементу данного вектора присваивается номер потока назначенного для обработки i -ого ключа. Популяция

представляет собой множество особей, выбранных согласно методам отбора генетического алгоритма. Размер популяции задается, как отдельный параметр алгоритма и может быть изменен в зависимости от специфических свойств решаемой задачи. Хромосома является упорядоченной последовательностью генов. Ген представляет собой атомарный элемент хромосомы. Начальная популяция выбирается случайной генерацией значений генов для хромосомы каждой особи. Функция приспособленности (фитнес-функция) определяет меру приспособленности определенной особи. Для нашей задачи оптимизации фитнес-функция рассчитывается как взвешенная сумма значений функционалов (1) и (3) для каждой особи в популяции. Задача генетического алгоритма, таким образом, выражается в нахождении особи с лучшим значением фитнес-функции. Генетический алгоритм для задачи распределения ключей в Mapreduce модели работает согласно следующей процедуре:

```
Load balancing procedure
Initialize algorithm
Generate initial random population  $p$  of chromosomes
While  $i \leq \text{MAX\_ITERATIONS}$ 
Update fitness values of each element of  $p$ 
For  $i=1..P\_NUM$ 
Choose two parent elements  $p_1$  and  $p_2$  from current population  $p$  by
applying tournament selection
Perform crossover on  $p_1$  and  $p_2$  to generate child chromosome  $c$ 
Perform mutation on child chromosome  $c$ 
Add child chromosome  $c$  to the list of new population elements  $np$ 
End Of For
Assign current population  $p$  to newly obtained population list  $np$ 
End of While
Choose list member with best fitness value and assign it to array
sol
```

Рисунок 2 - Генетический алгоритм распределения ключей в Mapreduce модели

Основной цикл алгоритма состоит из P_NUM шагов. На каждой итерации алгоритма сначала обновляются значения фитнес-функции для каждой особи в текущей популяции. Затем, выбираются две особи популяции p_1 и p_2 на основе турнирного отбора, цель которого заключается в том, чтобы выбрать из случайно выбранного множества особей из текущей популяции наиболее приспособленного (с наилучшим значением фитнес-функции). После этого, производится операция кроссинговер (crossover), и из двух особей, посредством данной операции, генерируется новая особь потомок, которая наследует свойства обоих родительских особей. После этого новая особь потомок добавляется в новую популяцию. Затем, производится операция мутации над новой особью потомком, после которой основной цикл алгоритма завершается и происходит переход на новую итерацию.

ЛИТЕРАТУРЫ

[1] Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107-113. doi:10.1145/1327452.1327492

- [2] Dean, J., & Ghemawat, S. (2010). Map reduce: A flexible data processing tool. *Communications of the ACM*, 53(1), 72-77. doi:10.1145/1629175.1629198
- [3] Lee, K. -, Lee, Y. -, Choi, H., Chung, Y. D., & Moon, B. (2011). Parallel data processing with MapReduce: A survey. *SIGMOD Record*, 40(4), 11-20. doi:10.1145/2094114.2094118
- [4] Jiang, D., Ooi, B. C., Shi, L., & Wu, S. (2010). The performance of mapreduce: An indepth study. *Proceedings of the VLDB Endowment*, 3(1), 472-483. doi: 10.14778/1920841.1920903
- [5] Dean, J. (2006). Experiences with MapReduce, an abstraction for large-scale computation. Paper presented at the *Parallel Architectures and Compilation Techniques - Conference Proceedings, PACT*, , 2006 1. doi:10.1145/1152154.1152155
- [6] Srinivas, M., & Patnaik, L. M. (1994). Genetic algorithms: A survey. *Computer*, 27(6), 17-26. doi:10.1109/2.294849
- [7] Konak, A., Coit, D. W., & Smith, A. E. (2006). Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering and System Safety*, 91(9), 992-1007. doi:10.1016/j.res.2005.11.018
- [8] Storn, R., & Price, K. (1997). Differential evolution - A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4), 341-359. doi:10.1023/A:1008202821328
- [9] Marler, R. T., & Arora, J. S. (2004). Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, 26(6), 369-395. doi:10.1007/s00158-003-0368-6
- [10] Gavish, B., & Pirkul, H. (1991). Algorithms for the Multi-Resource Generalized Assignment Problem. *Management Science*, 37(6), 695-713. doi: 10.1287/mnsc.37.6.695
- [11] Shmoys, D. B., & Tardos, E. (1993). Approximation algorithm for the generalized assignment problem. *Mathematical Programming, Series B*, 62(3-8), 461-474. doi:10.1007/BF01585178
- [12] Ross, G. T., & Soland, R. M. (1975). A branch and bound algorithm for the generalized assignment problem. *Mathematical Programming*, 8(1), 91-103. doi:10.1007/BF01580430
- [13] Cattrysse, D. G., & Van Wassenhove, L. N. (1992). A survey of algorithms for the generalized assignment problem. *European Journal of Operational Research*, 60(3), 260-272. doi:10.1016/0377-2217(92)90077-M
- [14] Chu, P.C. & Beasley, J.E. 1997, "A genetic algorithm for the generalised assignment problem", *Computers and Operations Research*, vol. 24, no. 1, pp. 17-23. doi: 10.1016/S0305-0548(96)00032-9
- [15] Liu, Y. Y., & Wang, S. (2015). A scalable parallel genetic algorithm for the generalized assignment problem. *Parallel Computing*, 46, 98-119. doi:10.1016/j.parco.2014.04.008

REFERENCES

- [1] Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107-113. doi:10.1145/1327452.1327492
- [2] Dean, J., & Ghemawat, S. (2010). Map reduce: A flexible data processing tool. *Communications of the ACM*, 53(1), 72-77. doi:10.1145/1629175.1629198
- [3] Lee, K. -, Lee, Y. -, Choi, H., Chung, Y. D., & Moon, B. (2011). Parallel data processing with MapReduce: A survey. *SIGMOD Record*, 40(4), 11-20. doi:10.1145/2094114.2094118
- [4] Jiang, D., Ooi, B. C., Shi, L., & Wu, S. (2010). The performance of mapreduce: An indepth study. *Proceedings of the VLDB Endowment*, 3(1), 472-483. doi: 10.14778/1920841.1920903
- [5] Dean, J. (2006). Experiences with MapReduce, an abstraction for large-scale computation. Paper presented at the *Parallel Architectures and Compilation Techniques - Conference Proceedings, PACT*, , 2006 1. doi:10.1145/1152154.1152155
- [6] Srinivas, M., & Patnaik, L. M. (1994). Genetic algorithms: A survey. *Computer*, 27(6), 17-26. doi:10.1109/2.294849
- [7] Konak, A., Coit, D. W., & Smith, A. E. (2006). Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering and System Safety*, 91(9), 992-1007. doi:10.1016/j.res.2005.11.018
- [8] Storn, R., & Price, K. (1997). Differential evolution - A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4), 341-359. doi:10.1023/A:1008202821328
- [9] Marler, R. T., & Arora, J. S. (2004). Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, 26(6), 369-395. doi:10.1007/s00158-003-0368-6
- [10] Gavish, B., & Pirkul, H. (1991). Algorithms for the Multi-Resource Generalized Assignment Problem. *Management Science*, 37(6), 695-713. doi: 10.1287/mnsc.37.6.695
- [11] Shmoys, D. B., & Tardos, E. (1993). Approximation algorithm for the generalized assignment problem. *Mathematical Programming, Series B*, 62(3-8), 461-474. doi:10.1007/BF01585178
- [12] Ross, G. T., & Soland, R. M. (1975). A branch and bound algorithm for the generalized assignment problem. *Mathematical Programming*, 8(1), 91-103. doi:10.1007/BF01580430
- [13] Cattrysse, D. G., & Van Wassenhove, L. N. (1992). A survey of algorithms for the generalized assignment problem. *European Journal of Operational Research*, 60(3), 260-272. doi:10.1016/0377-2217(92)90077-M

[14] Chu, P.C. & Beasley, J.E. **1997**, "A genetic algorithm for the generalised assignment problem", *Computers and Operations Research*, vol. 24, no. 1, pp. 17-23. doi: 10.1016/S0305-0548(96)00032-9

[15] Liu, Y. Y., & Wang, S. (**2015**). A scalable parallel genetic algorithm for the generalized assignment problem. *Parallel Computing*, 46, 98-119. doi:10.1016/j.parco.2014.04.008

А.С. Шоманов¹, Д.Ж. Ахмед-Заки¹, Е.Н. Амиргалиев¹, М.Е. Мансурова¹

¹әл-Фараби атындағы Қазақ Ұлттық Университеті, Алматы, Қазақстан Республикасы

КІЛТТЕРДІ MAPREDUCE ҮЛГІСІНДЕ ТАРАТУ ЕСЕБІ ТУРАЛЫ

Үлкен деректерді өңдеу саласында Mapreduce үлгіні ойлап шығаруы маңызды нәтиже деп есептелінеді. Осы үлгінің негізінде деректерді өңдеу тапсырмасын параллельді процесстер арасында бөлуі жатыр. Деректерді өңдеу тапсырмасы екі функция арқылы орындалады: map және reduce. Map кіріс мәтінді параллельді процесстермен түрлендіру тапсырмасын орындайды, нәтижесінде кілт/мағынасы жиынтығы құрастырылады. Белгілі кілт және осы кілтпен байланысты мағыналар жиынтығы негізінде reduce мағыналар жиынтығымен әртүрлі операцияларды орындайды, нәтижесі ретінде кілт және операциялардан алынған жинақ мағынасы боп табылады. Mapreduce жұмыс істеу ортасы бірнеше есептеу орталықтардан құрылған кластерден тұрады. Кластердің орталықтары коммуникациялық желілермен байланыстырылған болу керек. Әр есептеу орталығында map және reduce процесстердің іске қосылуы орындалады.

Сонымен, желілер бойынша деректерді жіберу азайту мақсатымен кілттерді reduce процесстер ішінде тарату есебі маңызды боп есептелінеді. Кілттерді тарату тапсырма дурыс емес жоспарылған кезде деректерді желілер арқылы жіберу мағыналы дәрежесінде деректерді өңдеу уақытын көбейтеді. Осы есеп NP-толық есептеріне жатады. NP-толықтығы түгіл ең кіші деректердің кіріс параметрлердің жағдайында осы есептің нақты шешімін табуын шектейді. Осы есепті тиімді түрде шешу үшін осы мақалада генетикалық алгоритм негізінде тәсілдеме ұсынылады. Осы тәсілдеме нақты шешімін жеткізе алмай мүмкін, бірақ өте жақсы жуықтама нәтижесін көрсете алады. Осы тәсілдеменің үлкен параметрлермен кіріс деректерден құрылған есептерді шешуі басқа артықшылық боп есептелінеді.

Тірек сөздер: Mapreduce, үйлесімділеу, деректерді параллельді өңдеу, Жалпылама тағайындау есебі.

МАЗМУНЫ

<i>Бердібай С.Б., Парецкая Н.А., Сабитов А.Н., Исламов Р.А., Тамазян Р.А., Токмолдин С.Ж., Ильин А.И., Мартиросян К.С.</i> Иод және оның құрылымымен фенилалалиннің кешенді комплексі	5
<i>Кабышев А.М., Кутербеков К.А., Пенионжкевич Ю.Э., Маслов В.А., Мендибаев К., Соболев Ю.Г., Лукьянов С.М., Кабдрахимова Г.Д., Азнабаев Д., Курманжанов А.Т.</i> Өлшеу кезіндегі модификацияланған трансмиссионды әдіс негізінде – реакциялардың толық өлшемдерінің кателіктерін және ұшып келуші бөлшектердің энергиясы анықтау.....	10
<i>Бердібай С.Б., Парецкая Н.А., Сабитов А.Н., Исламов Р.А., Тамазян Р.А., Токмолдин С.Ж., Ильин А.И., Мартиросян К.С.</i> Иод және оның құрылымымен фенилалалиннің кешені комплекс фенилаланина с иодом и его структура.....	19
<i>Жұмағұлова Қ.Н., Рамазанов Т.С., Машеева Р.У., Донко З.</i> Үш өлшемді Юкава жүйесінің диффузия коэффициентіне сыртқы магнит өрісінің әсері.....	25
<i>Грушевская Е.А., Лебедев И.А., Темиралиев А.Т., Федосимова А.И.</i> Асимметриялы ядролардың өзара әрекеттерінде снарядтың ядросының толық талқандану жағдайларының сипаттамаларын зерттеу.....	30
<i>Асқарова А., Жұмаханова А.С., Құдайкұлов А., Ташев А.А., Қалиева Г.С.</i> Айнымалы жылу ағынының қатысуымен көлденең қимасының жылу және жылу окшаулаумен бөлек тұрақты жылуфизикалық жай-күйін зерттеу энергиясының әдісі.....	38
<i>Абишев М., Кенжебаев Н., Кенжебаева С., Джанибеков А.</i> Реакторлық нейтрондармен әсерлесудегі катализдық қоспаның изотоптық құрамын және энергия шығаруын есептеу.....	48
<i>Абишев М., Хасанов Н.</i> Жылулық нейтрондардың катализдық қоспамен (Pb, Bi, Po) әсерлесуін "IBUS" компьютерлік бағдарламалау кешенімен жобалау.....	53
<i>Алдабергенова Т.М., Ганеев Г.З., Кислицын С.Б., Досболаев М.К.</i> Графит бетінің термиялық эрозиясы мен құрылымына импульстік плазмалық сәулелендірудің ықпалы.....	57
<i>Жақып К.Б.</i> Стокса және Навье теңдеулерінің генеалогиялары. Дәрежелік реологиялық заңдар және теңдеулер.....	64
<i>Жаугашева С.А., Валиолда Д.С., Джансейтов Д.М., Жусупова Н.К., Сериков Ж., Айтжан Ф.</i> Теоретическое исследование кулоновского развала гало ядер ¹¹ Be, ¹⁵ C.....	81
<i>Жаугашева С.А., Сайдуллаева Г.Г., Нурбакова Г.С., Хабыл Н., Турарбекова М.М.</i> В(Bs) Мезонның ауыр мезондарға ыдырау қасиетін релятивистік әсерлесуін ескере отырып анықтау	86
<i>Қошанов Б.Д., Нұрыкенова Ж.С.</i> Жоғарғы ретті эллиптикалық теңдеулер үшін жалпылаған Дирихле - Нейман есебінің шешілімі туралы.....	95
<i>Құралбаев З.К., Оразаева А.Р., Рахимжанова З.М.</i> Жоғары көтерілген магма заттарының әсерінен болатын астеносферадағы қозғалыстың механика-математикалық моделі.....	103
<i>Мұқашев К.М., Казаченок В.В., Алиева М.Е.</i> Ғарыштық бөлшектер тұрғысынан физиканың іргелі проблемаларын оқытудың парадигмасы туралы жаңа көзқарастар.....	112
<i>Мырзақұл Т.Р., Таушинова А.С., Белисарова Ф.Б., Мырзақұл Ш.Р.</i> Гаусс-Бонн инвариантымен минималды емес байланыс кезіндегі <i>k</i> - эссенцияның инфляциялық моделі.....	120
<i>Омашова Г.Ш., Спабекова Р.С., Қабылбеков К.А., Саудахметов П.А., Абдрахманова Х.К., Арысбаева А.С.</i> Изохоралық процесті зерттеуге арналған компьютерлік зертханалық жұмысты ұйымдастырудың бланкі үлгісінің тапсырмаларын өз бетінше құрастыру.....	127
<i>Рябкин Ю.А., Рақыметов Б.А., Айтмукан Т.</i> Көміртек қабықшасының ЭПР-мәліметі негізінде қатты отын жалынының парамагниттік қасиетін анықтау мүмкіндігі.....	134
<i>Спабекова Р.С., Омашова Г.Ш., Қабылбеков К.А., Саудахметов П.А., Серикбаева Г.С., Актурева Г.К.</i> Тоқ көзін қосқанда және ажыратқанда тізбектегі токкүшінің өзгеруін зерттеуге арналған компьютерлік зертханалық жұмысты ұйымдастыруда матлав бағдарламасын қолдану.....	139
<i>Ташенова Ж.М., Калдарова М., Мусайф М.</i> Жылу ағыны, жылу алмасу және жылу изоляциясы бар үшөлшемді есептің тұрақты температуралы күйіндегі сандық сипаттамасы.....	148
<i>Ташенова Ж.М., Мусайф М., Калдарова М.</i> Термосерпімділікті есептеудегі энергетикалық әдісі.....	155
<i>Тұрғанбай Қ.Е., Қалдыбекова С.У.</i> Жоғарғы мектепте информатика пән мұғалімнің ойлау қабілетін жетілдіру ерекшеліктері.....	163
<i>Шоманов А.С., Ахмед-Заки Д.Ж., Амирғалиев Е.Н., Мансурова М.Е.</i> Кілттерді Mapreduce үлгісінде тарату есебі туралы	167
<i>Бакирова Э.А., Исакова Н.Б., Уаисов Б.</i> Параметрі бар фредгольм интегралдық- дифференциалдық теңдеуі үшін сызықты шеттік есепті шешудің бір алгоритмі туралы	173
<i>Ақылбаев М.И., Сапрыгина М.Б., Шалданбаев А.Ш.</i> Коэффициенті тұрақты, бірінші ретті кәдімгі дифференциалдық теңдеудің сингуляр әсерленген Коши есебін аргументтің ауытқыту әдісі арқылы шешу.....	181
<i>Рустемова К.Ж., Шалданбаев А.Ш., Ақылбаев М.И.</i> Коэффициенттері тұрақты, екінші ретті кәдімгі дифференциалдықтеңдеудің сингуляр әсерленген Коши есебін аргументтің ауытқыту әдісі арқылы шешу.....	193
<i>Аширбаев Х.К., Қабылбеков К.А., Абдрахманова Х.А., Джумағалиева А.И., Кыдырбекова Ж.Б.</i> MATLAB бағдарлама пакетін қолданып электр және магнит өрістерін зерттеуге арналған компьютерлік зертханалық жұмыстарды ұйымдастыру.....	206

СОДЕРЖАНИЕ

<i>Бердибай С.Б., Парецкая Н.А., Сабитов А.Н., Исламов Р.А., Тамазян Р.А., Токмолдин С.Ж., Ильин А.И., Мартиросян К.С.</i> Комплекс фенилаланина с иодом и его структура.....	5
<i>Кабышев А.М., Кутербеков К.А., Пенионжкевич Ю.Э., Маслов В.А., Мендибаев К., Соболев Ю.Г., Лукьянов С.М., Кабдрахимова Г.Д., Азнабаев Д., Курманжанов А.Т.</i> Статистические и систематические погрешности, полное сечение реакции, γ -спектрометр.....	10
<i>Бердибай С.Б., Парецкая Н.А., Сабитов А.Н., Исламов Р.А., Тамазян Р.А., Токмолдин С.Ж., Ильин А.И., Мартиросян К.С.</i> Комплекс фенилаланина с иодом и его структура.....	19
<i>Джумагулова К.Н., Рамазанов Т.С., Машеева Р.У., Донко З.</i> Влияние внешнего магнитного поля на коэффициент диффузии трехмерной Юкава системы.....	25
<i>Грушевская Е.А., Лебедев И.А., Темиралиев А.Т., Федосимова А.И.</i> Исследование событий полного разрушения ядра снаряда во взаимодействиях асимметрических ядер.....	30
<i>Аскарова А., Жумаханова А.С., Кудайкулов А., Ташев А.А., Калиева Г.С.</i> Энергетический метод в исследовании установившегося теплофизического состояния стержня переменного сечения при наличии теплового потока, теплообмена и теплоизоляции.....	38
<i>Абишев М., Кенжебаев Н., Кенжебаева С., Джанибеков А.</i> Расчет изотопного состава каталитического материала при облучении реакторными нейтронами.....	48
<i>Абишев М., Хасанов М.</i> Моделирование взаимодействия тепловых нейтронов каталитическим составом (Pb,Bi,Po) с помощью программного комплекса "IBUS".....	53
<i>Алдабергенова Т.М., Ганеев Г.З., Кислицин С.Б., Досболаев М.К.</i> Влияние импульсного плазменного облучения на термическую эрозию и структуру поверхности графита.....	57
<i>Джакупов К.Б.</i> Генезис уравнений Стокса и Навье. Степенные реологические законы и уравнения.....	64
<i>Жаугашева С.А., Валиолда Д.С., Джансейтов Д.М., Жусупова Н.К., Сериков Ж., Айтжан Ф.</i> ^{11}Be , ^{15}C Гало ядроларының кулондық күйреуін теориялық зерттеу.....	81
<i>Жаугашева С.А., Сайдудлаева Г.Г., Нурбакова Г.С., Хабыл Н., Турарбекова М.М.</i> Определение свойств тяжелого V(Bs)-мезона в рамках релятивистского характера взаимодействия.....	86
<i>Кошанов Б.Д., Нурикенова Ж.С.</i> О разрешимости обобщенной задачи Дирихле - Неймана для эллиптического уравнения высокого порядка.....	95
<i>Куралбаев З.К., Оразаева А.Р., Рахимжанова З.М.</i> Механико-математическая модель движений в астеносфере под воздействием поднимающихся мантийных веществ.....	103
<i>Мукашев К.М., Казаченок В.В., Алиева М.Е.</i> О новых взглядах на парадигму обучения фундаментальным проблемам физики на примере частиц космического происхождения.....	112
<i>Мырзақұл Т.Р., Таукенова А.С., Белисарова Ф.Б., Мырзақұл Ш.Р.</i> Инфляционная модель k -эссенции при неминимальной связи с инвариантом Гаусса-Боннэ.....	120
<i>Омашова Г.Ш., Саббекова Р.С., Кабылбеков К.А., Саидахметов П.А., Абдрахманова Х.К., Арысбаева А.С.</i> Самостоятельное конструирование заданий для выполнения компьютерной лабораторной работы по исследованию изохорного процесса.....	127
<i>Рябкин Ю.А., Ракыметов Б.А., Айтмуқан Т.</i> О возможности определения парамагнитных характеристик пламени твердого топлива на основе ЭПР-данных углеродных пленок.....	134
<i>Саббекова Р.С., Омашова Г.Ш., Кабылбеков К.А., Саидахметов П.А., Серикбаева Г.С., Актуреева Г.К.</i> Организация компьютерных лабораторных работ по исследованию тока включения и выключения с использованием пакета программ MATLAB.....	139
<i>Ташенова Ж.М., Калдарова М., Мусайф М.</i> Численное обоснование одномерности некоторой трехмерной задачи установившегося температурного состояния при наличии теплового потока, теплообмена и теплоизоляции.....	148
<i>Ташенова Ж.М., Мусайф М., Калдарова М.</i> Энергетический метод в решении задач термоупругости.....	155
<i>Турганбай К.Е., Қалдыбекова С.У.</i> Особенности развития мышления учителя информатики в высшей школе.....	163
<i>Шоманов А.С., Ахмед-Заки Д.Ж., Амиргалиев Е.Н., Мансурова М.Е.</i> О задаче оптимизации распределения ключей в Mapreduce модели.....	167
<i>Бакирова Э.А., Искакова Н.Б., Уайсов Б.</i> Об одном алгоритме решения линейной краевой задачи для интегро-дифференциального уравнения Фредгольма с параметром.....	173
<i>Ақылбаев М.И., Сапрыгина М.Б., Шалданбаев А.Ш.</i> Решение сингулярно возмущенной задачи Коши для обыкновенного дифференциального уравнения первого порядка с постоянным коэффициентом методом отклоняющегося аргумента.....	181
<i>Рустемова К.Ж., Шалданбаев А.Ш., Ақылбаев М.И.</i> Решение сингулярно возмущенной задачи Коши, для обыкновенного дифференциального уравнения второго порядка с постоянными коэффициентами, методом отклоняющегося аргумента.....	193
<i>Аширбаев Х.К., Кабылбеков К.А., Абдрахманова Х.А., Джумасалиева А.И., Кыдырбекова Ж.Б.</i> Организация компьютерной лабораторной работы по исследованию электрического и магнитного полей с использованием пакета программ MATLAB.....	206

CONTENTS

<i>Berdibay S.B., Paretskaya N.A., Sabitov A.N., Islamov R.A., Tamazyan R.A., Tokmoldin S.Zh., Ilin A.I., Martirosyan K.S.</i> Phenylalanine - iodine complex and its structure.....	5
<i>Kabyshv A.M., Kuterbekov K.A., Penionzhkevich Yu.E., Maslov V.A., Mendibayev K., Sobolev Yu.G., Lukyanov S.M., Kabdrakhimova G. D., Aznabayev D. T., Kurmanzhanov A. T.</i> Errors in the total reaction cross sections and energies of incident particles measured using modified transmission technique	10
<i>Berdibay S.B., Paretskaya N.A., Sabitov A.N., Islamov R.A., Tamazyan R.A., Tokmoldin S.Zh., Ilin A.I., Martirosyan K.S.</i> Phenylalanine complex with iodine and its structure.....	19
<i>Dzhumagulova K.N., Ramazanov T.S., Masheyeva R.U., Donkó Z.</i> Effect of magnetic field on diffusion coefficients of the three-dimensional yukawa systems.....	25
<i>Grushevskaya E.A., Lebedev I.A., Temiraliev A.T., Fedosimova A.I.</i> Study on events with complete destruction of projectile nucleus in interactions of asymmetric nuclei	30
<i>Askarova A., Zhumakhanova A.S., Kudaykulov A., Tashev A.A., Kaliyeva G.S.</i> The energy method in the study of steady-state thermophysical condition of a rod of variable cross section in the presence of heat flow, heat exchange and thermal insulation.....	38
<i>Abishev M., Kenzhebeyev N., Kenzhebeyeva S., Dzhanybekov A.</i> Calculation of isotopic composition of catalytic material under radiation by reactor neutrons.....	48
<i>Abishev M., Khassanov M.</i> Simulation of the thermal neutronsinteraction with catalytic composition (Pb, Bi, Po) by "IBUS" software.....	53
<i>Aldabergenova T.M., Ganeyev G.Z., Kislitsin S.B., Dosbolaev M.K.</i> Effect of pulsed plasma irradiation on thermal erosion and structure of graphite surface.....	57
<i>Jakupov K.B.</i> Genealogy of the Stokes and Navier equations. Degree rheological laws and equations.....	64
<i>Zhaugasheva S.A., Valiolda D.S., Janseitov D.M., Zhussupova N.K., Serikov Zh., Aitzhan F.</i> Theoretical study of the coulomb breakup of the halo nuclei ^{11}Be , ^{15}C	81
<i>Zhaugasheva S.A., Saidullaeva G.G., Nurbakova G.S., Khabyl N., Turarbekova M.M.</i> Determination properties of heavy decay in the B(Bs) meson in the framework of the relativistic character of the interaction.....	86
<i>Koshanov B.D., Nurikenova J.</i> On solvability of the generalized Dirichlet-Neiman problem for a high order elliptic equation.....	95
<i>Kuralbaev Z.K., Orazaeva A.R., Rahimzhanova Z.M.</i> Mechanical-mathematical model of kinematics in the asthenosphere under the influence of rising mental substances.....	103
<i>Mukashev K.M., Kazachenok V.V., Alieva M.E.</i> About new look at the paradigm of study fundamental problems of physics of cosmic the example of origin.....	112
<i>Myrzakul T.R., Taukenova A.S., Belisarova F.B., Myrzakul S.R.</i> Inflation model of k -essence for non minimally coupled Gauss-Bonnet invariant.....	120
<i>Omashova G. Sh., Spabekova R.S., Kabylbekov K.A., Saidakhmetov P.A., Abdrakhmanova KH.K., Arysbaeva A.S.</i> Independent designing of tasks for performance of computer laboratory work on the investigation of the isophoric process...	127
<i>Ryabikin Yu.A., Rakymetov B.A., Aitmukan T.</i> On the possibility of determination of paramagnetic characteristics of flame of solid fuel on the basis of epr-data carbon films.....	134
<i>Spabekova R. S., Omashova G.SH., Kabylbekov K. A., Saidakhmetov P. A., Serikbaeva G.S., Aktureeva G.K.</i> Organization of computer laboratory works on the research of turnonand turnoff current with the use of matlab program package	139
<i>Tashenova Zh., Kaldarova M., Mussaif M.</i> One-dimensional numerical substantiation of some three-dimensional problem steady state temperature in the presence of heat flow, heat exchange and thermal insulation.....	148
<i>Tashenova Z., Mussaif M., Kaldarova M.</i> Energy method in decision problems thermoelasticity.....	155
<i>Turganbay K.E., Kaldibekoba S.U.</i> Features of thinking of the teacher of Informatics in high school.....	163
<i>Shomanov A.S., Akhmed-Zaki D.Zh., Amirgaliyev E.N., Mansurova M.E.</i> About the problem of key distribution in Mapreduce model	167
<i>Bakirova E.A., Iskakova N.B., Uaisov B.</i> On the algorithm for solving of a linear boundary value problem for fredholm integro-differential equation with parameter.....	173
<i>Akylbaev M.I., Saprigina M.B., Shaldanbaeva A.Sh.</i> Solution of a singularly perturbed Cauchy problem, for an ordinary differential equation of the first order with a constant coefficient, by the method of a deviating argument.....	181
<i>Rustemova K.Zh., Shaldanbaeva A.Sh., Akylbaev M.I.</i> Solution of a singularly perturbed Cauchy problem for an ordinary second-order differential equation with constant coefficients by the method of a deviating argument.....	193
<i>Ashirbaev H.A., Kabylbekov K. A., Abdrahmanova H. K., Dzhumagalieva A.I., Kydyrbekova Zh.B.</i> Organization of computer laboratory works to study electric and magnetic fields using the software package matlab.....	206

**Publication Ethics and Publication Malpractice
in the journals of the National Academy of Sciences of the Republic of Kazakhstan**

For information on Ethics in publishing and Ethical guidelines for journal publication see <http://www.elsevier.com/publishingethics> and <http://www.elsevier.com/journal-authors/ethics>.

Submission of an article to the National Academy of Sciences of the Republic of Kazakhstan implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see <http://www.elsevier.com/postingpolicy>), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The National Academy of Sciences of the Republic of Kazakhstan follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct (http://publicationethics.org/files/u2/New_Code.pdf). To verify originality, your article may be checked by the Cross Check originality detection service <http://www.elsevier.com/editors/plagdetect>.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the National Academy of Sciences of the Republic of Kazakhstan.

The Editorial Board of the National Academy of Sciences of the Republic of Kazakhstan will monitor and safeguard publishing ethics.

Правила оформления статьи для публикации в журнале смотреть на сайтах:

www.nauka-nanrk.kz

<http://www.physics-mathematics.kz>

ISSN 2518-1726 (Online), ISSN 1991-346X (Print)

Редакторы *М. С. Ахметова, Д.С. Аленов, Т.А. Апендиев*
Верстка на компьютере *А.М. Кульгинбаевой*

Подписано в печать 10.04.2017.
Формат 60x88¹/₈. Бумага офсетная. Печать – ризограф.
6,5 п.л. Тираж 300. Заказ 3.

Национальная академия наук РК
050010, Алматы, ул. Шевченко, 28, т. 272-13-18, 272-13-19