

INTRODUCTION TO DATABASES

Curriculum (Syllabus)

Course details

Level of higher education	<i>Second (Master's)</i>
Field of knowledge	<i>05 Social and behavioural sciences</i>
Specialisation	<i>054 Sociology</i>
Educational programme	<i>Social Data Analytics</i>
Status of discipline	<i>Elective</i>
Form of study	<i>Full-time (day)</i>
Year of study, semester	<i>1st year of master's programme, spring semester</i>
Scope of the discipline	<i>5 credits, 150 hours (18 hours – lectures, 36 hours – practical classes, 96 hours – independent study)</i>
Semester control/control measures	<i>Exam, Modular control work</i>
Class schedule	<i>http://roz.kpi.ua/</i>
Language of instruction	<i>Ukrainian</i>
Information about the course leader/teachers	<i>Lectures and practical classes: Candidate of Physical and Mathematical Sciences, Senior Lecturer at the Department of PMA, Yulia Petrivna Bai ju.p.bai@gmail.com</i>
Course location	<i>Lecture files, practical work, video recordings of online classes on Google Drive (link provided during the first class)</i>

Course programme

1. Description of the course, its purpose, subject matter and learning outcomes

In most cases, structured data is stored in databases and data warehouses, so it is very important to understand how it can be organised, represented, accessed, used and its integrity ensured. After studying the discipline "Introduction to Databases," students will be able to make basic queries to the database to obtain the necessary information without the help of IT specialists, which can significantly speed up the process of preparing and analysing available sociological and related data.

The aim of the educational component is to develop students' abilities to:

- *understand the main modern information technologies for data processing;*
- *understand the place and role of databases in information processing systems, methods of database design and management;*
- *understand the basic models of data structures (lists, hierarchies, relations, network structures);*
- *apply methods for developing an infological model of a subject area;*
- *understand the classification of database management systems (DBMS) according to supported data models, types of stored information, methods of access organisation, and system architecture; select a DBMS for solving*

standard professional tasks;

- *know the basic concepts of the relational data model;*
- *understand normal forms of relations up to and including 4NF, identify violations of normal forms;*
- *know the basic constructs of the SQL language for defining database objects, modifying them, and constructing queries of various types and levels of complexity;*
- *visualise query results;*
- *import and export data from a database to files of various formats;*
- *be able to create and use virtual and materialised database views;*

Main tasks of the educational component.

After completing the course, students should demonstrate the following learning outcomes:

knowledge:

- *principles of organisation, main types and characteristics of database management systems (DBMS);*
- *data models (lists, hierarchies, relations, network structures);*
- *stages of database design;*
- *normal forms of relations 1NF – 4NF, their characteristics and purpose;*
- *basic constructs of the SQL data processing language;*
- *methods of ensuring data integrity;*
- *methods of physical database organisation;*
- *the variety of modern DBMS, their areas of application and features;*

skills:

- *design a database from the stage of analysing the subject area of the system to the implementation of the physical model of the database;*
- *execute queries in SQL;*
- *visualise query results;*
- *import data from various sources and export data to various file types.*

According to the educational and scientific programme, mastering this discipline contributes to strengthening:

I. general competencies:

- *GC01. Ability to think abstractly, analyse and synthesise.*
- *GC02. Ability to work autonomously.*
- *GC05. Ability to evaluate and ensure the quality of work performed.*
- *GC06. Ability to make informed decisions.*

II. Professional competencies:

- *PC01. Ability to analyse social phenomena and processes.*
- *PC03. Ability to design and implement sociological research, develop and justify their methodology.*
- *FC04. Ability to collect and analyse empirical data using modern methods of sociological research.*
- *FK11. Ability to analyse open source data (OSINT), analyse qualitative information, text data, use intelligent analysis for social data.*
- *FK12. Ability to apply modern statistical methods, models, digital technologies, and specialised software for modelling social processes.*

The programme learning outcomes are:

- *PP01. Analyse social phenomena and processes using empirical data and modern concepts and theories of*

sociology.

- *PR04. Apply scientific knowledge, sociological and statistical methods, digital technologies, and specialised software to solve complex problems in sociology and related fields of knowledge.*
- *PR05. Search for, analyse and evaluate the necessary information in scientific literature, databases and other sources.*
- *PR09. Plan and carry out scientific research in the field of sociology, analyse results, and justify conclusions.*
- *PR12. Analyse open source intelligence (OSINT), analyse qualitative information and text data, and use intelligent analysis for social data.*
- *PR13. Apply modern methods of sociological research in the context of the digitalisation of social relations.*

2. Prerequisites and post-requisites of the discipline (place in the structural-logical scheme of training under the relevant educational programme)

Prerequisites: to study the discipline "Introduction to Databases", students need basic knowledge of programming and computer analysis of information. To perform certain practical tasks, students need the knowledge and skills acquired while studying the discipline "Programming in Python".

Post-requisites: studying the discipline deepens students' analytical training and creates a basis for further mastering professional components related to data processing, storage and analysis, as well as for writing master's theses using databases and analytical tools.

3. Course content

1. *Database concept and technology. The concepts of a database, data bank, and DBMS. DBMS functions. DBMS architecture. DBMS architecture components. Data models in databases. Characteristics of data model components (lists, hierarchies, network structures, relations).*

2. *Relational data model. Basic definitions. Interpretation of relations in the form of tables. Properties of tabular representation. Definition of the concept of a relation and its elements. Keys. Objects and relationships of the infological model in the relational data model.*

3. *Stages of database design and their characteristics (conceptual, logical, physical). Entities and relationships between them, attributes of entities and relationships. Data types, null values. Creating constraints and checks for table fields. Ensuring database integrity at the domain and table level. CASE tools for database design.*

4. *Normalisation of relations, purpose and general characteristics of normalisation steps. Normal forms of relations 1NF - 4NF. Examples of violations of normal forms. Methods of bringing relations to 1NF - 4NF.*

5. *General characteristics of the SQL language. SQL standards, methods of their implementation. Structure of the SQL language. Data Definition Language. Database object definition operators. Creating and deleting database objects.*

6. *Data Manipulation Language. Data manipulation language operators in SQL: composition and purpose. Operators for inserting data into tables, deleting and updating data. Searching for data using the select operator. Structure of the select command. Where, between, in, like constructions. Aggregate functions of the SQL language. Group by, having, order by constructions, features of their use. Implementing a quantitative query on one or more columns using SQL.*

7. *Nested queries. Combined queries using union and except operations. Queries based on the use of inner join, left outer join, right outer join, and full outer join.*

8. *Views. Adding, deleting, and updating data in views. Virtual and materialised views. Improving query processing performance. Creating and using indexes in a database.*

9. *Transactions, their definition and purpose. Transaction properties. Anomalies of parallel transaction execution. Transaction isolation levels. Introduction to database administration. Database administration tools in PostgreSQL.*

4. Training materials and resources

Basic literature:

1. *Databases in Information Systems: Textbook / V. I. Gaidarji, I. V. Izvarin. - Kyiv: University "Ukraine", 2018. - 418 p.*
2. *Database organisation: textbook / O. G. Trofimenko [et al.]. 2nd ed. Odessa: Phoenix, 2019. 245 p.*
3. *Khariv N. O. Databases and Information Systems: Textbook / N. O. Khariv. – Rivne: NUVGP, 2018. – 127 p.*

<https://ep3.nuwm.edu.ua/9129/3/%D0%A5%D0%B0%D1%80%D1%96%D0%B2%20%D0%9D.%D0%9E.pdf>

4. Dotsenko S. I. *Organisation and Management Systems for Databases: Textbook*. – Kharkiv: UkrDUZT, 2023. – 117 p., fig. 92, table 3.

<http://lib.kart.edu.ua/bitstream/123456789/13596/1/%D0%9D%D0%B0%D0%B2%D1%87%D0%B0%D0%BB%D1%8C%D0%BD%D0%B8%D0%B9%20%D0%BF%D0%BE%D1%81%D1%96%D0%B1%D0%BD%D0%B8%D0%BA.pdf>

5. Berko A.Yu., Veres O.M., Pasichnyk V.V. *Database and knowledge systems. Book 2. Database and knowledge management systems: Textbook*. – Lviv: Magnolia 2006, 2012. – 584 p.

Additional literature:

1. *An Introduction to Database Systems, 8th edition* / C. J. Date. – Pearson Education, 2003. – 1040 p.
2. *Database systems. A practical approach to design, implementation and management* / T. M. Connolly, C. E. Begg. – Addison-Wesley, 2005. – 1374 p.
3. *Relational Database Design and Implementation, 4th Edition* / J. Harrington. – Morgan Kaufmann, 2016. – 712 p.
4. *Beginning Databases with PostgreSQL, 2nd edition* / N. Matthew, R. Stones. – Berkeley: Apress, 2005. – 637 p.
5. Bui D.B., Silveystruk L.M. *Formalisation of the "entity-relationship" model*. – Kyiv: VPC "Kyiv University", 2011. – 175 p.
6. B. Forta. *SQL in 10 Minutes, 4th ed. (Sams Teach Yourself)*. – Sams Publishing, 2012. – 288 p.
7. *PostgreSQL Documentation*. <https://www.postgresql.org/docs/>

Educational content

5. Methodology for mastering the academic discipline (educational component)

5.1. Lectures

No No	Lecture topic and list of key questions	Number lecture hours
1	<i>Database concept and technology. The concept of a database, data bank, DBMS. DBMS functions. DBMS architecture. DBMS architecture components. Data models in databases. Key concepts and definitions. Characteristics of data models (lists, hierarchies, network structures, relationships).</i>	2
2	<i>Relational data model. Basic definitions. Interpretation of relations in the form of tables. Properties of tabular representation. Definition of the concept of a relation and its elements. Keys. Objects and relationships of the infological model in the relational data model.</i>	2
3	<i>Stages of database design and their characteristics (conceptual, logical, physical). Entities and relationships between them, attributes of entities and relationships. Data types, null values. Creating constraints and checks for table fields. Ensuring database integrity at the domain and table level. CASE tools for database design.</i>	2
4	<i>Normalisation of relations, purpose and general characteristics of normalisation steps. Normal forms of relations (1NF - 4NF). Examples of violations of normal forms. Methods for bringing relations to 1NF - 4NF.</i>	2
5	<i>General characteristics of SQL. SQL standards, methods of implementation. Structure of SQL. Data Definition Language. Database object definition operators. Creating and deleting database tables.</i>	2
6	<i>Data Manipulation Language. Data manipulation language operators in SQL: composition and purpose. Operators for inserting data into tables, deleting and updating data. Searching for data using the select operator. Structure of the select command. Where, between, in, like constructions. SQL aggregate functions. Group by, having, order by constructions, features of their use. Implementation of a quantitative query on one or more columns using SQL.</i>	2

7	<i>Nested queries. Combined queries using union and except operations. Queries based on the use of inner join, left outer join, right outer join, and full outer join.</i>	2
8	<i>Views. Adding, deleting, and updating data in views. Virtual and materialised views. Improving query processing performance. Creating and using indexes in a database.</i>	2
9	<i>Transactions, their definition and purpose. Transaction properties. Anomalies of parallel transaction execution. Transaction isolation levels. Introduction to database administration. Database administration tools in PostgreSQL.</i>	2

5.2. Practical classes

No No	Name and objectives of practical work	Number of classroom hours
1	<i>Levels of database design: conceptual, logical, physical.</i>	2
2	<i>Entities, types of relationships between entities. Attributes of entities and relationships. Data types, null values.</i>	2
3	<i>Using CASE tools for database design (using the example of the ERDplus web service).</i>	2
4	<i>Installing the PostgreSQL DBMS. Creating new users. Configuring the connection. Creating a database and its objects. Configuring permissions.</i>	2
5	<i>Executing queries in SQL. Filtering data. WHERE, BETWEEN, IN, LIKE constructions.</i>	2
6	<i>Using aggregate functions of the SQL language in queries.</i>	2
7	<i>Grouping and sorting data. GROUP BY, HAVING, ORDER BY constructs.</i>	2
8	<i>Nested queries. Combined queries using union and except operations. Table joins. Queries based on inner join.</i>	2
9	<i>Queries based on outer left (right) join and outer full join. Modular control work#1.</i>	1 + 1
10	<i>Executing queries to the PostgreSQL database from programmes using Python code as an example.</i>	2
11	<i>Visualisation of data obtained from the database using the matplotlib package. Construction of dependency graphs, histograms, and pie charts.</i>	2
12	<i>Inserting, updating, and deleting data. Modifying table structures.</i>	2
13	<i>Creating and using views.</i>	2
14	<i>Extension of the SQL language: conditional operators, loops, cursors.</i>	2
15	<i>Importing data into a database from various sources.</i>	2
16	<i>Exporting data from the database to .csv and .json files.</i>	2
17	<i>Improving query processing performance. Creating and using indexes in the database.</i>	2
18	<i>Transactions. Anomalies during transaction execution. Transaction isolation levels. Modular control work No. 2.</i>	1

6. Independent work by students

No No	Topic title and independent work assignment	Number hours
1	Study the concepts of database, data bank, DBMS, database concept, DBMS functions, and the main components of DBMS architecture.	4
2	Study the basic data models in databases (lists, hierarchies, relationships, network structures) and their characteristics.	4
3	Study the basic definitions of the relational data model, the interpretation of relations in the form of tables, the properties of tabular representation, the concept of a relation and its elements, primary and foreign keys.	4
4	Analyse the stages of database design: conceptual, logical, physical.	4
5	Study the main types of relationships in the infological model in the relational data model	4
6	Study the use of CASE tools for database design using the example of the ERDplus web service.	3
7	Study relation normalisation, methods of bringing relations to first normal form (1NF). Analyse the concept of functional dependency (FD) in relations, the definition of complete functional dependency and second normal form (2NF). Study the algorithm for reducing a relation to 2NF.	4
8	Analyse the definition of transitive dependency and third normal form (3NF), the algorithm for reducing to 3NF. Study multi-valued dependencies, characteristics of relations in 4NF, and methods for reducing relations to 4NF.	4
9	Study SQL language operators for defining database objects, creating and deleting database tables. Examine data types and problems associated with data types in databases. Study database integrity at the domain and table levels.	3
10	Study SQL data manipulation language operators, know their structure and purpose. Analyse operators for inserting data into tables, deleting and updating table data.	4
11	Study data search using the select operator, analyse the structure of the select command, where, between, in, like constructions, aggregate functions, group by, having, order by constructions, features of their use, quantitative queries on one or more columns using SQL.	3
12	Work on creating nested queries. Analyse combined queries using union and except operations.	3
13	Work on creating queries based on the use of inner inner join, left outer join, right outer join, and full outer join.	3
14	Work on views, adding, deleting, and updating data in views, virtual and materialised views	3
15	Work on PL/SQL language extensions: conditional operators, loops, cursors, their purpose, description, application.	3
16	Study methods of accessing data in a database, methods of measuring query processing performance, improving performance by using indexes.	3
17	Study the concept of transactions, transaction properties, anomalies in parallel transaction execution, and transaction isolation levels.	3
18	Study database administration tools in PostgreSQL, command and object privileges, grant and revoke commands.	3
19	Preparation for Modular control work	4

20	Preparation for the exam	30
	Total	96

Policy and control

7. Policy on academic discipline (educational component)

- **Attendance at lectures and practical classes.** *Absence from lectures and practical classes without a valid reason is not permitted.*
- **Rules of conduct in class.** *During lectures and practical classes, students must turn off their mobile phones or switch them to silent mode.*
- **Rules for defending practical work.** *All practical work is formalised and submitted by students in the form of reports – doc or pdf files – and uploaded to Github. The defence of practical work takes the form of an oral interview with the student based on the results of the formalised report. Students may also be asked questions from the list of control questions contained at the end of each practical work assignment.*
- **Rules for awarding incentive and penalty points.** *For active participation in lectures and practical classes, incentive points are awarded in the amount of up to 4 points per semester. Penalty points are awarded for late completion of practical work (see below RSO).*
- **Procedure for appealing the results of assessment measures.** *Students have the opportunity to raise any issue related to the assessment procedure and expect it to be considered in accordance with pre-defined procedures. To appeal the assessment measures, the student must submit a statement indicating the reason for the appeal and providing evidence of the teacher's bias. The teacher must discuss this application with the student in person during a consultation. If there is no agreement on the result of the assessment, a commission of teachers from the department is formed to evaluate the assessment procedure and the student's claims. The commission may decide to repeat the assessment or reject the application. The commission's decision is final and cannot be appealed.*
- **Academic integrity policy.** *According to the KPI Student Code of Honour, when completing practical work, it is prohibited to use other people's completed practical work or fragments thereof. In the event of plagiarism, the practical work may be assessed from 0 to 1/3 of the total number of points planned for it.*
- **Informal education.** *At the request of the applicant, in conditions that do not facilitate regular attendance of classes, it is permissible to study individual content-rich parts of the discipline in asynchronous mode, in particular through distance learning courses and other forms of informal education. In order to be included in the rating system for assessing points for such courses, they must correspond in content to certain topics of the syllabus, and their completion must be agreed with the teacher of the discipline. To confirm completion of informal learning, the student must provide a relevant document (certificate) indicating the name of the courses and their duration in hours. Recognition of informal education results is carried out in accordance with the procedure set out in the relevant Regulations of Igor Sikorsky KPI: <https://osvita.kpi.ua/node/179>*

8. Types of control and rating system for assessing learning outcomes (RSO)

A student's rating for a credit module consists of points awarded for:

- 1) for completing and defending practical work (20 points);
- 2) for completing the Modular control work (30 points);
- 3) for completing the exam control work (50 points).

RATING POINT SYSTEM

8.1. Points for completing and defending practical assignments

During the semester, students complete 4 practical assignments:

- 1) Conceptual, logical and physical database design.
- 2) Execution of queries using SQL.
- 3) Designing their own database. Normalisation of tables.
- 4) Execution and visualisation of queries using the Python programming language. Import and export of data.

Maximum number of points for each practical assignment: **5 points**.

Points are awarded for:

- quality of work: 0-3 points;
- answers during the defence of the practical work: 0-2 points.

*Criteria for assessing **the quality of the work**:*

3 points — the work is done well and in full;

2 points — the work is done well and in full, but has flaws;

1 point — the work is completed in full, but contains errors;

0 points — the work is not completed in full or contains significant errors.

*Criteria for evaluating **answers during the defence of practical work**:*

2 points — answers during the defence are complete and well-reasoned;

1 point — the answers contain errors;

0 points — no answers or incorrect answers.

Maximum number of points for completing and defending practical work:

5 points × 4 practical assignments = 20 points.

8.2. Points for completing the Modular control work

The Modular control work is divided into two 45-minute tests, each of which contains 15 test questions related to both the theoretical and practical parts of the course.

Maximum number of points for answering each question: **1 point**.

Assessment criteria:

1 point — correct answer;

0 points — incorrect answer.

Maximum number of points for modular control: 1 point

× 15 questions × 2 CR = 30 points.

8.3. Points for completing the exam control work

The exam test contains 10 tasks. Weighted score: **5 points** for each task. Maximum number of points for the exam test:

5 points × 10 tasks = 50 points.

Assessment criteria:

5 points – complete and correct answer;

4 points – complete and generally correct answer with minor errors; 3 points – answer is incomplete or contains significant errors;

0 points – incorrect answer or no answer.

8.4. Calculation of the rating scale (R):

The rating scale for the discipline is **R = 100 points**.

The sum of weighted points during the semester is:

$R_D = 50 + 50 = 100$ points.

A prerequisite for a student's admission to the exam is the completion of all practical work.

8.5. Current assessment

At the first assessment, in the 8th week, the student receives a "pass" if their current rating is at least **12 points**, i.e. 50% of the maximum number of points that a student can receive before the first assessment.

At the second assessment, in week 14, students receive a "pass" if their current rating is at least **25 points**, i.e. 50% of the maximum number of points that a student can receive before the second assessment.

Table of correspondence between rating points and grades on the university scale:

<i>Total rating, R_b</i>	<i>Grade</i>
100-95	Excellent
94	Very good
84	Good
74-65	Satisfactory
64-60	Sufficient
< 60	Unsatisfactory
Admission requirements not met	Not admitted

Work programme for the academic discipline (syllabus):

Prepared by *Y. P. Bai, Ph.D., senior lecturer at the Department of PMA*

Approved by the Department of Sociology (Minutes No. 18 of 10.06.2024)

Approved by the Methodological Council of Igor Sikorsky KPI (Minutes No. 12 of 21.06.2024)