COURSE DESCRIPTION CARD - SYLLABUS

Course name
Advanced Database Technology [N2Inf1-AMiWdIP>ZTBD]

Course

Field of study
Computing

Year/Semester
1/2

Area of study (specialization)
Mobile and Embedded Applications for the Internet of Things

Profile of study
general academic

Level of study
second-cycle

Course offered in
Polish

Form of study
part-time

Requirements
compulsory

Number of hours

Lecture
16

Laboratory classes
16

Other
0

Tutorials
0

Projects/seminars
0

Number of credit points

4,00

Coordinators

dr inż. Paweł Boiński
pawel.boinski@put.poznan.pl

Lecturers

Prerequisites

Educational outcomes from first-cycle studies defined in the Resolution of the PUT Senate, especially K_W1-2, K_W4, K_W6-15, verified in the recruitment process for second-cycle studies - these effects are presented on the faculty's website. Educational outcomes from first-cycle studies defined in the Resolution of the Senate of the State University of Technology, especially the outcomes K_U1-2, K_U4, K_U7-8, K_U14-20, K_U22-23, K_U26, verified in the recruitment process for second-cycle studies - these effects are presented on the website faculty online. Educational outcomes from first-cycle studies defined in the Resolution of the PUT Senate, especially K_K1-9, verified in the recruitment process for second-cycle studies - these effects are presented on the faculty's website. Moreover, in terms of social competences, the student must demonstrate such skills attitudes such as honesty, responsibility, perseverance, cognitive curiosity, creativity, personal culture, respect for other people.

Course objective

Providing students with knowledge regarding the design and implementation of data warehouses and BI applications, in the field of: architectures, data modeling, designing the integrating and power layer - ETL, SQL extensions for BI applications, physical and index structures of data warehouses, development trends of data warehouses and systems B.I. Providing students with basic knowledge about NoSQL databases. Developing students' problem-solving skills in the field of: designing and implementing a data warehouse system, assessing the suitability of data warehouse and BI technologies, and data mining for a specific application. Developing students' skills in implementing data warehouse, BI and data mining projects. Developing students' skills in using sources of knowledge (e.g. English-language publications), self-education and time management.

Course-related learning outcomes

Knowledge
has advanced and in-depth knowledge of data warehouses and NoSQL databases, the theoretical foundations of their construction
has detailed knowledge of methods, tools and development environments used to implement data warehouses
has knowledge of the development trends of NoSQL databases
has advanced and detailed knowledge of the processes occurring in the life cycle of a data warehouse
knows advanced methods, techniques and tools used in solving complex engineering tasks and conducting research work in the field of data warehousing and data analysis
Skills
is able to acquire information on advanced database systems from the literature and other sources (in the native language and in English), integrate them, interpret and critically evaluate them
is able to use experimental methods to formulate and solve engineering tasks and research problems
is able - when formulating and solving engineering tasks in the field of advanced database systems - to integrate knowledge from different areas of computer science (e.g., software engineering, information systems administration, databases)
can assess the usefulness and applicability of new developments (methods and tools) and new IT products (in particular, database management systems, ETL development environments)
is able to determine the directions of further learning and realize the process of self-education (among other things, using the available guides to ETL tools)
Social competences
understands that in computer science, database knowledge and skills are rapidly becoming obsolete
understands the importance of using the latest database knowledge in solving research and practical problems (including selecting appropriate tools and methods for loading and refreshing data warehouses)

Methods for verifying learning outcomes and assessment criteria

Learning outcomes presented above are verified as follows:

In the field of lectures, verification of the assumed learning outcomes is carried out by assessing the knowledge and skills demonstrated during a final colloquium with problem questions and/or
open. The test is considered passed after obtaining more than 50% of the possible points.

The following grading and point scale is adopted:
<0;50%>: n/a, (50%;60%>: dst, (60%;70%>: dst+, (70%;80%>: db, (80%;90%>: db+, (90%;100%>: very good

In the field of laboratories, verification of the assumed learning outcomes is carried out by:
- assessment of the implementation of tasks assigned during each class,
- assessment of knowledge and skills related to the implementation of laboratory tasks by completing a test (in the form of a test and open questions) at the end of the semester.
- obtaining additional points for activity during classes, especially for:
- discussion of additional aspects of the issue,
- comments related to the improvement of teaching materials.

The condition for passing the laboratory is to send the project completed as part of the laboratory to the ekursa system.

In the laboratory, the following grading scale is adopted depending on the number of tests obtained points: <0;50%>: n/a, (50%;60%>: dst, (60%;70%>: dst+, (70%;80%>: db, (80%;90%>: db+, (90%;100%>: very good.

Programme content

The lectures cover the development of data storage methods in computer science from the first databases, through the popular relational model to the modern ideas of Big Data including the rise and development of data warehouses and NoSQL database systems.

The program of laboratory classes includes practical use of ETL tools to build a ROLAP data warehouse.

Course topics

The lecture program covers the following topics:
- BigData processing issues, including: BI;
- data warehouse system architecture (basic ETL, with ODS layer, with data mart layer, ELT architecture, BigData),
- data modeling (conceptual multidimensional model, relational implementation - diagrams star, snowflake, constellation of facts with their evaluation, multidimensional implementation),
- physical structures of data warehouse files and index structures,
- NoSQL databases for web and cloud applications (Why NoSQL? Key-value databases, column families, document databases, graph databases)

The laboratory curriculum is divided into the following parts:

1. Introduction to the training environment
- case study,
- data sources, data warehouse scheme,
- basics of the Agile BI methodology.

2. Introduction to the Pentaho Data Integration tool
- basic concepts,
- repository,
- transformation based on one data source,
- sub transformation.

3. Support for multiple data sources
- expansion of existing transformations and sub-transformations with an additional data source,
- data flow path control,
- methods of combining data.

4. Additional transformations
- methods of eliminating duplicates,
- automatic data generation for dimensions,
- feeding the fact table.
- basics of the Agile BI methodology.

5. Advanced transformations
- data sources based on CSV files, detection of changes in data sources,
- operational data store, refreshing the data warehouse.

6. Modern data sources
- XML documents, web services.

7. Data profiling and cleaning, historical data
- detecting errors in data (reference data, data patterns),
- automatic error correction, fixing errors in data sources,

- modifying the transformation to store historical data for changing dimensions.

8. Improving the efficiency of the ETL process, thematic data warehouses
- bulk data loading (Oracle, PostgreSQL, MySQL)
- calculating aggregates from data, example of a thematic data warehouse.

9. Data processing in data warehouses using SQL and its extensions.

Classes are conducted in the form of exercises using computers, with each student working independently. Each task is preceded by a short presentation and then the discussed issues are practiced in practice.

Teaching methods

1. lecture: multimedia presentation, the presentation is supplemented with short examples presented in a traditional way using a blackboard.

2. laboratories: multimedia presentation, the presentation is supplemented with short examples presented in a traditional way using a blackboard, performing exercises in the data warehouse, discussing more difficult exercises at the blackboard, answering questions on an ongoing basis, solving problems on an ongoing basis

Bibliography

Basic
1. Z.Królikowski, Data warehouses - Logical and physical structures, Wydawnictwo Naukowe PP, 2008
2. A.Chodkowska-Gyurics, Data warehouses, Theory and practice, PWN 2014
3. Guy Harrison - NoSQL, NewSQL and BigData. Next Generation Databases, Helion, 2019
4. P.J.Sadalage, M.Flower, "NoSQL - Knowledge Compendium", Helion, 2015

Supplementary
1. Jiang B.: Constructing Data Warehouses with Metadata-driven Generic Operators, and more:
2. Pentaho Data Integration Documentation https://wiki.pentaho.com/
3. Matt Casters, Roland Bouman, Jos Van Dongen: Pentaho Kettle Solutions, John Wiley & Sons 2010
4. A. Pelikant, Data warehouses - from analytical processing to reporting, Ed. Helion, 2011

Breakdown of average student's workload

	Hours	ECTS
Total workload	100	4,00
Classes requiring direct contact with the teacher	34	1,50
Student's own work (literature studies, preparation for laboratory classes/ tutorials, preparation for tests/exam, project preparation)	66	2,50