Ministry of Science and Higher Education of the Russian Federation
NATIONAL RESEARCH
TOMSK STATE UNIVERSITY (NR TSU)

Institute of Applied Mathematics and Computer Science

APPROVE:
Director
_____ A. V. Zamyatin

Work program of the discipline

High-performance computing

in the major of training

01.04.02 Applied mathematics and informatics

Orientation (profile) of training:

Big Data and Data Science

Form of study
full-time

Qualification
Master

Y ear of admission
2024

Code of discipline in the curriculum: B1.O.11

AGREED:
Head of EP
_____ A.V. Zamyatin

Chairman of the EMC
_____ S.P. Sushchenko

Tomsk – 2024

1. Purpose and planned results of mastering the discipline

The purpose of mastering the discipline is the formation of the following competencies:
- UK-1 - the ability to carry out a critical analysis of problem situations based on a systematic approach, to develop an action strategy;
- PK-6 - the ability to manage getting, storing, transferring, processing of big data.

The results of mastering the discipline are the following indicators of the achievement of competencies:
IUK-1.1 Identifies a problem situation, on the basis of a systematic approach, carries out its multifactorial analysis and diagnostics.
IUK-1.2 Carries out the search, selection and systematization of information to determine alternative options for strategic solutions in a problem situation.
IUK-1.3 Suggests and justifies the strategy of action, taking into account the limitations, risks and possible consequences.
IPK-6.1 Does monitoring and assess performance of data processing
IPK-6.2 Uses methods and instruments of retrieving, saving, transferring, processing of big data.
IPK-6.3 Prepare suggestions about increasing performance of big data processing.

2. Tasks of mastering the discipline

– Teach students methods of high-performance computing organization
– Form skills of using different programmatic tools for organizing parallel computing

3. The place of discipline in the structure of the educational program

Discipline belongs to the mandatory part of the educational program.

4. Semester of mastering and form of intermediate certification in the discipline

Third semester, examination.

5. Entrance requirements for mastering the discipline

For the successful mastering of the discipline, learning outcomes are required in the following disciplines: "Basics of Programming", "Computer Science".

6. Implementation language

English.

7. Scope of discipline

The total labor intensity of the discipline is 5 credits, 180 hours, of which:
- lectures: 20 hours
- laboratory: 40 hours
  including practical training: 0 h.

The volume of independent work of the student is determined by the curriculum.

8. The content of the discipline, structured by topics

Topic 1. Actuality, basic terminology and tendencies of development. Parallelism in computing.

Concept of high-performance computing. Ways of increasing performance of computer system. Concept of parallel computing. Reasons which lead to parallelism of computing. Multiprocessing. Classification of parallelism levels. Vector-based data processing. Multifunctional data processing. Pipeline of commands.

Topic 2. Architecture of systems for high-performance computing.

Flinn classification of computers. Systems with one command flow and one data flow. Systems with once command flow and multiple data flows. Systems with multiple command flows and multiple data flows. Classification of computers with multiple command flows and multiple data flows. Systems with shared memory. Systems with distributed memory. Systems with centralized shared memory. Systems with decentralized shared memory. Symmetrical multiprocessing. Architecture CC-NUMA. Architecture COMA. Architecture NCC-NUMA. Massively parallel processing. Clusters and their types.

Topic 3. Cloud technologies, their properties and types.

Concept of cloud technologies and computing. Properties of cloud technologies. Classification of cloud technologies by type of accessed resource. Classification of cloud technologies by deployment model.

Topic 4. Technology MapReduce.

Technology of data processing MapReduce, her scope. Implementations of MapReduce. Example of a program implemented with the help of MapReduce.

Topic 5. Distributed file systems.

Concept of distributed file system. Properties of distributed file systems. Google File System, Hadoop distributed file system, their architecture, similarities and differences.

Topic 6. Programming for high-performance computing. Methodology of designing parallel algorithms.

Problems of parallel programming. Methodology for organizing parallel computations for SIMD architecture. Methodology for organizing parallel computations for MIMD architecture. Quality metrics for parallel algorithms. Libraries for message exchange. MPI (Message Passing Interface). OpenMP (Open Multi-Processing). Example of a program, which uses OpenMP.

9. Ongoing evaluation

The ongoing evaluation is carried out by monitoring attendance, conducting tests, tests on lecture material, performing laboratory work, and is recorded in the form of a control point at least once a semester.

10. The procedure for conducting and criteria for evaluating the intermediate certification

The test with an assessment is carried out in writing by tickets and then answering orally.

An approximate list of theoretical questions:

1. Types of computing parallelism
2. Flinn’s classification of computers
3. Classification of MIMD systems
4. Properties and types of cloud technologies
5. Technology of distributed computation MapReduce
6. Concept of distributed file system
7. Google File System
8. Hadoop Distributed File System
9. Methodology of organizing parallel computations for SIMD and MIMD architectures
10. Quality metrics for parallel algorithms
11. OpenMP and MPI: purpose of each technology and their comparative analysis

The results of the assessment with an assessment are determined by the marks "excellent", "good", "satisfactory", "unsatisfactory". It is necessary to have 2 laboratory works done before taking the examination. If less than 3 laboratory works are done, than the maximal

mark on the exam is “good”. To get "excellent" mark it is also necessary to answer additional oral question (without having time to prepare). If a student has answered only the question from the ticket, but gave the complete answer, he gets mark "good". If a student has only answered the question from the ticket partially, he gets mark "satisfactory".

11. Educational and methodological support

a) Electronic training course on the discipline at the electronic university "Moodle" - https://moodle.tsu.ru/course/view.php?id=00000

b) Estimated materials of the current control and intermediate certification in the discipline.

c) Guidelines for laboratory work.

d) Guidelines for organizing independent work of students.

12. List of educational literature and Internet resources

№ п/п	Авторы / составители	Заглавие	Издательство	Год издания
Main literature				
1	S. Srinivasan	Cloud Computing Basics electronic resource	New York, NY : Springer New York : Imprint: Springer,	2014
2	A Ohri	R for Cloud Computing electronic resource : An Approach for Data Scientists	New York, NY : Springer New York : Imprint: Springer	2014
3	Zaigham Mahmood	Cloud Computing: Methods and Practical Approaches	London : Springer London : Imprint: Springer	2013
Additional literature				
1	Christoph Fehling, Frank Leymann, Ralph Retter, Walter Schupeck, Peter Arbitter	Cloud Computing Patterns electronic resource : Fundamentals to Design, Build, and Manage Cloud Applications	Vienna : Springer Vienna : Imprint: Springer	2014
2	Xiaolin Li, Judy Qiu	Cloud Computing for Data-Intensive Applications electronic resource	New York, NY : Springer New York : Imprint: Springer	2014
3	Zaigham Mahmood	Cloud Computing electronic resource : Challenges, Limitations and R&D Solutions	New York, NY : Springer New York : Imprint:	2014

			Springer	
Resources from the internet				
1		MapReduce Tutorial	https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html	
1		HDFS Architecture Guide	https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html	

13. List of information technologies

a) licensed and freely distributed software:

Microsoft Visual Studio Community Edition 2013 (C++), library of classes, which implements technology MapReduce (for example, https://github.com/cdmh/mapreduce). All these applications and libraries are freely distributed.

b) information reference systems:

– Electronic catalog of the TSU Scientific Library – http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system
– TSU electronic library (repository) – http://vital.lib.tsu.ru/vital/access/manager/Index
– EBS Lan – http://e.lanbook.com/
– EBS Student Advisor – http://www.studentlibrary.ru/
– Educational platform Urayt – https://urait.ru/
– EBS ZNANIUM.com – https://znanium.com/
– EBS IPRbooks – http://www.iprbookshop.ru/

c) professional databases:

– University Information System RUSSIA – https://uisrussia.msu.ru/

14. Logistics

Halls for conducting lecture-type classes.

Classrooms for conducting seminar-type classes, individual and group consultations, current control and intermediate certification.

Rooms for independent work, equipped with computer technology and access to the Internet, to the electronic information and educational environment and to information reference systems.

Classrooms for doing laboratory works must be equipped with Nvidia GPUs with CUDA technology support.

15. Authors information

Denis Druzhinin. Candidate of Technical Sciences, Docent, Department of Theoretical Foundations of Informatics, Institute of Applied Mathematics and Computer Science, TSU.