

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ СЕМЕНА КУЗНЕЦЯ

ЗАТВЕРДЖЕНО

на засіданні кафедри
інформаційних систем
Протокол № 1 від 22.08.2023 р.

ПОГОДЖЕНО

Проректор з навчально-методичної роботи

Каріна НЕМАШКАЛО



СИСТЕМИ БІЗНЕС-ІНТЕЛЕКТУ
робоча програма навчальної дисципліни (РПНД)

Галузь знань
Спеціальність
Освітній рівень
Освітня програма

12 "Інформаційні технології"
121 "Інженерія програмного забезпечення"
перший (бакалаврський)
"Інженерія програмного забезпечення"

Статус дисципліни
Мова викладання, навчання та оцінювання

вибіркова
англійська

Розробник:
к.е.н., доцент

підписано КЕП

Сергій ЗНАХУР

Завідувач кафедри
інформаційних систем

Дмитро БОНДАРЕНКО

Гарант програми

Олег ФРОЛОВ

Харків
2024

INTRODUCTION

The purpose of the course "Business intelligence systems" is the formation of a system of knowledge about the types of tasks that arise in the intellectual analysis of data, the ability to analyze business information in order to identify new knowledge and skills necessary for decision-making; study of basic methods and models of modern data processing; formation of practical skills for working with application program packages for solving practical tasks of analysis and interpretation of business data, consideration of practical examples of application of business intelligence systems; preparation of applicants for independent work on solving problems by means of business intelligence systems and development of intelligent systems.

The tasks of the course are:

- the ability to ensure the processes of obtaining and interpreting hidden knowledge using Pandas, Dask, Spark;
- the ability to develop software to combine data sources with the appropriate software environment;
- the ability to implement computing processes based on cloud services and technologies;
- acquisition of knowledge, practical skills and the ability to use modern methods and data processing algorithms.

Course will help graduates become effective specialists in the field of information technology, capable of implementing business intelligence systems and ensuring the competitiveness of organizations.

The subject of the course is the design and development of components of business intelligence systems.

The learning outcomes and competencies formed by the course are defined in table 1

Table 1

Learning outcomes and competencies formed by the course

Learning outcomes	Competences that must be mastered by a student of higher education
LO 9	GC6, SC7
LO 13	SC2, SC3, SC4
LO 18	GC1, GC5, GC6, SC7

where, LO09. Know and be able to use methods and tools for collecting, formulating and analyzing software requirements.

LO13. Know and apply methods of developing algorithms, designing software and data and knowledge structures.

LO18. Know and be able to apply information technologies for data processing, storage and transmission.

GC01. Ability to think abstractly, analyze and synthesize.

GC05. Ability to learn and master modern knowledge.

GC06. Ability to search, process and analyze information from various sources.

SC02. Ability to participate in the design of software, including modelling (formal description) of its structure, behaviour and processes of operation.

SC03. Ability to develop architectures, modules and components of software systems.

SC04. Ability to formulate and ensure software quality requirements in accordance with customer requirements, terms of reference and standards.

SC07. Knowledge of data information models, ability to create software for data storage, extraction and processing.

COURSE CONTENT

Content module 1. Basics of business intelligence systems

Topic 1. Introduction. Business intelligence system. Basic concepts. Classification.

Topic 2. Principles of organizing big data storage. Characteristics of big data: 5 “V” – volume, velocity, variety, value, veracity. Modern technologies and tools for storing and processing big data.

Topic 3. Basic tasks of data analysis and business intelligence. Modern technologies of data analysis and business intelligence. Architectural solutions of business intelligence systems.

Topic 4. Fundamentals of Dask and Apache Spark.

Topic 5. Peculiarities of data organization and implementation of SQL queries in Apache Spark.

Topic 6. Machine learning with Apache Spark: classification problems.

Topic 7. Machine learning with Apache Spark: prediction tasks.

Topic 8. Cloud services and solutions of business intelligence systems.

The list of laboratory studies in the course is given in table 2.

Table 2

The list of laboratory studies

Name of the topic and / or task	Content
Topic 1-3. Task 1	Data analysis based on Dask
Topic 4. Task 2	Main features of Apache Spark
Topic 5. Task 3	Development of data schemas and database queries in Apache Spark
Topic 6. Task 4	Data visualization
Topic 7. Task 5	ML Apache Spark (Classification)
Topic 7-8. Task 6	ML Apache Spark (Prediction)

The list of self-studies in the course is given in table 3.

Table 3

List of self-studies

Topic name	Content
Topic 1 - 8	Studying lecture material
Topic 1 - 8	Preparation for laboratory classes
Topic 1 - 8	Preparation for current control works
Topic 1 - 8	Preparation for the exam

The number of hours of lectures, laboratory studies and hours of self-study is given in the technological card of the course.

TEACHING METHODS

In the process of teaching the course, in order to acquire certain learning outcomes, to activate the educational process, it is envisaged to use such learning methods as:

Verbal (lecture (Topic 1-8, problem lecture (Topic 1, 8), provocative lecture (Topic 1)).

In person (demonstration (Topic 1-8)).

Practical (laboratory work (Topic 1 - 8), case method (Topic 8).

ASSESSMENT FORMS AND METHODS

The University uses a 100-point cumulative system for assessing the learning outcomes of students.

Current control is carried out during lectures and laboratory classes and is aimed at checking the level of readiness of the student to perform specific job and is evaluated by the amount of points scored:

– for courses with a form of semester control as an exam: maximum amount is 60 points; minimum amount required is 35 points.

The final control includes current control and an exam.

Semester control is carried out in the form of a semester exam.

The maximum number of points that a student of higher education can receive during the examination (examination) is 40 points. The minimum amount for which the exam is considered passed is 25 points.

The final grade in the course is determined summing up points for current and final control.

During the teaching of the course, the following control measures are used:

Current control: defense of laboratory work (50 points), written control work (10 points).

More detailed information on the assessment system is provided in technological card of the course.

An example of an exam card and assessment criteria.

Simon Kuznets Kharkiv National University of Economics
First (bachelor) study cycle
Specialty "Software engineering"
Study program "Software engineering"
Semester V
Course "Business intelligence systems"

EXAM TICKET 1

Task 1 (diagnostic):

Provide an example of implementing a request for calculating the average value and grouping data by category using DASK

Task 2 (diagnostic)

Describe the principles of organizing storage of big data

Task 3. (heuristic,

1. Create a csv file with fields name, age, salary (at least 7 records, example record: Alex, 45, 2100) and calculate the average age for persons with the corresponding name using PySpark
2. Implement the regression model of the dependence of salary on age using ML PySpark and justify the adequacy of the model.

Add screenshots of filled data, program code and result to the report.

Approved at the meeting of the Department of Information Systems № ____ of
"____" _____ 20____.

Examiner

Serhii ZNAKHUR

Head of Department, Phd

Dmytro BONDARENKO

Evaluation criteria

Final scores for the exam consist of the sum of points for each task.

The algorithm for solving each task includes separate stages that differ in complexity, time-consuming, and importance for solving the task. Therefore, individual tasks and stages of their solution are evaluated separately from each other as follows:

Task 1 (diagnostic, 8 points)

This task is evaluated from 0 to 8 points according to the following criteria:

8 points. The theoretical question is described in full, with conclusions. Conclusions to the theoretical question are reasoned and substantiated.

7 points. The theoretical question is described in full, the material is presented in accordance with the program material of the discipline. When performing the comparison, in-depth knowledge of the material provided by the curriculum is used. However, the acquirer assumes certain inaccuracies.

6 points. The theoretical question is fully disclosed, the program material is presented in accordance with the material of the discipline. When completing the task, the student applies the basic knowledge of the educational material provided by the curriculum. The comparison is performed correctly in general, while performing the acquisition, some minor errors are assumed.

5 points. The theoretical question is fully disclosed, but minor mistakes were made in the presentation of the material. The student applies the basic knowledge of the educational material provided by the curriculum. When performing it, the acquirer makes minor mistakes.

4 points. The theoretical question is incompletely disclosed, with significant errors. When performing a theoretical task, the student applies educational material without sufficient understanding, makes a significant number of mistakes, faces significant difficulties in analyzing and interpreting situations.

3 points. The theoretical question is incompletely disclosed or not disclosed at all. When completing the task, the applicant makes quite a large number of gross mistakes, faces significant difficulties in analyzing and interpreting situations, and shows the ability to express an opinion at an elementary level.

2 points. The applicant cannot present a theoretical question, faces significant difficulties in analyzing and interpreting the situation, shows the ability to present an opinion at an elementary level. The theoretical question is not disclosed.

1 point The applicant cannot present a theoretical question, he faces significant difficulties in analyzing and interpreting the situation. The theoretical question is not disclosed.

0 points are given for an uncompleted task at all.

Task 2 . (diagnostic, 8 points)

This task is evaluated from 0 to 8 points according to the following criteria:

8 points. The theoretical question is described in full, with a presentation of the conclusions obtained on the basis of the program, additional material. Conclusions to the theoretical question are reasoned and substantiated.

7 points. The theoretical question is described in full, the material is presented in accordance with the program material of the discipline. When performing the comparison, in-depth knowledge of the material provided by the curriculum is used. However, the acquirer assumes certain inaccuracies.

6 points. The theoretical question is fully disclosed, the program material is presented in accordance with the material of the discipline. When completing the task, the student applies the basic knowledge of the educational material provided by the curriculum. The comparison is performed correctly in general, while performing the acquisition, some minor errors are assumed.

5 points. The theoretical question is fully disclosed, but minor mistakes were made in the presentation of the material. The student applies the basic knowledge of the educational material provided by the curriculum. When performing it, the acquirer makes minor mistakes.

4 points. The theoretical question is incompletely disclosed, with significant errors. When performing a theoretical task, the student applies educational material without sufficient understanding, makes a significant number of mistakes, faces significant difficulties in analyzing and interpreting situations.

3 points. The theoretical question is incompletely disclosed or not disclosed at all. When completing the task, the applicant makes quite a large number of gross mistakes, faces significant difficulties in analyzing and interpreting situations, and shows the ability to express an opinion at an elementary level.

2 points. The applicant cannot present a theoretical question, faces significant difficulties in analyzing and interpreting the situation, shows the ability to present an opinion at an elementary level. The theoretical question is not disclosed.

1 point The applicant cannot present a theoretical question, he faces significant difficulties in analyzing and interpreting the situation. The theoretical question is not disclosed.

0 points are given for an uncompleted task at all.

Task 3. (heuristic, 24 points).

This task is evaluated on a 24-point scale.

The first task is evaluated from 0 to 12 points according to the following criteria:

12 points . The task was completed in full accordance with the individual task.

0 points. Task not completed

In the event that the task is completed in full, but mistakes were made during its execution, a part of the points proportional to the completed in the examination paper is deducted from the maximum score for the task, namely:

for each insignificant error, up to 0.5 points are deducted, but no more than 1.5 points for each group of homogeneous insignificant errors;

up to 2 points are deducted for each group of homogeneous significant errors (for example, lack of sufficient data).

The second task is evaluated from 0 to 12 points according to the following criteria:

12 points . The task was completed in full accordance with the individual task.

0 points. Task not completed

In the event that the task is completed in full, but mistakes were made during its execution, a part of the points proportional to the completed in the examination paper is deducted from the maximum score for the task, namely:

for each insignificant error, up to 0.5 points are deducted, but no more than 1.5 points for each group of homogeneous insignificant errors;

up to 2 points are deducted for each group of homogeneous significant errors (for example, the absence of model metrics).

RECOMMENDED LITERATURE

Main

1. Голубничий Д. Ю. Багатоетапні алгоритми рішення задачі оптимізації структури сховища даних у вузлах мережі хмарного середовища / Д.Ю. Голубничий, О.В. Коломійцев, В.Ф. Третяк та ін. // Scientific Collection «InterConf», 1(37): with the Proceedings of the 1st International Scientific and Practical Conference «Recent Scientific Investigation» (December 6-8, 2020). – Oslo, Norway: Dagens naeringsliv forlag, 2020. – p. 1071–1078. <http://repository.hneu.edu.ua/handle/123456789/24617>

2. Data Science для бізнесу. Як збирати, аналізувати і використовувати дані / Фостер Провост, Том Фоусет. Видавництво: "Наш Формат" 2019, 400 с.

3. Benjamin S. Baumer, Daniel T. Kaplan , Nicholas J. Horton. Modern Data Science. 2nd Edition. Chapman and Hall/CRC, 2021. 632 pp.

4. Hadley Wickham, Mine Çetinkaya-Rundel, Garrett Golemund. Data Science: Import, Tidy, Transform, Visualize, and Model. O'Reilly Media, Inc., 2023. 576 pp.

5. Parallel Programming Technologies : work program for students of the first (bachelor's) level the specialty 121 Software Engineering / compilers S. G. Udovenko, O. O. Tiutiunyk, I. A Ushakova, N. H. Aksak. – Kh.: Publisher S. Kuznets KhNUE, 2020. – 13 p. (English language) <http://repository.hneu.edu.ua/handle/123456789/26517>

Additional

6. Литвин В.В., Інтелектуальні системи: Підручник / В.В. Литвин, В.В. Пасічник, Ю.В. Яцишин. – Львів: “Новий Світ – 2000”, 2020 – 406 с

Information resources

7. Apache Spark (a project managed by the Apache Spark Committee). URL: <https://projects.apache.org/project.html?spark>

8. KDNuggets: Data Mining Community Top Resource for Analytics, Data Mining, and Data Science Software, Companies, Data, Jobs, Education, News, and more. URL: <http://www.kdnuggets.com>

9. Kaggle URL: <http://www.kaggle.com>

10. The Data Mine URL: <http://www.the-data-mine.com>

11. Statista. Big data analytics market revenue worldwide in 2019 and 2025. URL: <https://www.statista.com>

12. Gartner. 12 Data and Analytics Trends to Keep on Your Radar. URL: <https://www.gartner.com>