

**Federal State Autonomous Educational Institution of Higher Education "Moscow
Institute of Physics and Technology
(National Research University)"**

APPROVED
Vice Rector for Academic Affairs

A.M. Raygorodskiy

Work program of the course (training module)

course:	Distributed and Cloud Computing/Распределенные и облачные вычисления
major:	Applied Mathematics and Informatics
specialization:	Modern State of Artificial Intelligence/Современные методы искусственного интеллекта
	“Pusk” Online and Supplementary Education Centre
	Chair of Machine Learning and Digital Humanities
term:	2
qualification:	Master

Semester, form of interim assessment: 3 (fall) - Exam

Academic hours: 60 AH in total, including:

lectures: 30 AH.

seminars: 30 AH.

laboratory practical: 0 AH.

Independent work: 45 AH.

Exam preparation: 30 AH.

In total: 135 AH, credits in total: 3

Author of the program: R.G. Neychev, professor

The program was discussed at the Chair of Machine Learning and Digital Humanities 05.03.2021

Annotation

Novel approaches in the field of AI require huge amounts of data and computational resources. This course focuses on techniques of storing and accessing the data efficiently and distributing the computations between several instances. Special attention is paid to the Deep Learning models serialization. This course accompanies the Software Development and Machine Learning courses.

1. Study objective

Purpose of the course

- Learn how to speed up the computations
- Learn how to distribute data between instances effectively
- Gain essential experience with distributed computing
- Get ready to run the Machine Learning and Deep Learning techniques on several machines

Tasks of the course

- Developing the serializable models
- Working with distributed file systems
- Speeding up the computations with specific hardware

2. List of the planned results of the course (training module), correlated with the planned results of the mastering the educational program

Mastering the discipline is aimed at the formation of the following competencies:

Code and the name of the competence	Competency indicators
Gen.Pro.C-1 Address current challenges in fundamental and applied mathematics	Gen.Pro.C-1.2 Consolidate and critically assess professional experience and research findings
Gen.Pro.C-5 An understanding of current scientific and technical problems in the field of informatics and computer technology, and is able to formulate professional tasks in scientific language	Gen.Pro.C-5.1 An understanding of the current state of research within his/her professional thematic area
	Gen.Pro.C-5.2 Able to assess the relevance of research in informatics and computer technology and its practical relevance
	Gen.Pro.C-5.3 A good command of the professional terminology used in modern scientific and technical literature, and is able to present the results of scientific work orally and in writing as part of professional communication
Gen.Pro.C-6 Capable of selecting and/or developing approaches to solving typical and new problems in informatics and computer technology, taking into account the characteristics and limitations of different solution methods	Gen.Pro.C-6.1 Able to analyse the problem, plan the solution, suggest and combine ways of solving it
	Gen.Pro.C-6.2 Capable of developing and upgrading software and hardware for information and automated systems
	Gen.Pro.C-6.3 Able to use research methods to solve new problems by applying knowledge from different fields of science (technology)
	Gen.Pro.C-6.4 Proficient in analytical and computational solution methods, and understands and takes into account in practice the limits of applicability of the solutions obtained

3. List of the planned results of the course (training module)

As a result of studying the course the student should:

know:

- WLCG project as a data source: project work stages as a data provider, resource requirements; principles of organization of hierarchical levels (Tiers) WLCG - functional differences; the concept of SLA agreements;
- the main types of Grid infrastructures;
- functional (basic) scheme of Grid computing infrastructure; modern meta-scheduler schemes (Gridway, Unicore project, Community Scheduler Framework), general functioning;
- resource management systems (RMS); types of local (LRMS) and distributed (DRMS) systems; functions of the Grid component of the WS GRAM architecture;
- types and possibilities of planners; types of policies for queuing, changing priorities, managing access lists; the principle of organization and operation of Grid SE, Grid FTP / FTS;
- the concept of the Grid task; JDL task description language;
- the main components of the Grid for the Resource Center (RC); logic diagram of the RC; central services of regional infrastructures;
- main projects and components of middleware Grid (middleware); principles of modular installation and configuration of site packages;
- principles of monitoring systems; RGMA basics; Grid infrastructure monitoring sites;
- ways of organizing security in Grid infrastructures; GSI protocol, X509 certificates, MyProxy service; virtual communities and roles, reasons for sharing computing and storage resources, VOMS, VO box; shortcomings of the existing security system model;
- using virtualization in Grid; First WNoDeS Dynamic Provisioning Project (WeNMR Project), Achievements and Disadvantages; a scheme with a virtualization resource management interface; similarities and differences between Grid and cloud computing models;
- key features of the creation and operation of the PanDA (Production and Distributed Analysis) project system; managing the workflow of jobs over datasets; major infrastructure components;
- basic requirements for cloud computing systems; basic models (IaaS, PaaS, SaaS); principles of virtualization;
- principles of organizing cloud computing based on the OpenStack system; a diagram of the work of the subsystem for managing virtual machines Nova, managing object storage Swift, images of Glance virtual machines.

be able to:

- Install and configure a combination of Torque Resource Manager and MAUI Task Scheduler. Configure three (or more) queues with different duration of tasks execution in them and binding to different values of the "acl group" and "acl user" fields;
- develop a task file and check the correctness of the configuration, task status;
- set a policy for scheduling tasks using MAUI (or other schedulers).
- execute the configuration chain for YAIM CREAM-CE and WN targets on two nodes;
- configure the resource managers Condor and SLURM;
- create your own certification authority, which includes - self-signed root CA certificate; creating a signature request; Signing the request with your own CA obtaining a signed certificate and verifying it; export to pki form for import to web browser;
- create a program to work with the OpenStack API.

master:

- The concepts of SLA agreements;
- basic skills of working with the application for organizing Grid infrastructures Globus Toolkit;
- principles of organization and work with Grid Storage Element, Grid FTP / FTS;
- the language for describing Grid tasks JDL, including the classic description and the minimum required functional fields;
- options for organizing the movement of input and output files (stage IN / stage OUT) of Grid tasks; the concept of a workflow; capabilities of task flow management systems (WMS);
- the principles of modular installation and configuration of site packages using the YAIM utility;
- the ability to work with monitoring systems for the execution of Grid tasks; the concepts of availability and reliability; the concept of performance of the base computing core HepSpec06;
- principles of operation of GSI protocols, X509 certificates, MyProxy service;
- ways to manage the workflow of tasks in PanDA;
- SAGA application programming interface;
- the principles of organizing cloud computing based on the OpenStack system.

4. Content of the course (training module), structured by topics (sections), indicating the number of allocated academic hours and types of training sessions

4.1. The sections of the course (training module) and the complexity of the types of training sessions

№	Topic (section) of the course	Types of training sessions, including independent work			
		Lectures	Seminars	Laboratory practical	Independent work
1	Thread. Multithreading	10	10		15
2	SQL with Big Data. Hive	10	10		15
3	Hadoop and MapReduce	10	10		15
AH in total		30	30		45
Exam preparation		30 AH.			
Total complexity		135 AH., credits in total 3			

4.2. Content of the course (training module), structured by topics (sections)

Semester: 3 (Fall)

1. Thread. Multithreading

Consensus Algorithms CI/CD in distributed computing

2. SQL with Big Data. Hive

Spark NoSQL

3. Hadoop and MapReduce

Distributed optimization for Deep Learning

5. Description of the material and technical facilities that are necessary for the implementation of the educational process of the course (training module)

A standard classroom.

6. List of the main and additional literature, that is necessary for the course (training module) mastering

Main literature

Комбинаторная логика в программировании. Вычисления с объектами в примерах и задачах [Текст] : [учеб. пособие для вузов] / В. Э. Вольфенгаген ; НОУ Ин-т Актуального образования "ЮрИнфоР-МГУ, Каф. перспективных компьт. исслед. и информ. технологий .— 3-е изд., доп. и перераб. — М. : Ин-т "ЮрИнфоР-МГУ, 2008 .— 384 с.

Additional literature

Машинное обучение: новый искусственный интеллект [Текст]/Э. Алпайдин, -М., Изд. группа "Точка", 2017

7. List of web resources that are necessary for the course (training module) mastering

<http://dm.fizteh.ru/>

8. List of information technologies used for implementation of the educational process, including a list of software and information reference systems (if necessary)

Multimedia technologies can be employed during lectures and practical lessons, including presentations.

9. Guidelines for students to master the course

To successfully master the course, in addition to attending laboratory work, students are required to work independently in the amount of at least those hours that are indicated for each section of the program. Basically, this time is devoted to the independent solution of additional practical tasks necessary to consolidate theoretical knowledge and practical skills acquired during the course. Self-study also includes repetition of the material of the classes.

Assessment funds for course (training module)

major: Applied Mathematics and Informatics
specialization: Modern State of Artificial Intelligence/Современные методы искусственного интеллекта
“Pusk” Online and Supplementary Education Centre
Chair of Machine Learning and Digital Humanities
term: 2
qualification: Master
Semester, form of interim assessment: 3 (fall) - Exam
Author: R.G. Neychev, professor

1. Competencies formed during the process of studying the course

Code and the name of the competence	Competency indicators
Gen.Pro.C-1 Address current challenges in fundamental and applied mathematics	Gen.Pro.C-1.2 Consolidate and critically assess professional experience and research findings
Gen.Pro.C-5 An understanding of current scientific and technical problems in the field of informatics and computer technology, and is able to formulate professional tasks in scientific language	Gen.Pro.C-5.1 An understanding of the current state of research within his/her professional thematic area
	Gen.Pro.C-5.2 Able to assess the relevance of research in informatics and computer technology and its practical relevance
	Gen.Pro.C-5.3 A good command of the professional terminology used in modern scientific and technical literature, and is able to present the results of scientific work orally and in writing as part of professional communication
Gen.Pro.C-6 Capable of selecting and/or developing approaches to solving typical and new problems in informatics and computer technology, taking into account the characteristics and limitations of different solution methods	Gen.Pro.C-6.1 Able to analyse the problem, plan the solution, suggest and combine ways of solving it
	Gen.Pro.C-6.2 Capable of developing and upgrading software and hardware for information and automated systems
	Gen.Pro.C-6.3 Able to use research methods to solve new problems by applying knowledge from different fields of science (technology)
	Gen.Pro.C-6.4 Proficient in analytical and computational solution methods, and understands and takes into account in practice the limits of applicability of the solutions obtained
	Gen.Pro.C-6.5 Able to independently acquire, develop and apply mathematical, natural science, socio-economic and professional knowledge to solve non-standard problems, including in new or unfamiliar environments and in an interdisciplinary context

2. Competency assessment indicators

As a result of studying the course the student should:

know:

- WLCG project as a data source: project work stages as a data provider, resource requirements; principles of organization of hierarchical levels (Tiers) WLCG - functional differences; the concept of SLA agreements;
- the main types of Grid infrastructures;
- functional (basic) scheme of Grid computing infrastructure; modern meta-scheduler schemes (Gridway, Unicore project, Community Scheduler Framework), general functioning;
- resource management systems (RMS); types of local (LRMS) and distributed (DRMS) systems; functions of the Grid component of the WS GRAM architecture;
- types and possibilities of planners; types of policies for queuing, changing priorities, managing access lists; the principle of organization and operation of Grid SE, Grid FTP / FTS;
- the concept of the Grid task; JDL task description language;
- the main components of the Grid for the Resource Center (RC); logic diagram of the RC; central services of regional infrastructures;
- main projects and components of middleware Grid (middleware); principles of modular installation and configuration of site packages;
- principles of monitoring systems; RGMA basics; Grid infrastructure monitoring sites;
- ways of organizing security in Grid infrastructures; GSI protocol, X509 certificates, MyProxy service; virtual communities and roles, reasons for sharing computing and storage resources, VOMS, VO box; shortcomings of the existing security system model;
- using virtualization in Grid; First WNoDeS Dynamic Provisioning Project (WeNMR Project), Achievements and Disadvantages; a scheme with a virtualization resource management interface; similarities and differences between Grid and cloud computing models;
- key features of the creation and operation of the PanDA (Production and Distributed Analysis) project system; managing the workflow of jobs over datasets; major infrastructure components;
- basic requirements for cloud computing systems; basic models (IaaS, PaaS, SaaS); principles of virtualization;
- principles of organizing cloud computing based on the OpenStack system; a diagram of the work of the subsystem for managing virtual machines Nova, managing object storage Swift, images of Glance virtual machines.

be able to:

- Install and configure a combination of Torque Resource Manager and MAUI Task Scheduler. Configure three (or more) queues with different duration of tasks execution in them and binding to different values of the "acl group" and "acl user" fields;
- develop a task file and check the correctness of the configuration, task status;
- set a policy for scheduling tasks using MAUI (or other schedulers).
- execute the configuration chain for YAIM CREAM-CE and WN targets on two nodes;
- configure the resource managers Condor and SLURM;
- create your own certification authority, which includes - self-signed root CA certificate; creating a signature request; Signing the request with your own CA obtaining a signed certificate and verifying it; export to pki form for import to web browser;
- create a program to work with the OpenStack API.

master:

- The concepts of SLA agreements;
- basic skills of working with the application for organizing Grid infrastructures Globus Toolkit;
- principles of organization and work with Grid Storage Element, Grid FTP / FTS;
- the language for describing Grid tasks JDL, including the classic description and the minimum required functional fields;
- options for organizing the movement of input and output files (stage IN / stage OUT) of Grid tasks; the concept of a workflow; capabilities of task flow management systems (WMS);
- the principles of modular installation and configuration of site packages using the YAIM utility;
- the ability to work with monitoring systems for the execution of Grid tasks; the concepts of availability and reliability; the concept of performance of the base computing core HepSpec06;
- principles of operation of GSI protocols, X509 certificates, MyProxy service;
- ways to manage the workflow of tasks in PanDA;
- SAGA application programming interface;
- the principles of organizing cloud computing based on the OpenStack system.

3. List of typical control tasks used to evaluate knowledge and skills

1. What is thread?

2. What is race condition?
3. What are the main features of distributed file systems?
4. What is the difference between multiprocessing and multithreading?
5. How NoSQL is different from SQL?
6. Why do neural networks work faster on GPU?

4. Evaluation criteria

Questions for the exam

1. Prove that if m, n are two coprime integers of different parity, then the numbers $m^2 - n^2$ and $2mn$ are also coprime.
2. Write and prove the general formula for the number of different representations of a given integer n as the sum of two squares. (Representatives that are not obtained from each other by changing signs and the order of the terms are considered different.)
3. Based on the obtained formula, derive the lower bound for the maximum number of equal distances among the given n points on the plane using a regular rectangular lattice.
4. Build a regular pentagon using a compass and a ruler.
5. Build a regular 15-gon using a compass and a ruler.
6. You are given a single segment. It is required to construct using a compass and a ruler a segment of length x satisfying the equation
7. Based on the previous task, prove that a regular heptagon cannot be built using a compass and a ruler.
8. Prove that trisection of the angle is impossible.
9. Describe all possible combinations of the amounts of black and white balls in the ballot box, so that if two balls are randomly fished in a sample without returning, the probability of fishing two white balls is exactly 0.5.
10. Consider the relation on the sides a, b, c of the triangle, in which a triangle with vertices at the bases of the bisectors is isosceles. Assuming that the sides converging on side c of the large triangle are equal, reduce this relation to the following
11. In what follows, we consider the cube defined by the first of the three equations (refusing the requirement that a, b, c be sides of a triangle). Show that the resulting cube is indecomposable, that is, the polynomial that defines it does not factor.
12. In addition to this, show that our cube is nonsingular, that is, there is not a single point on its projectivization at which each direction is tangent (or the same thing at which all three first partial derivatives of the polynomial defining it degenerate).

Exam ticket examples

Ticket number 1

1. Write and prove the general formula for the number of different representations of a given integer n as the sum of two squares.
2. Prove that trisection of the angle is impossible.

Ticket number 2

1. Consider the relationship on the sides a, b, c of the triangle, in which a triangle with vertices at the bases of the bisectors is isosceles.
2. Describe all kinds of combinations of the numbers of black and white balls in the ballot box, so that if two balls are randomly fished in the sample without returning, the probability of fishing two white balls is exactly 0.5.

Assessment “excellent (10)” is given to a student who has displayed comprehensive, systematic and deep knowledge of the educational program material, has independently performed all the tasks stipulated by the program, has deeply studied the basic and additional literature recommended by the program, has been actively working in the classroom, and understands the basic scientific concepts on studied discipline, who showed creativity and scientific approach in understanding and presenting educational program material, whose answer is characterized by using rich and adequate terms, and by the consistent and logical presentation of the material;

Assessment “excellent (9)” is given to a student who has displayed comprehensive, systematic knowledge of the educational program material, has independently performed all the tasks provided by the program, has deeply mastered the basic literature and is familiar with the additional literature recommended by the program, has been actively working in the classroom, has shown the systematic nature of knowledge on discipline sufficient for further study, as well as the ability to amplify it on one’s own, whose answer is distinguished by the accuracy of the terms used, and the presentation of the material in it is consistent and logical;

Assessment “excellent (8)” is given to a student who has displayed complete knowledge of the educational program material, does not allow significant inaccuracies in his answer, has independently performed all the tasks stipulated by the program, studied the basic literature recommended by the program, worked actively in the classroom, showed systematic character of his knowledge of the discipline, which is sufficient for further study, as well as the ability to amplify it on his own;

Assessment “good (7)” is given to a student who has displayed a sufficiently complete knowledge of the educational program material, does not allow significant inaccuracies in the answer, has independently performed all the tasks provided by the program, studied the basic literature recommended by the program, worked actively in the classroom, showed systematic character of his knowledge of the discipline, which is sufficient for further study, as well as the ability to amplify it on his own;

Assessment “good (6)” is given to a student who has displayed a sufficiently complete knowledge of the educational program material, does not allow significant inaccuracies in his answer, has independently carried out the main tasks stipulated by the program, studied the basic literature recommended by the program, showed systematic character of his knowledge of the discipline, which is sufficient for further study;

Assessment “good (5)” is given to a student who has displayed knowledge of the basic educational program material in the amount necessary for further study and future work in the profession, who while not being sufficiently active in the classroom, has nevertheless independently carried out the main tasks stipulated by the program, mastered the basic literature recommended by the program, made some errors in their implementation and in his answer during the test, but has the necessary knowledge for correcting these errors by himself;

Assessment “satisfactory (4)” is given to a student who has discovered knowledge of the basic educational program material in the amount necessary for further study and future work in the profession, who while not being sufficiently active in the classroom, has nevertheless independently carried out the main tasks stipulated by the program, learned the main literature but allowed some errors in their implementation and in his answer during the test, but has the necessary knowledge for correcting these errors under the guidance of a teacher;

Assessment “satisfactory (3)” is given to a student who has displayed knowledge of the basic educational program material in the amount necessary for further study and future work in the profession, not showed activity in the classroom, independently fulfilled the main tasks envisaged by the program, but allowed errors in their implementation and in the answer during the test, but possessing necessary knowledge for elimination under the guidance of the teacher of the most essential errors;

Assessment “unsatisfactory (2)” is given to a student who showed gaps in knowledge or lack of knowledge on a significant part of the basic educational program material, who has not performed independently the main tasks demanded by the program, made fundamental errors in the fulfillment of the tasks stipulated by the program, who is not able to continue his studies or start professional activities without additional training in the discipline in question;

Assessment “unsatisfactory (1)” is given to a student when there is no answer (refusal to answer), or when the submitted answer does not correspond at all to the essence of the questions contained in the task.

5. Methodological materials defining the procedures for the assessment of knowledge, skills, abilities and/or experience

During examination the student are allowed to use the program of the discipline.