Formation of competencies in parallel computing in databases by future information specialists

Meruert Serik†, Saule K. Zhumagulova†, Dinara A. Kazimova‡ & Gulnur B. Abildaeva*

L.N. Gumilyov Eurasian National University, Astana, Kazakhstan†
E.A. Buketov Karaganda University, Karaganda, Kazakhstan‡
A. Saginov Karaganda Technical University, Karaganda, Kazakhstan*

ABSTRACT: Currently, parallel computing is an integral part of the field of database software development. The article describes the current problems of improving the professional training of future IT specialists. In this article, the authors present a methodology for teaching students parallel programming aimed at forming their parallel style of thinking, as well as acquiring practical skills in creating parallel computing in a database in the process of subject training at a university in Kazakhstan. The developed methodology was expected to play a crucial role in skills acquisition, and relevant data were collected to assess its impact. The results of the data obtained allowed the authors to assert that the applied methodology for preparing future IT specialists for professional activity contributes to the formation of students as highly qualified IT specialists. Thus, the approach proposed by the authors to the selection and application of methods at different stages of training will ensure that students successfully master the practical skills of organising parallel computing in databases.

INTRODUCTION

Today, every IT specialist should have the required skills of writing programs, and skills in the field of databases (DB), as well as their management [1].

The content of educational programmes for training programmers in universities includes a considerable list of courses, leading students to the formation of the required competencies [2]. Among them, competence in the field of parallel computing in DB can be distinguished as an attribute of a person who has the ability to create parallel queries to the DB.

In the article, the authors present the characteristics of an experimental study aimed at the introduction of a new methodological approach for the training of future IT specialists, in educational institutions, and outline the analysis of the data obtained. Also, within the framework of the developed methodology, the article includes conclusions on the effectiveness of the formation of students' competence in the field of parallel computing in the database [3].

The research was based on a literature review of scientific and specialised sources containing information about the principles of creating databases and parallel computing in them, including a work by Korneev [4].

Voevodin et al considered the tasks of the methodology of teaching parallel computing in higher educational institutions and noted that the study of parallel programming by students is a complex pedagogical problem due to the fact that it is important for the process of forming their mental activity [5]. In order to cope with the task of parallel data processing, a person must have a parallel style of thinking that allows creating a parallel algorithm [6].

Sadikov has made a significant contribution to the study of parallel computing in databases [7]. He investigated two ways of creating parallel programs, identified their pros and cons and solved the problem of database access by deploying a virtual private network (VPN) connection between cluster nodes [7].

Also, in the course of this research, experience has been accumulated in teaching parallel programming to students.

RESEARCH METHODOLOGY

In view of the consulted literature and own experience, and in order to form students' competence in the field of database programming and parallel computing in databases, the authors proposed a new methodological approach for teaching students parallel programming. It was envisaged that this training system would allow students to form a parallel style of thinking and gain skills in creating parallel computing in the database in the educational process at university.

The new methodology was applied in the form of an experiment in some universities in Kazakhstan, including the verification of its effectiveness in regard to the formation of students' competencies in parallel computing in

databases. In the organisation and implementation of the experimental work plan, the authors adopted the principles set out in the works of Voevodin [2], Antonov [3], Korneev [4] and others.

A 100-point scale was used as a tool for determining the level of competence in the field of DB and parallel computing in the database, which allowed to assess the level of students' ability to work with parallel computing in the database.

The experiment was implemented using the parallel method with two groups involved: control and experimental. In the control group the independent and dependent variables were measured in parallel with their counterparts in the experimental group, and the control group was not involved in the experiment, but had their regular, traditional classes in the subject *Database*.

The experimental group of students participated in lectures, which were explanatory and illustrative, and in the course of their practical work, problem-based research projects were undertaken. As part of their practical work, the students were offered to build a database with a virtual server on the Azure cloud platform and a local server on MS SQL server, in accordance with the methodological recommendations they had received earlier. The students were faced with the task of creating and sending requests capable of processing data both on the virtual and local server in parallel from several client stations. To do this, it was proposed to use query parallelisation methods based on a parallel computing library. The students performed the installation and configuration of virtual and local servers to work with the developed database, as well as its testing.

In the process of work, they became acquainted with the scheme of completing tasks, compiling programs, running sequential and parallel queries. In addition, by using a query execution timer when writing a program, they were able to compare the execution time of sequential and parallel algorithms for implementing database queries. This method of comparison allows to demonstrate the potential efficiency of parallel database queries using examples.

Stages of the experiment:

1. The ascertaining stage of scientific research is designed to determine the main characteristics of the research object, including the formation of the control and experimental group. At this stage, the authors specified the problem and determined the degree of its influence on the activity of the participants. The problem was the enhancement of parallel thinking skills of students through the application of a new methodology. A group of computer science students from Kazakh universities participated in this research. The data were analysed according to specific indicators through mathematical methods. The assessment of the level of competence formation in the field of parallel computing in DB was given to the participants of these groups with the identification of disadvantages.
2. Formative stage, with the use of the forms, methods and means proposed by the authors for the experimental group. The formative experiment was designed to determine the behaviour of the participants, taking into account the influence of specific factors on them. The formative experiment included development of an action plan and selection of the main parameters. Next, was the organisation of the experiment that involved immersion of the participants in specific conditions and assessment of the impact - what has changed and why, level of change in the indicators and results. At this stage, the hypothesis put forward by the authors was tested (confirmed/refuted) based on the evidence collected.
3. At the control stage, the control experiment was designed to record the results obtained and compare them with the initially obtained data leading to fully fledged conclusions. At this time, secondary diagnostics of competence formation in parallel computing in databases were implemented. In addition, by means of statistical criteria, the analysis of the obtained data of the ascertaining and formative stages was carried out.

More specifically, third-year students in computer science from Kazakhstan universities took part in the experiment. Further, a comparative analysis of the obtained assessment with the diagnostic results was performed in order to obtain a picture of the dynamics of the skills formation process.

Below, are highlighted key results of the conducted research in the field of parallel computing in databases.

1. Professional competence of IT specialists [8].

In order to create a structure of professional competence of future IT specialists, the authors have formed a general list of competencies and defined general and professional competencies. The general competencies included the ability to assess the degree of influence of external factors on the effectiveness of professional activity, to write complex programs in groups, to put into practice the acquired knowledge, etc [9].

The professional competencies included the ability to write programs using modern methods of data access and processing; the ability to write effective algorithms for parallel computing in a database; the ability to apply modern paradigms of parallel programming in order to process large amounts of data, etc.

As pointed out earlier by Serik et all, knowledge, skills and abilities make up the professional competencies of future IT specialists: competence in the field of DB programming and parallel computing in the database, communication skills, mathematical, engineering, managerial and professional competencies are needed to form a high-level IT specialist [10].

A significant role among the professional competencies of future IT specialists is played by competence in the field of parallel computing in databases, which consists of knowledge and skills in writing parallel query processing programs.

2. The system of training future IT specialists for professional activity.

The proposed system can be presented in the form of a system, the basis of which is the analysis of the labour market requirements for the qualification of specialists, the integration of methods of their education at university [11]. The key idea is to create a system of professional training of IT specialists during the educational process, which provides:

- formation of professional competence of students on the basis of regular analysis of the labour market;
- implementation of the proposed system for students to study programming of parallel computing in databases.

3. A combination of general didactic and specific teaching methods and tools [12].

The immediate and more distant future of university graduates is to be fully and professionally engaged in the global society, enterprise and commerce. The quality of education that they obtained will impact directly on their recruitment drive, and the capacity to further develop professionally and achieve high living standards. Hence, the understanding of these challenges creates the tendency for educational institutions to revisit, reevaluate and reform the educational programmes that they offer [13].

At the ascertaining stage of the experiment, the assessment of the level of students' competence in parallel computing in databases was carried out. The academic performance of students was analysed on the basis of data from periodic checks of students' knowledge in DB disciplines from records of students' academic performance in study groups.

RESULTS

The level of knowledge and skills at the ascertaining stage was as follows: the average and high levels of formation of the required competence were demonstrated by 43 students of the control group, which was 62.94%, and 45 students (64.28%) in the experimental group. Low levels of the indicators were shown by 24 students (37.07%) of the control group and 25 (35.71%) of the experimental group. Thus, more than 35% of the participants in the experiment demonstrated a low level of competence formation in parallel computing in database, that is, students at that stage had a fairly low level of training.

In Table 1, the data on the initial state of students' knowledge and skills of working with parallel computing in databases are presented, based on five criteria and levels defined to reflect the students' readiness to work with these resources.

Table 1: The state of the initial level of preparation for working with parallel computing in databases (ascertaining stage).

No.	Criteria for readiness to work with parallel computing in databases	Low level		Sufficient level		High level	
		Groups					
		Control	Experimental	Control	Experimental	Control	Experimental
		No.	No.	No.	No.	No.	No.
		%	%	%	%	%	%
1.	Ability to self-educate	24	26	37	36	6	8
		35.82%	37.14%	55.22%	51.43%	8.96%	11.43%
2.	Possession of the required volume of special theoretical knowledge	26	24	36	40	5	6
		38.81%	34.29%	53.73%	57.14%	7.46%	8.57%
3.	Ability to create an algorithm for processing large amounts of information in databases	28	24	35	39	4	7
		41.79%	34.29%	52.24%	55.71%	5.97%	10.0%
4.	Knowledge of effective implementation methods of parallel data processing in the design of databases	23	25	38	38	6	7
		34.33%	35.71%	56.72%	54.29%	8.96%	10.0%
5.	Creating databases using parallel computing	25	26	39	39	3	5
		37.31%	37.14%	58.21%	55.71%	4.48%	7.14%
6.	Knowledge of effective implementation methods of parallel data processing in database development	23	25	41	38	3	7
		34.33%	35.71%	61.19%	54.29%	4.48%	10.0%
7.	Average value	24	25	38	38	5	7
		37.07%	35.71%	56.22%	54.76%	6.72%	9.52%

In the formative stage of the experiment, the following activities were undertaken:

1. Analysis of employers' qualification requirements for IT specialists in regard to setting up the task of forming professional competence.
2. Implementation of the proposed system of teaching through the formation of professional competence.
3. Updating the content of training of future IT specialists with emphasis on the formation of skills and abilities essential for solving problems in professional activity.

During the formative stage of the experiment, students in the experimental group formed educational material according to the developed programmes, by independently analysing electronic sources and recommended literature. The emphasis was on independent work, and knowledge and skills formation by students themselves.

At the control stage of the experiment, the students' academic performance was analysed based on the data of the second control period. The results showed average and high indicators for 42 students in the control group (63.19%), with 25 students (36.82%) in the low range. Sixty-four participants (91.43%) of the experimental group had average and high indicators, while six students were in the low range (8.57%) (Table 2).

These data indicate that the participants of the experimental group had a higher level of competence in parallel computing in databases, and that their work was done efficiently.

Table 2: Formation of readiness to work with parallel computing in databases (formative experiment).

No.	Criteria for readiness to work with parallel computing in databases	Low level		Sufficient level		High level	
		Groups					
		Control	Experimental	Control	Experimental	Control	Experimental
		No.	No.	No.	No.	No.	No.
		%	%	%	%	%	%
1.	Ability to self-educate	22	5	38	50	7	15
		32.84%	7.14%	56.72%	71.43%	10.45%	21.43%
2.	Possession of the required volume of special theoretical knowledge	25	8	37	50	5	12
		37.31%	11.43%	55.22%	71.43%	7.46%	17.14%
3.	Ability to create an algorithm for processing large amounts of information in databases	28	6	35	51	4	13
		41.79%	8.57%	52.24%	72.86%	5.97%	18.57%
4.	Knowledge of effective implementation methods of parallel data processing in the design of databases	24	6	39	52	4	12
		35.82%	8.57%	58.21%	74.29%	5.97%	17.14%
5.	Creating databases using parallel computing	24	6	40	53	3	11
		35.82%	8.57%	59.7%	75.71%	4.48%	15.71%
6.	Knowledge of effective implementation methods of parallel data processing in database development	25	5	38	54	4	11
		37.31%	7.14%	56.72%	77.14%	5.97%	15.71%
7.	Average value	25	6	38	52	4	12
		36.82%	8.57%	56.47%	73.81%	6.72%	17.62%

In general, the results of the control stage of the experiment allowed to observe the absence of changes in the level of competence formation in the control group. At the same time, in the experimental group, the low indicator had decreased (by 0.271), while the medium and high increased (by 0.19 and 0.081, respectively). The dynamics of the indicators in regard to the formation of students' readiness to work is shown in Table 3.

Table 3: Dynamics of the indicators of the formation of students' readiness to work with parallel computing in databases.

Level of formation	Quantitative indicators (experimental group)		Difference	Quantitative indicators (control group)		Difference
	Ascertaining stage	Control stage		Ascertaining stage	Control stage	
Low	0.357	0.086	-0.271	0.371	0.368	-0.003
Sufficient/medium	0.548	0.738	0.190	0.562	0.565	0.003
High	0.095	0.176	0.081	0.067	0.067	0

The reliability of the experimental results has been proved by means of the Fisher criterion φ (phi) and the Pearson uniformity criterion $\chi 2$ (chi square). At the ascertaining stage of the experiment, the identity of the average and low

indicators of competence formation in parallel computing in databases was determined. The fact that the indicators of the empirical values of the criterion are less than its critical value (φcritical = 1.64) is part of the argument. In the context of the individual criteria: *Ability to self-educate* (the average level of φempirical = 0.45, the low level of φempirical = 0.16); *Possession of the required volume of special theoretical knowledge* (the average level of φempirical = 0.4, the low level of φempirical = 0.55); *Ability to create an algorithm for processing large amounts of information in databases* (the average level of φempirical = 0.41, the low level of φempirical = 0.907); *Knowledge of effective implementation methods of parallel data processing in the design of databases* (the average level of φempirical = 0.287, the low level of φempirical = 0.907); *Creating databases using parallel computing* (the average level of φempirical = 0.298, the low level of φempirical = 0.023); *Knowledge of effective implementation methods of parallel data processing in database development* (the average level of φempirical = 0.819, the low level of φempirical = 0.176).

The control stage of the experiment showed differences in low and average indicators of competence formation in parallel computing in databases, as evidenced by the greater empirical values obtained than the critical value of the criterion, equal to φcritical = 1.64. In the context of the individual criteria: *Ability to self-educate* (the average level of φempirical = 1.8, the low level of φempirical = 3.98): *Possession of the required volume of special theoretical knowledge* (the average level of φempirical = 1.98, the low level of φempirical = 3.65); *Ability to create an algorithm for processing large amounts of information in databases* (the average level of φempirical = 2.7, the low level of φempirical = 4.57); *Knowledge of effective implementation methods of parallel data processing in the design of databases* (the average level of φempirical = 2.007, the low level of φempirical = 4.025); *Creating databases using parallel computing* (the average level of φempirical = 2.563, the low level of φempirical = 4.534); *Knowledge of effective implementation methods of parallel data processing in database development* (the average level of φempirical = 2.019, the low level of φempirical = 4.025).

The assessment results of the reliability of generalised indicators using the Fisher criterion showed that at the ascertaining stage there was no significant difference between the low and average indicators in both groups (the average level of φempirical = 0.287 < φcritical, the low level of φempirical = 0.012 < φcritical). The data obtained for the two groups involved in the experiment show the presence of significant differences between the participants of the control and experimental groups (average - φempirical = 2.182 > φcritical, low - φempirical = 4.207 > φcritical).

This is also confirmed by the results of comparing the indicators of competence formation using the Pearson criterion. The ascertaining stage of the experiment did not reveal any differences between the studied samples ($\chi 2$ empirical < $\chi 2$ (chi square) critical = 5.99). At the same time, the control stage showed significant differences between them ($\chi 2$ (chi square) empirical > $\chi 2$ critical = 5.99). The statistics were: at the ascertaining stage, $\chi 2$ empirical = 0.37, at the control stage - $\chi 2$ empirical = 18.62.

Thus, the data of the conducted experiment allow to assert that the indicators of competence in parallel computing in databases among the students of the control and experimental groups have significant differences. It can also be stated that due to the implementation of the new methodology for training future IT specialists for professional activity, the graduates will have the potential to become highly qualified employees.

DISCUSSION

The key goal of the study was that students would gain knowledge about parallelism in computing by writing parallel queries in databases, as well as develop competencies for creating parallel query algorithms.

In the course of practical work, students will be able to implement a parallel algorithm for processing information in a database.

Taking into account the knowledge gained by students during lectures, the established methodological guidelines for the implementation of practical work make it possible for students to study parallel programming in practice.

The effectiveness of the formation of future IT specialists' competence profile in parallel computing in databases is provided by a set of methods and training tools.

The participants of the experiment were third-year students from Kazakh universities. At the first, ascertaining stage, the primary diagnostics of the level of competence formation were carried out. At the formative stage, the content of training was updated by introducing methods, forms and means of teaching parallel computing in databases. At the final, control stage, secondary diagnostics, analysis of the data obtained and their comparison with the initial level of formation of the specified competence were conducted.

The result of the formative stage of the experiment demonstrated an increase in the level of formation of the analysed competence among students. At the same time, there was no particular change in the same indicators for the control group of students. The effectiveness of the applied methodology is evidenced by the significant difference between the levels of competence formation among students of both groups.

The results of the experiment allowed to formulate a number of recommendations to ensure the quality of training of IT specialists in parallel computing in databases; thus it is proposed to:

- modernise the content and methodology of training in parallel computing in databases for IT specialists in accordance with international standards;
- analyse the quality of training in working with databases of IT students;
- analyse the labour market to determine employers' requirements for training students to work with databases.

In the course of their work, students study the scheme of task implementation, program compilation, implementation of sequential and parallel queries. In addition, they will be able to compare the implementation time of sequential and parallel algorithms for executing queries to the database, which can give a motivational effect. Such a comparative method allows to show the potential efficiency of parallel database queries by demonstrating examples.

The results can be used as a basis for further research in this area in order to obtain more scientific materials for comparing the results.

CONCLUSIONS

The article proposed a methodology for teaching parallel computing in a database to IT students. Within the framework of the developed methodology, the authors draw conclusions about the effectiveness of the formation of students' competence in the field of parallel computing in databases.

The results of the presented study can be of practical value in the development of educational programmes for training IT specialists, and for those studying at universities.

Depending on the needs of students, the following results of the proposed training system are determined:

1. Formation of a parallel style of thinking.
2. Obtaining the skills to create a parallel algorithm for developing database queries.
3. Acquisition of knowledge about parallel systems, principles of writing parallel database queries.
4. Getting skills in writing parallel computing programs in databases.

Thus, the proposed methodological system at different stages of training will allow students to successfully master the practical skills of organising parallel computing in databases.

REFERENCES

1. Bogachev, K.J., *Osnovy Parallel'nogo Programmirovanija.* Moscow: Laboratorija Znanij, 127-131 (2016) (in Russian).
2. Voevodin, V.V. and Voevodin, V.V., *Parallel'nye Vychislenija*. Saint-Petersburg: BHV-Peterburg, 311-315 (2018) (in Russian).
3. Antonov, A.S., *Vvedenie v Parallel'nye Vychislenija.* Moscow: Fizicheskogo Fakul'teta MGU, 70-75 (2016) (in Russian).
4. Korneev, V.V., *Parallel'nye Vychislitel'nye Sistemy.* Moscow: Nolidzh, 302-309 (2016) (in Russian).
5. Batlle, J., Martı, J., Ridao, P. and Amat, J., A new FPGA/DSP-based parallel architecture for real-time image processing. *Real-Time Imaging*, 8, **5**, 345-356 (2016).
6. Feo, J.T. (Ed), *A Comparative Study of Parallel Programming Languages: the Salishan Problems.* North Holland: Elsevier, 233-235 (2018).
7. Sadikov, M.G., *Parallel'nye Vychisleniya i Bazy Dannyh*. Moscow: Nolidzh, 263-268 (2014) (in Russian).
8. Flynn, M., *Computer Architecture: PipelinAed and Parallel Processor Design*. London: Jones and Bartlett Publishers, 181-185 (2017).
9. Serik, M., Yerlanova, G., Karelkhan, N. and Temirbekov, N., The use of the high-performance computing in the learning process. *Inter. J. of Emerging Technologies in Learning*, 16, **17**, 240-254 (2021).
10. Serik, M., Karelkhan, N., Kultan, J. and Zulpykhar, Z., Setting up and implementation of the parallel computing cluster in higher education. *Inter. J. of Emerging Technologies in Learning*, 14, **6**, 4-17 (2019).
11. Kazimova, D.A., Abdrasheva, B.Z., Baimukanova, M.T., Rolupan, K.L. and Musatai, N.A., Innovative approaches in training specialists in conditions of digitalization of education. *Bulletin of the Karaganda University. Pedagogy Series*, 4, **92**, 89-94 (2018).
12. Golovachyova, V.N., Tomilova, N.I. and Abildaeva, G.B., The expert system of control and knowledge assessment. *Scientific J. of Astana IT University*, 1, 21-29 (2020).
13. Danilova, D.A. and Pudlowski, Z.J.,. Engineering education reform: the imperatives for ensuring its quality and outcomes. *World Trans. on Engng. and Technol. Educ.*, 6, **2**, 239-243 (2007).