Syllabus
«Big data processing technologies»

Lecturer: *Irina Kalinina*
PhD, Associate Professor, Associate Professor of the Intelligent Information Systems Department

Expected learning outcomes
As a result of studying the discipline, the student
must know:
- methods and means of aggregation and integration of big data of various types;
- methods and means of designing storage facilities for storing big data;
- methods, means and tools of intelligent processing of big data;
- methods and means of visualization of big data;
- methods and means of obtaining knowledge from big data.

must be able to:
- use tools for integrating big data of various types;
- correctly choose and use methods and algorithms for filtering, validating and storing big data;
- correctly choose the most informative ways of visualizing big data;
- gain knowledge by analysing big data;
- design and develop applied information programs for analysing big data and making decisions based on the information obtained.

Prerequisites
"Higher mathematics", "Fundamentals of programming", "Intelligent data analysis", "Theory of probability and mathematical statistics", "Theory of algorithms and data structures", "Theory of decision-making", "Methods and systems of machine learning".

The scope of the discipline: 5 ECTS credits (16 hours of lectures, 48 hours of practice).

Purpose: theoretical and practical training of future masters in the field of information systems and technologies for working with big data. The knowledge gained as a result of mastering the discipline will help in collecting and analysing huge volumes of structured or unstructured information, in developing data models and obtaining new knowledge. All this is necessary for a graduate who has mastered the master's program to solve various tasks of practical and research activities.

The content of the discipline

Topic 1. Introduction to big data analysis. Basic concepts and definitions. Characteristics of big data. Types of big data. Application in economy, business, agriculture, industry. Examples of use. Big data in scientific fields. Features of application. Requirements for the profession of a big data analyst.
Topic 2. Business motivation and incentives for switching to big data processing. Market dynamics. Business architectures. Management of business processes. Information and communication technologies.
Topic 3. Corporate technologies and Business Intelligence for big data.
Topic 4. Concepts of big data storage. Clusters. File systems and distributed file systems. NoSQL, sharding, replication, CAP theorem, ACID, BASE.
Topic 5. Concepts of big data processing. Parallel data processing. Distributed data processing. Hadoop. Processing work tasks. Cluster. Processing in batch mode. Real-time processing.
Topic 6. Batch processing of big data using MapReduce. Examples of use.
Topic 7. Big data storage technologies. Disk storage devices. RAM storage systems. Examples of use.
Topic 8. Basic methods of big data analysis. Quantitative analysis. Qualitative analysis. Data Mining. Statistical analysis. Machine learning. Semantic analysis. Visual analysis.

Consequences
The knowledge that students will acquire while studying the course "Technology of big data processing" will be necessary for further education, for pre-diploma practice and preparation of qualification work, as well as for production activities in a professional specialty.

Technical support
Laboratory work on the discipline is conducted in computer classes using languages and software environments: R Studio, R, Python**.**

Deadline Policy
Works that are submitted late without good reason will be assigned a lower grade**.**

Academic Integrity Policy
Presupposes independent performance of laboratory work. Copying during the exam (including using mobile devices) is prohibited. If plagiarism or plagiarism is detected, the work will not be counted**.**

Topic 9. Machine learning on big data.

Evaluation criteria of laboratory works / practices / individual works / reports / projects

Maximum number of points a student with high quality independently completed the entire amount of work, answers all questions related to the completed work, and makes additional calculations offered by the teacher. The teacher has no complaints about the implementation and requirements for the performance of the work.
Approximately 70%-99% of the maximum number of points – a student completed all tasks with sufficient quality, but in the process of work he made some mistakes, which, after being pointed out by the teacher, he corrected himself. He answers some questions incorrectly. Additional calculations proposed by the teacher are done with some effort. Not all requirements for performance of work are met.
Approximately 40%-69% of the maximum number of points – a student independently completed all the work, but the quality of the implementation is insufficient (calculation errors, not all work requirements are met). The answers to questions about the performance of work are not quite clear. There are mistakes in the answers.
Approximately 1%-39% of the maximum number of points – a student did not complete all the work independently, while the quality of the implementation was insufficient (errors in calculations, does not comply with the requirements for the design of the work). He does not answer questions about the performance of work clearly. There are gross mistakes in the answers.
0 points – a student did not complete the entire amount of work, or did it with gross errors. He has problems with calculations, does not know the theoretical material, the software implementation does not meet the requirements.