MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
ODESSA I.I. Mechnikov NATIONAL UNIVERSITY
Department of mathematical support of computer systems

"APPROVED"
Vice-rector for scientific and pedagogical work

"_____" _____ 20___

WORKING PROGRAM OF EDUCATIONAL COURSE

OK7 "On-line analytical processing systems"
(course name)

Level of higher education Second (master's)

Field of knowledge 12 – Information technologies

Specialty 126 – Information systems and technologies
(code and name of specialty)

Educational and professional program Information systems and technologies
(EPP/ESP name)

ONU - 2022

The working program of the educational course "On-line analytical processing systems". – Odesa: ONU, 2022. – 10 p.

Developers:
Malakhov E.V., Sc.D. (Tech.), professor, head of department of MSCS

The working program was approved at the meeting of the Department of Mathematical Support of Computer Systems

Protocol No. 1 from "25" 08 2022 year

Head of the department _____ (signature) (Eugene MALAKHOV) (First Name Surname)

Agreed with the guarantor of the OPP "Information systems and technologies"

_____ (signature) (Eugene MALAKHOV) (First Name Surname)

Approved by the educational and methodical commission (EMC) for IT specialties of the FMPhIT

Protocol No. 1 from "31" 08 2022 year

Head of EMC _____ (signature) (Alla RACHYNSKA) (First Name Surname)

Reviewed and approved at the meeting of the department _____

Protocol No. 1 from "29" 08 2023 year

Head of Department _____ (signature) (_____) (First Name Surname)

Reviewed and approved at the meeting of the department _____

Protocol No. ___ from "____"_____20__ year

Head of Department _____ (signature) (_____) (First Name Surname)

1. Course description

<table>
<tr><td rowspan="2">Name of indicators</td><td rowspan="2">Field of knowledge, specialty, specialization, level of higher education</td><td colspan="2">Course characteristics</td></tr>
<tr><td>full-time education</td><td>external form of education</td></tr>
<tr><td rowspan="15">The total number of:

credits - 3.5

hours - 105

content modules – 5</td><td rowspan="15">Branch of knowledge
<u>12 - Information technologies</u>

Specialty
<u>126 – Information systems and technologies</u>

Level of higher education:
<u>Second (master's)</u></td><td colspan="2">Mandatory</td></tr>
<tr><td colspan="2">Year of training:</td></tr>
<tr><td>1st</td><td></td></tr>
<tr><td colspan="2">Semester</td></tr>
<tr><td>2nd</td><td></td></tr>
<tr><td colspan="2">Lectures</td></tr>
<tr><td>14 hours</td><td>6 hours</td></tr>
<tr><td colspan="2">Practical, seminar</td></tr>
<tr><td>hours</td><td>hours</td></tr>
<tr><td colspan="2">Laboratory</td></tr>
<tr><td>16 hours</td><td>6 hours</td></tr>
<tr><td colspan="2">Independent work</td></tr>
<tr><td>75 hours</td><td>93 hours</td></tr>
<tr><td>including IERT*: 30 hours</td><td></td></tr>
<tr><td colspan="2">Final control form: test</td></tr>
</table>

* - in the presence

2. The purpose and tasks of the educational course

The purpose is the study of basic concepts and principles of planning, designing, building, filling and maintaining data warehouses.

Task:

– familiarization with the basics of the theory of information repositories;
– study of means of building multidimensional data structures;
– learning the MDX language;
– preparation for the implementation of diploma projects and qualification works, the subject of which is related to the solution of analytical and forecasting problems.

The process of studying the course is aimed at forming elements of the following **competencies**:

a) general (GC): —

c) special professional (SC):

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

SC05. The ability to utilize modern data analysis technologies for optimizing processes in information systems.

SCM04. The ability to apply modern models and methods of fuzzy logic inference based on knowledge representation forms and decision-making organization methods; design and develop expert systems.

SCM07. The ability to conduct information analysis and create multi-dimensional models of subject areas.

Program learning outcomes:

LO08. Develop models of information processes and systems of different classes, using methods of modeling, formalization, algorithmization, and model implementation using modern computer tools.

LO09. Develop and use data repositories, and perform data analysis to support decision-making.

LOM04. Utilize modern fuzzy models, methods, and artificial intelligence tools in decision-making systems, and apply intelligent algorithms using fuzzy models to solve artificial intelligence problems.

LOM08. Create optimized pipelines for data preparation for subsequent storage and processing.

Expected learning outcomes. As a result of studying the academic course, the student should

know: *basic concepts and terminology of data warehouse technology, technologies and areas of application of OLAP tools and data mining methods (Data Mining), the role of data warehouses in decision-making automation systems.*

be able: *design data stores and data warehouses,* apply modern tools for building data warehouses and analytical data processing.

3. Course content

Content module 1. Concept of OLTP and OLAP systems.

Topic 1. Concept of OLTP and OLAP systems. Evolution of data warehouses (DWH). DWH concept. Definition of DWH. Advantages of DWH technology. Comparison of OLTP and DWH. Problems of development and maintenance of DWH.
Literature: [1, 2, 10, 12].

Content module 2.DWH architecture.

Topic 1. DWH architecture. The main components of DWH and their purpose. Information flows in DWH.
Literature: [1, 2, 5, 12].

Topic 2. DWH tools and technologies. DBMS for DWH and features of its architecture. Using metadata in DWH. Management and administration tools.
Literature: [1, 5, 8, 9, 13].

Topic 3. Data windows: general characteristics and connection to DWH. Architecture of data stores. Data store support issues.
Literature: [1, 5, 8, 9, 13].

Content module 3.DWH design.

Topic 1. DWH design tasks. Modeling measurements. Schemes "star", "snowflake", "star-snowflake". Advantages of these schemes. Comparison of DM and ER models.
Literature: [1, 3, 8, 9].

Topic 2. Nine-step database design methodology for DWH. Essence, description and characteristics of all stages. Implementation of various schemes. Evaluation criteria for DWH measurements.
Literature: [1, 5, 8, 9, 13].

Content module 4.OLAP and data mining.

Topic 1. OLAP and data mining. Concept of OLAP system, its purpose and functions. Benchmark testing of OLAP tools. Areas of application of OLAP. Basic requirements for OLAP systems. Advantages of OLAP systems. Options for presenting multidimensional data. Basic analytical operations.
Literature: [1, 2, 7, 12].

Topic 2. OLAP tools. Codd's rules for evaluating OLAP systems.
Literature: [1, 5, 7, 11, 12].

Topic 3. SQL extension to support OLAP. Operations DECODE, CUBE, MOVINGAVG, MOVINGSUM, RANK...WHEN, RATIONTOREPORT, TERTILE, CREATE MACRO. Examples.
Literature: [1, 3, 6, 7].

Topic 4. Basic concepts of data mining technology. Examples of applications. Data development methods. Data mining tools.
Literature: [1, 7, 13].

Content module 5.DWH generation and processing using MS Power BI or MS Excel pivot tables.

IERT*(independent work)*

4. Course structure

Names of content modules and topics	Number of hours									
	Full-time					Correspondence form				
	That's all	including				That's all	including			
		l	p	lab	W ed		l	p	lab	Wed
1	2	3	4	5	6	7	8	9	10	11
Content module 1. Concept of OLTP and OLAP systems.										
Topic 1.	5	1			4	7	1		1	5
Content module 2. DWH architecture.										
Topic 1.	5	1		4	3	8	1		1	8
Topic 2.	6	1			4	11				10
Topic 3.	7	1			4	11				10
Content module 3.DWH design.										
Topic 1.	9	2		6	4	12	2		2	10
Topic 2.	11	2			6	12				10
Content module 4. OLAP and data mining.										
Topic 1.	7	2		6	4	11	2		2	10
Topic 2.	7	2			4	11				10
Topic 3.	9	1			6	11				10
Topic 4.	9	1			6	11				10
Total hours without ISRT	**75**	**14**		**16**	**45**		**6**		**6**	**93**
Content module 5. DWH generation and processing using MS Power BI or MS Excel pivot tables.										
IERT*	30	–	–	–	30	–	–	–	–	–
Total hours	**105**	**14**		**16**	**75**	**105**	**6**		**6**	**93**

* - in the presence

5. Topics of seminar classes

Seminar classes are not provided

6. Topics of practical classes

Practical classes are not provided

7. Topics of laboratory classes

No s/p	Topic name	Number hours
1	Designing a showcase of the data of a separate business process in a specific application area	4
2	Data storage model design.	6
3	Implementation of the project by means of MS SQL Analytical Services	6
	Total hours	**16**

8. Independent work

No s/p	Title of the topic / types of tasks	Number hours
1	Concept of OLTP and OLAP systems.[1]	4
2	DWH architecture. Information flows in DWH.[1]	4
3	DWH tools and technologies.[1]	4
4	Data windows.[1]	3
5	DWH design.[1]	10
6	OLAP and data mining.[1]	6
7	OLAP tools. Codd's rules for evaluating OLAP systems.[1]	4
8	The MDX language.[1]	10
	Total hours (without IERT)	**45**

Independent work includes:
[1] – preparation for lectures, practical, seminar, laboratory classes;

8.1. Individual educational and research task

An individual task is to design and create a data warehouse for the analysis of a specific subject area.

The student must:

– give a description of the subject area;
– give a list of analytical problems that can or should be solved in this subject area;
– determine the structure of the data warehouse and create it using SSAS;
– implement the solution of the specified analytical problems and export the solution results to MS Excel summary tables.

Note: Correspondence students do not complete IERT.

9. Teaching methods

Lectures using multimedia presentation material.

10. Control methods

During the defense of an individual assignment, the student must:

– explain the created data storage structure;
– demonstrate and explain the solution of the specified analytical problems.

During the final control, the student must answer 2 questions of the examiner from the list given in point 12.

10.1. Evaluation criteria at the final inspection:

1. The answer should be complete and short. It should not contain material that does not relate to the essence of the question.
2. Clearly formulate statements, skillfully apply the necessary formulas and knowledge of the main issues of the program.

3. Answers with false statements are evaluated based on the closeness of the answer to the correct one.
4. Omissions in the justification of statements are taken into account and this leads to a decrease in the number of points.
5. Small flaws, inaccuracies in the presentation of the material, reduce the number of points.
6. Ignorance and misunderstanding of the main idea of a theoretical question or problem leads to the withdrawal of up to 90% of points.
7. If there is no answer to the question, zero points are assigned.

11. Questions for final control

1. The concept and architecture of the decision support system.
2. Concepts: data storage, data showcase, operational databases.
3. Concept and model of OLAP data. OLAP tools.
4. Analytical data processing methods.
5. Categories of data in SD. Information flows in SD.
6. OLAP cube structure. Fact table. Tables of measurements. Projecting measurements. Deployment of the project. Attribute links. Creating custom hierarchies. Connections of dimensions. Definition of calculations.
7. Operations performed on a hypercube
8. Architecture of OLAP systems. Data mining, transformation and loading layer. Data storage layer. Data analysis layer.
9. Client OLAP tools. Server OLAP tools.
10. Technical aspects of multidimensional data storage: MOLAP, ROLAP, HOLAP.
11. SSAS capabilities.
12. Integration Services capabilities for working with OLAP. Merging data from disparate data stores. Populating data warehouses and data showcases. Cleaning and standardization of data.
13. Architecture of SSIS services. SSIS package.
14. Introduction to MDX. Scope and main purpose. Basic language terms. Description of the axes of the resulting set. MDX functions and expressions.
15. MDX. Navigation in the dimension tree.
16. MDX. Using external libraries.
17. String conversion to MDX types.
18. MDX. User-defined functions. Arrangement based on aggregates. Calculation of moving average.
19. MDX. Operations on sets. Named plurals.

12. Distribution of points received by students

Content module #1	Content module No. 2			Content module #3		Content module No. 4				Content module No. 5	
T1	T1	T2	T3	T1	T2	T1	T2	T3	T4	T1	
8	8	8	10	8	10	8	8	10	8	14	100

T1, T2 ... - topics of content modules, KR - course work, IERT - individual educational and research task

Evaluation scale: national and ECTS

The sum of points for all types of educational activities	ECTS assessment	Evaluation on a national scale	
		for an exam, course project (work), practice	for credit
90-100	A	perfectly	counted
85-89	B	fine	
75-84	C		
70-74	D	satisfactorily	
60-69	E		
35-59	FX	unsatisfactory with the possibility of reassembly	not counted with the possibility of retaking
0-34	F	unsatisfactory with mandatory repeated study of the course	not enrolled with mandatory repeated study of the course

13. Educational- methodological support

Synopsis of lectures; a complex of educational and methodological support of the course; regulations; presentation materials.

14. Recommended Books

14.1. Basic literature

1. Reis J., Housley M., Fundamentals of Data Engineering. Plan and Build Robust Data Systems. - O'Reilly Media, 2022. - 446 p.
2. Inmon WH, Krishnan K., Building the Unstructured Data Warehouse: Architecture, Analysis, and Design. - Technics Publications, 2010. - 216 p.
3. Codd EF, Codd SB, Salley CT Providing OLAP (On-Line Analytical Processing) to User-Analysts: An IT Mandate. - EF Codd & Associates, 1993.
4. Jukic N., Database Systems: Introduction to Databases and Data Warehouses. - Prospect Press, 2016. - 400 p.
5. Wrembel R., Koncilia C., Data Warehouses and OLAP: Concepts, Architectures and Solutions. - IGI Global, 2006. - 332 p.
6. Kimball R., Ross M., Becker B., Mundy J., Thornthwaite W., The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence Remastered Collection. 2nd Edition. - Wiley, 2015. - 912 p.
7. Rainardi V., Building a Data Warehouse: With Examples in SQL Server (Expert's Voice). - Apress, 2008. - 539 p.
8. Bhatia P., Data Mining and Data Warehousing: Principles and Practical Techniques. - Cambridge University Press, 2019. - 506 p.
9. Gorelik A., The Enterprise Big Data Lake: Delivering the Promise of Big Data and Data Science. - O'Reilly Media, 2019. - 224 p.

10. Corr L., Agile Data Warehouse Design: Collaborative Dimensional Modeling, from Whiteboard to Star Schema. - DecisionOne Press, 2011 - 328 p.
11. Crickard P., Data Engineering with Python. - Birmingham: Packt Publishing, 2020. - 357 p.

14.2. Supporting literature

12. Spirly E. Corporate data warehouses. Planning, development, realization. Volume 1. - M.: Izdatelsky dom "Williams", 2001.
13. Larson B. Development of business analytics in Microsoft SQL Server 2005 / St. Petersburg: Peter, 2008 - 684 p.
14. Barsegyan A.A., Kupriyanov M.S., Stepanenko V.V., Kholod I.I. Methods and models of data analysis: OLAP and Data Mining - St. Petersburg: BHV-Petersburg, 2004. - 336 p.: ill.
15. Connolly T., Begg K. Databases. Design, implementation and support. Theory and practice. 3rd edition. - M.: "Williams" Publishing House, 2003.
16. Poluboyarov V.V. Using MS SQL Server Analysis Services 2008 to build data warehouses. [Electronic resource] / NNTU, 2008. – Access method: http://www.intuit.ru/department/database/mssqlsas2008/. - 488 p.
17. Nekrasov V. 30 ideas of OLAP application. [Electronic resource] / Intersoft Lab. – Access method: http://www.iso.ru/rus/document5808.phtml.

15. Electronic information resources

1. Analysis Services | Microsoft SQL Server 2012 - Access mode: http://www.microsoft.com/en-us/sqlserver/solutions-technologies/business-intelligence/analysis.aspx
2. AdventureWorks Databases – 2019, 2016, 2012, 2008R2 and 2008 – Access Mode: http://msftdbprodsamples.codeplex.com/releases/view/93587