Projekt "Perspektywy Współpraca Synergia Zarządzanie w Tarnowie" współfinansowany jest przez Unię Europejską ze środków Europejskiego Funduszu Społecznego w ramach Programu Operacyjnego Wiedza Edukacja Rozwój. Projekt realizowany w ramach konkursu Narodowego Centrum Badań i Rozwoju z III Osi priorytetowej: Szkolnictwo wyższe dla gospodarki i rozwoju; Działanie 3.5 Kompleksowe programy szkół wyższych. Nr umowy o dofinansowanie projektu: POWR.03.05.00-00-Z087/17-00.

Module SYLLABUS

Organizational unit name	The Polytechnic Institute – Department of Computer Science		
Field of study	Computer science		
Module name	Data modeling, analysis and warehousing		
Module code	POWER.IP.1	**Erasmus code**	11.3
ECTS	3	**Module type**	Optional
Year of study	3	**Semester**	6

Form of classes	**Hours total**	**Form of assessment**
Project classes	30	Graded credit

Coordinator teacher	mgr inż. Tomasz Potempa		
Academic teacher	mgr inż. Tomasz Potempa		
Language of instruction	English		
Basic courses	No	**Open course / course at he another field of study**	No
Profile of education	Practical profile	**Study level**	First-cycle level

Prerequisites and additional requirements

1. Basic knowledge of discrete mathematics in the field of algebra of sets and relation algebra.
2. Basic knowledge of statistics.
3. Basic knowledge of the propositional logic.
4. Knowledge of the basics of relational databases.
5. Knowledge of programming in object-oriented language.
6. Knowledge of basic terminology in the field of data warehousing.
7. Knowledge of the objectives and principles of data analysis and data mining.
8. Basic knowledge of database management systems.
9. Knowledge of English to a degree that allows communication and studying professional literature.

Learning outcomes for module

No.	**Student after module completion has the knowledge/knows how to/is able to Learning outcome code**	**Learning outcome type**	**Method of learning outcomes verification**	**Form of classes**
				Project
1.	The student knows and understands the basic concepts and methods of data analysis.	Knowledge	project	Y

Projekt "Perspektywy Współpraca Synergia Zarządzanie w Tarnowie" współfinansowany jest przez Unię Europejską ze środków Europejskiego Funduszu Społecznego w ramach Programu Operacyjnego Wiedza Edukacja Rozwój. Projekt realizowany w ramach konkursu Narodowego Centrum Badań i Rozwoju z III Osi priorytetowej: Szkolnictwo wyższe dla gospodarki i rozwoju; Działanie 3.5 Kompleksowe programy szkół wyższych. Nr umowy o dofinansowanie projektu: POWR.03.05.00-00-Z087/17-00.

2.	The student knows and understands the goals and principles of the data warehouse work.	Knowledge	project	Y
3.	The student can use basic methods of data analysis.	Skills	technical documentation	Y
4.	The student can design and implement a simple data warehouse.	Skills	technical documentation	Y
5.	The student is aware of the role and importance of large data sets and their analysis in social applications and in the company.	Social competence	project	Y

Didactic methods

Class forms:
Classes have a form of project under which the uncomplicated warehouse or the simple BIG DATA system is modelled and implemented with the use of selected tools (PostgreSQL, GNU R, IBM, Oracle, Apache Hadoop, Apache Spark).

Teaching methods:
Motivation to consistent and systematic making subsequent stages of the project, brainstorming, project method.

Rules of assessment

Preparation and presentation of a self-made project which consists of documentation and implementation. Assessment must comply with the current university regulations.

Module content (brief)

1. Analysis of requirements, conceptual design.
2. Implementation of the data warehouse or BIG DATA system.
3. Methods of importing data and preliminary data processing.
4. Programming advanced SQL queries and implementation of the statistical methods.
5. Performing analysis using an external analytical package.
6. Preparation of a full technical documentation.

Module content (comprehensive)

1. Analysis and description of requirements. Exploration and discovery modelled data (analysis and synthesis) based on interviews and documentation. Preparation of a conceptual design.
2. Collecting of the large sets of data (e.g. datasets, facts) and preparing the data for use in database/warehouse or BIG DATA system.
3. Design and implementation of the database/warehouse with using of the database management system (e.g. PostgreSQL, Oracle, IBM DB2) or with using of the BIG DATA system (e.g. Apache Hadoop).
4. Importing datasets with use of ETL procedure (Extract, Transform, Load) or creating access to data (e.g. connecting datasets). Preliminary data processing.
5. Programming advanced SQL queries for processing the data and implementation of the statistical methods for data analysis and data mining.
6. Performing analysis using an external analytical package (e.g. R Language, RapidMiner, Pentaho, Apache Spark).
7. Preparation of a complete technical documentation of the project using the adopted conventions, including the SQL standards.

Projekt "Perspektywy Współpraca Synergia Zarządzanie w Tarnowie" współfinansowany jest przez Unię Europejską ze środków Europejskiego Funduszu Społecznego w ramach Programu Operacyjnego Wiedza Edukacja Rozwój. Projekt realizowany w ramach konkursu Narodowego Centrum Badań i Rozwoju z III Osi priorytetowej: Szkolnictwo wyższe dla gospodarki i rozwoju; Działanie 3.5 Kompleksowe programy szkół wyższych. Nr umowy o dofinansowanie projektu: POWR.03.05.00-00-Z087/17-00.

Recommended literature and teaching resources	
1. Paulraj Ponniah: Data Warehousing Fundamentals for IT Professionals. Wiley, 2010. 2. Kuis Torgo: Data Mining with R. Learning with Case Studies. CRC Press, 2010. 3. Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy, Bob Becker: The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence. John Wiley & Sons, 2010. 4. Krish Krishnan: Data Warehousing in the Age of Big Data. Morgan Kaufmann, 2013. 5. Chuck Ballard, Daniel M. Farrell, Amit Gupta, Carlos Mazuela, Stanislav Vohnik: Dimensional Modeling: In a Business Intelligence Environment. IBM Redbooks, 2012. 6. Anthony David Giordano: Data Integration Blueprint and Modeling: Techniques for a Scalable and Sustainable Architecture. IBM Press, 2010. 7. Christopher Adamson: Star Schema The Complete Reference. McGraw-Hill Osborne Media, 2010.	
Connection with area of study	engineering sciences
Student workload (ECTS credits balance)	
Student workload form	Student workload [hours]
Participation in project classes	30
Completion of a project	35
Individual consultations and final project presentation	10
Summary student workload	75
Module ECTS credits	
Workload of the direct assistance of the academic teacher	**1.6**
Workload of the practical classes	**3**

Annotation:
1 hour = 45 minutes