**ЗАТВЕРДЖЕНО**
на засіданні кафедри
інформаційних систем
Протокол № 1 від 27.08.2024 р.

**ПОГОДЖЕНО**
Проректор з навчально-методичної
роботи

Каріна НЕМАШКАЛО

# УПРАВЛІНСЬКІ ІС ТА СХОВИЩА ДАНИХ

### робоча програма навчальної дисципліни (РПНД)

| | |
|---|---|
| Галузь знань | **12 "Інформаційні технології"** |
| Спеціальність | **126 "Інформаційні системи та технології"** |
| Освітній рівень | **другий (магістерський)** |
| Освітня програма | **"Інформаційні системи та технології"** |

| | |
|---|---|
| Статус дисципліни | **обов'язкова** |
| Мова викладання, навчання та оцінювання | **англійська** |

Розробник:
к.е.н., доцент                                    Сергій ЗНАХУР

Завідувач кафедри
інформаційних систем                    Дмитро БОНДАРЕНКО

Гарант програми            підписано КЕП            Олександр КОЛГАТІН

**Харків**
**2024**

**APPROVED**

at the meeting of the informational systems
department.
Protocol № 1 of 27.08.2024

**AGREED**

Vice-rector for educational and
methodological work

Karina NEMASHKALO

# MANAGEMENT INFORMATION SYSTEMS AND DATA WAREHOUSE

## Program of the course

| | |
|---|---|
| Field of knowledge | **12 "Information technologies"** |
| Specialty | **126 "Information systems and technologies"** |
| Study cycle | **second (master's)** |
| Study programme | **"Information Systems and Technologies"** |

| | |
|---|---|
| Course status | **mandatory** |
| Language | **english** |

Developers:
PhD in Economics,
Associate Professor     _____      Serhii ZNAKHUR

Head of the Department
information systems     _____      Dmytro BONDARENKO

Head of Study Programme      digital signature      Oleksandr KOLGATIN

Kharkiv
2024

# INTRODUCTION

The course "Management information systems and Data Warehouse" belongs to the group of educational and professional courses for masters in speciality 126 "Information Systems and Technologies". The course "Management information systems and Data Warehouse" is a basic course and is studied in accordance with the curriculum for the second (master's) level of training in specialty 126 "Information Systems and Technologies".

**Objective of the course:** the purpose of teaching the course "Management information systems and Data Warehouse" is to acquaint masters with existing methodological approaches and technological means of developing Data Warehouse s and analytical information systems BI (business intelligence), to study methods of building and maintaining such systems. The course covers the principles of building data analytics-oriented systems, various data models used in Data Warehouse and data lakes. The course also covers the issues of building BI systems, using data mining technologies, visualising results and other issues. Pandas, Postgresql, Mysql, Big Query, and PowerBI are studied as tools in the course. In the PowerBI environment, students learn how to develop information and analytical systems, integrate data from various sources into a Data Warehouse and data lakes. Mastering such knowledge will allow you to implement the tasks of automating information processing, automating facility management, and creating decision support systems.

**The object** is the process of building analytical information systems based on modern data processing and data storage technologies.

**The subject of** the course is technologies, principles and methods of data mining, data visualisation based on the use of Data Warehouses and lakes.

The programme of study includes lectures and laboratory work.

For the practical mastering of the main topics of the course, laboratory work is carried out using computers, local area networks and the Internet in the computer classes of HNUE named after S. Kuznets.

The material taught in this course is used by master's students to write their thesis and in the study of courses in the master's programme.

The learning outcomes and competences that the course forms are defined in table 1.

Table 1

### Learning outcomes and competencies formed by the course

| Learning Outcomes | Competencies |
|---|---|
| LO07 | GC01,GC04, GC05,SC01,SC02,SC03 |

| LO09 | IC, SC04, SC05 |
|------|----------------|
| LO12 | SC05 |

LO07. Making a grounded choice of project solutions and design a service-oriented information architecture of the enterprise (institution, organisation, etc.).

LO09. Developing and use data warehouses, to perform data analysis for supporting decision-making.

LO12. Improving the information system on the base of business processes analysis.

GC01. Ability to abstract thinking, analysis and synthesis.

GC04. Ability to develop and manage projects.

GC05. Ability to evaluate and provide the quality of the work performed.

SC01. Ability to develop and apply IST necessary for solving strategic and current tasks.

SC02. Ability to formulate requirements for life cycle stages of service-oriented information systems.

SC03. Ability to design information systems taking into account the specifics of their purpose, incomplete/insufficient information and conflicting requirements.

SC04. The ability to develop mathematical, information and computer models of objects and informatization processes.

SC05. Ability to use modern data analysis technologies to optimize processes in information systems.

IC. The ability to solve problems of a research and innovation nature in the field of information systems and technologies.

# COURSE CONTENT

**Content module 1: Modern management IS and BI**

**Topic 1: Introduction to the course**

Introduction to the course. The purpose and objectives of the course, its place in the educational process. Structure of the course, recommendations for its study. Organisational and methodological support of the course. Properties of 5V data. The difference between data analytics and AI. Strategy for building a data infrastructure. The data pipeline. Basic concepts of BI. Application of intelligent systems in BI. Basic standards for designing BI systems. General requirements and functional responsibilities of Data Engineer. General requirements and functional responsibilities of Data Scientist.

**Topic 2.** Data management systems and technologies

Data storage in a modern data architecture on the example of AWS. Types of data. Data storage layer in modern architectures. Comparison of data lakes and Data Warehouse s. Amazon data storage service (Amazon S3). Formation of an AWS data lake. Choosing a database. Factors for choosing a database. Common cases of using databases. Classical relational databases. Disadvantages of relational databases.

## Topic 3. PowerBI system

Main provisions and history of PowerBI. The architecture of the PowerBI system. Access to data in Power BI. Using gateway for data access security. Organisation of work with data. Data formats. Importing data. Preliminary data transformation. Organising entity-relationships. Building dashboards and reports. Using Data Analysis Expressions (DAX). Building dimensions for OLAP reports. Using aggregates based on DAX. Building reports.

### Content module 2. BI Systems

## Topic 4. Data Warehouse and BI

Data Warehouse. Data Warehouse and practical cases. The structure of the Data Warehouse. Commercial Data Warehouse solutions. Open-source Data Warehouse solutions. Basic concepts of ETL and ELT. ETL and ELT pipelines. Comparison of Data Warehouse architectures for ETL and ELT pipelines. Data Warehouse in Amazon Redshift. Amazon Redshift parallel data processing architecture. Automation and scaling in Amazon Redshift. Examples of building a Data Warehouse in Amazon Redshift. Relational and non-relational databases. Amazon RDS. AWS Aurora. Amazon DynamoDB. Features of integrating data from relational and non-relational databases into a Data Warehouse. Data integration with the AWS QuickSight BI service.

## Topic 5. OLAP systems

OLTP and OLAP. Comparison of system characteristics of OLTP and OLAP. Codex rules for OLAP systems. Multidimensional DBMS. Data organisation in multidimensional DBMS. Multidimensional databases. FASMI test. Hypercube. Operations with data in a hypercube. Technical aspects of multidimensional data

storage. Features of OLAP solutions. Disadvantages of OLAP solutions. Architectures of OLAP systems.

**Topic 6: BigQuery analytical system**

Modern cloud services for analytics and data processing (AWS, AZURE, Google Platform). Cloud technology platforms and services. Categories of Google Cloud Platform services. Google Cloud Platform data storage services. BigData service BigQuery. BigQuery architecture. Uploading data to BigQuery. Organising queries in BigQuery. BigQuery ML (BigQuery machine learning). Building reports based on BigQuery and Google Cloud Platform services.

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. Lab 1 | Data analysis basics with the Pandas library |
| Topic 2. Lab 2 | Data visualisation based on Pandas, Matplotlib, Seaborn libraries |
| Topic 3. Lab 3 | Analytical data processing and visualisation in PowerBI |
| Topic 4. Laboratory work 4 | Using Data Mining in PowerBI |
| Topic 5. Lab 5 | Multivariate analysis in PowerBI |
| Topic 6. Laboratory work 6 | Using BigQuery to store and analyse data |

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

Table 3

**List of self-studies**

| Name of the topic and/or task | Content |
|---|---|
| Topic 1: Task 1. | Researching modern strategies for building data infrastructure |
| Topic 2. Task 2. | Analysis of AWS data storage and processing services |

| Topic 3. Task 3. | Research of modern BI tools |
|---|---|
| Topic 4. Task 4. | Analysing the features of ETL and ELT pipelines |
| Topic 5. Task 5. | ROLAP and MOLAP studies |
| Topic 6. Task 6. | Exploring Google Cloud Platform services |

The number of hours of lectures, practical (seminar) 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 teaching methods as:
Problem lecture (Topics 1, 2, 4, 5), mini-lecture and discussion (Topics 3, 6).
Visual (demonstration (Topics 1-6)).
Individual laboratory work (Topics 1 - 6)

# FORMS AND METHODS OF ASSESSMENT

The University uses a 100-point cumulative system for assessing the learning outcomes of students.
**Current control** is carried out during lectures, practical, laboratory and seminar classes and is aimed at checking the level of readiness of the student to perform a specific job and is evaluated by the amount of points scored:
− for courses with a form of semester control as grading: maximum amount is 100 points; minimum amount required is 60 points.
**The final control** includes current control and assessment of the student .
*The final grade in the course* is determined:
− for courses with a form of grading, the final grade is the amount of all points received during the current control.
During the teaching of the course, the following control measures are used:
Current control: proving the laboratory work (70 points); current assessment (30 points).
Semester control: Grading.
More detailed information on the assessment system is provided in technological card of the course.

# RECOMMENDED LITERATURE

## Main

1.Foreman, John W. Data Smart: Using Data Science to Transform Information into Insight / John W. Foreman; DB books24x7. - John Wiley & Sons, 2014. - 432 pages. - ISBN 978-1-118-03496-5: http://common.books24x7.com/toc.aspx?bookid=58144.

2.Davy S. Fundamentals of Data Science and BigData. Python and Data Science." S.Davy, M.Arno, A.Mohamed - St. Petersburg: Peter, 2017. - 336 p.

3.Intelligent data analysis: Textbook / Cherniak O.I., Zakharchenko P.V. / K.: Znannya, 2014. 599 p.

## Additional

1.        Znakhur S.V. Integration of PowerBI and Data Mining tools / S.V. Znakhur // Proceedings of the International Scientific and Practical Conference "Economic Development and Heritage of Semen Kuznets" - 2019 - Kh.: KhNEU – P. 325-326. http://repository.hneu.edu.ua/bitstream/123456789/27384/1/%D0%92%D1%8B%D0%BF%D1%83%D1%81%D0%BA%2096.pdf

2.        Zavadskyi I.O. Fundamentals of databases: [Study guide] / I.O. Zavadskyi - Kyiv: I.O. Zavadskyi Publisher, 2011. - 192 p.

## Information resources

1.POWERBI [Electronic resource]: https://www.microsoft.com/en-us/power-platform/products/power-bi.

2. Databases and information systems. Training course [Electronic resource]. - Access mode: http://simulation.kiev.ua/dbis/lection06.htm.

3.PANDAS [Electronic resource]. - Access mode: https://pandas.pydata.org/.

4. AWS [Electronic resource]: https://aws.amazon.com/about-aws/?nc2=h_header.

4. Website of personal learning systems of KhNUE named after S. Kuznets. course "Management IS and Data Warehouses." - Access mode: https://pns.hneu.edu.ua/course/view.php?id=5290.