Ministry of Education and Science of Ukraine
Ivan Franko National University of Lviv
Faculty of Electronics and Computer Technologies
Department of System Design

Approved
Department of System Design
Faculty of Electronics and Computer Technologies
Ivan Franko National University of Lviv
(Meeting Minutes #1 from August 28, 2023)

Head of Department _____ Roman SHUVAR

Academic Discipline Syllabus
"Cloud Computing and Big Data"
which is taught within the Education Program «Computer Science»
for the second (master) higher education level
in the Specialty 122 – Computer Science

Lviv 2023

Subjects	Cloud Computing and Big Data
The address of teaching the discipline	50 Drahomanov Street, 79005 Lviv
The faculty and The faculty and department under which the discipline is established	Faculty of Electronics and Computer Technologies, Department of System Design
Field of knowledge, code and name of speciality	12 – Information Technologies 122 – Computer Science
Teachers of the discipline	Pavlyshenko Bohdan Mykhailovych, Dr. Sci., Professor, Docent
Contact information of teachers	bohdan.pavlyshenko@lnu.edu.ua https://electronics.lnu.edu.ua/employee/pavlyshenko-bohdan-myhajlovych
Consultations on issues of training in the discipline are taking place	Consultations on the day of lectures (by prior arrangement). The address is Room 317, 50 Drahomanov Street, Lviv. Online consultations are also possible through the Moodle system. To agree on the time of online consultations, you should write to the e-mail address of the teacher.
Discipline page	https://moodle.elct.lnu.edu.ua/course/view.php?id=148 https://electronics.lnu.edu.ua/course/cloud-computing-and-big-data-122-kn
Information about the discipline	The discipline "Cloud Computing and Big Data" is an optional discipline from the Specialty 122 "Computer Science", which is taught in the 1st semester in the amount of 7.0 credits (according to the European Credit Transfer System ECTS).
A brief abstract of the discipline	The course is designed to provide participants with the knowledge to accumulate, process and analyze Big Data.
Purpose and objectives of the disciplines	*The main aim*: to acquire knowledge on the big data and cloud computing. *The secondary aims*: to be able mastering the modern technologies of big data and use them for processing in cloud environments.
Literature for studying the discipline	Basic literature: 1. Apache Hadoop documentation [Electronic resource] // Apache Hadoop. – 2021. – Resource access mode: https://hadoop.apache.org/docs/stable/. 2. Apache Spark documentation [Electronic resource] // Apache Spark. – 2019. – Resource access mode: https://spark.apache.org/docs/latest/. 3. HBase documentation [Electronic resource] // HBase. – 2019. – Resource access mode: https://hbase.apache.org/book.html. 4. RabbitMq [Electronic resource] // RabbitMq. – 2020. – Resource access mode: https://www.rabbitmq.com/documentation.html. 5. Ifeyinwa A. A. Big Data and Business Analytics: Trends, Platforms, Success Factors and Applications / A. A. Ifeyinwa, H. N. Friday. - Nigeria: Abakaliki, 2019. - 30 p. 6. Cloud Computing for Science and Engineering. - [Access mode]: https://cloud4scieng.org/chapters/
The scope of the course	The course has totally 210 hours of teaching, with 32 hours of lectures, 48 hours of practical work and 130 hours of self-education
Expected learning outcomes	After the course finishing the student should: ● to know: ○ Apache Hadoop, Pig ○ Apache Spark ○ ELK, EFK

	○ HDFS ○ RabbitMQ, Kafka, ZeroMQ ○ the basic concepts and terminology of cloud and fog technologies, areas of application of cloud technologies, basic principles of cloud computing, principles and methods of developing applications for cloud systems using various platforms, infrastructure of cloud services ● be able: ○ to have the ability for abstract thinking, analysis and synthesis; ○ to have the ability to apply knowledge in practical situations; ○ to have the ability to search, process and analyze information from various sources; ○ to have the ability to act on the basis of ethical considerations; ○ to have the ability to develop software using different programming paradigms (including parallel, object-oriented, functional programming, etc.); ○ to develop a human-machine interface; ○ to deploy and administer applications in cloud environments, evaluate the effectiveness of the application of certain cloud solutions ○ to have the skills to program interfaces and applications for the cloud and distributed systems. After studying the course, applicants will acquire the following competencies (ZK, SK) and program results (PH): ZK1. Ability to abstract thinking, analysis and synthesis. ZK2. Ability to apply knowledge in practical situations. ZK6. The ability to be critical and self-critical. ZK7. Ability to generate new ideas (creativity). SK1. Understanding the theoretical foundations of computer science. SK4. Ability to collect and analyze data (including large data) to ensure the quality of project decisions. SK6. Ability to apply existing and develop new algorithms for solving problems in the field of computer science. SK9. Ability to develop and administer databases and knowledge bases. SK11. Ability to initiate, plan and implement the development processes of information and computer systems and software, including its development, analysis, testing, system integration, implementation and support. SK13. Ability to apply methods and approaches of artificial intelligence, intellectual analysis and data science and optimization approaches to solving specific computer science problems. PH1. Have specialized conceptual knowledge that includes modern scientific achievements in the field of computer science and is the basis for original thinking and conducting research, critical thinking of problems in the field of computer science and at the border of the fields of knowledge. PH2. Have specialized computer science problem-solving skills necessary for conducting research and/or conducting innovative activities to develop new knowledge and procedures. PH3. It is clear and unambiguous to convey one's own knowledge, conclusions and arguments in the field of computer science to specialists and non-specialists, in particular to persons who are studying.

	PH7. Develop and apply mathematical methods for the analysis of information models. PH8. Develop mathematical models and data analysis methods (including large ones). PH9. Develop algorithmic and software for data analysis (including large data). PH11. Create new algorithms for solving problems in the field of computer science, evaluate their effectiveness and limitations on their application. PH12. Design and support databases and knowledge. PH15. Identify the needs of potential customers regarding the automation of information processing. PH16. Conduct research in the field of computer science. PH19. To analyze the current state and global trends in the development of computer sciences and information technologies. PH20. To have the methods and means of artificial intelligence, engineering and data analysis, pattern recognition and adaptive processing of information, analysis and processing of natural language, modeling and optimization. PH21. Create new data systems, high-level embedded systems, specialized computer systems and intelligent systems using basic knowledge of hardware and software of microcontrollers and microcomputers.
Keywords	Big data, cloud computing, fog computing
Type of training	Full-time
Course format	Conducting lectures, practical work and consultations for a better understanding of the topics
Topics	See COURSE SCHEME
Final control, form	Assessment at the end of the semester
Prerequisites	To study the discipline, knowledge of the following subjects is necessary: higher mathematics, computer science, which are related to the basic knowledge in the Brunch 12 – Information Technologies.
Teaching methods and techniques that will be used during the teaching of the course	Presentations, lectures, practical work, discussion and debate.
Necessary equipment	Multimedia, Moodle and MS Teams platforms, computer software: AWS free resources (EC2, S3, Lex, CloudFormation), Git (free), Slack (free), Facebook Messenger (free).
Evaluation criteria (separately for each type of educational activity)	Assessment is carried out throughout the semester on a 100-point scale. Points are awarded for the following types of work with the following ratios: • 12 practical work: 12x5=60 points maximally, i.e. 60% of the semester grade (the maximum number of points is 60). • control measurements (2 modules): 2x20=40 points maximally, i.e. 40% of the semester grade (the maximum number of points is 40). A total of 100 points during the semester. _____. **Control measurements are carried out in the form of test tasks. Academic integrity**: Students' labs and tests are expected to be their own original research or reasoning. Failure to cite used sources, fabrication of sources, plagiarism, and interference with the work of other students are, but are not limited to, examples of possible academic dishonesty. The detection of signs of academic dishonesty in the student's work is grounds for its non-credit by the teacher, regardless of the scale of plagiarism or deception. **Literature.** All literature that students cannot find on their own will be provided by the teacher for educational purposes only, without the right to

	transfer it to third parties. Students are also encouraged to use other literature and sources that are not among the recommended ones. **Scoring policy.** The points scored on the current test, independent work and the points on the final test are taken into account. At the same time, attendance at classes and the student's activity during practical classes must be taken into account; inadmissibility of absences and lateness to classes; using a mobile phone, tablet or other mobile devices during class for non-educational purposes; plagiarism and plagiarism; untimely performance of the assigned task, etc. Any form of breach of academic integrity will not be tolerated.
Questions of control works	The list of questions and tasks for the final assessment of knowledge of certain topics for control papers is posted on the web page.
Poll	An evaluation questionnaire for the purpose of assessing the quality of the course will be provided at the end of the course.

COURSE SCHEME

Week	Topic, plan, short theses	Form of activity (class)	Literature. Resources on the Internet	Assignment (laboratory work), hours	Deadline
1	Big data. General use. Principle of operation. Problems of Big Data. Common characteristics. The Map-Reduce paradigm.	Lecture	1, 2, 3, 4, 5	Introduction. Academic integrity. Installation and review of big data technologies.	1st week of the semester
2	Review of Hadoop, Spark and Kafka technologies. Variety, Velocity, Volume and their application.	Lecture	1	Review of Hadoop technology. Realization of a simple payment of data.	2,3 weeks of semester
3	An overview of Hadoop technology. HDFS file system. Architecture.	Lecture	1	Using Apache Pig in Apache Hadoop.	4,5 week of semester
4	Consideration of the Map Reduce system. A review of the Apache Pig platform.	Lecture	1	Using Apache Hive in Apache Hadoop.	6th week of semester
5	Apache Hive. Hive QL. MRv1 architecture. YARN scheduler.	Лекція	1	Integrating HBase with Apache Hadoop.	7th week of semester
6	Review of the Hbase database. Hbase data model. Life cycle. Architecture.	Lecture	3	Integrating Cassandra with Apache Hadoop.	8th week of semester
7	MongoDB and Neo4j. Database architecture and their use in Big Data.	Lecture	3	Development of the project with the studied technologies.	9th week of semester
8	Cassandra database disorder. Its principle of operation	Lecture	3	Review of Apache Spark technology. Realization of a simple payment of data.	10th week of semester
9	Apache Spark. Consideration. Architecture. Comparison with Hadoop technology. Principle of operation.	Lecture	2	Working with the EKL stack	11th week of semester

10	Apache Spark. Consideration. Architecture. Comparison with Hadoop technology. Principle of operation.	Lecture	2	Working with the EPK stack	12th week of semester
11	Manufacturing Operations: MLOps. Problems solved by MLOps. Applications of Artificial Intelligence in industry. ML model delivery process. AI services in the clouds. AI functions in the clouds. ML tools from Amazon, Google, IBM and Microsoft. An example of MLOps with AWS.	Lecture	5	Setting up a cloud environment. Launching an instance.	13th week of semester
12	Manufacturing Operations: DataOps. The advent of DataOps. Difficulties in production. Risky operations. Big Data: An Interaction Paradigm. Data integration over time. Data integration over time.	Lecture	5	Storage S3. Automation of data loading	14th week of semester
13	Parallelism. Parallel processing. Decomposition. Communication. Synchronization. Limitations and preliminary performance evaluation. Choice of parallel architecture. Amdahl's law. Examples of parallelism.	Lecture	4, 5	Parallelization when working with big data	15th week of semester
14	Use of cloud technologies in business. Cloud technologies and business processes. Facilitating business innovation using cloud computing. New business service. Advantages of digital business with cloud technologies.	Lecture	4, 5	Project review	16th week of semester