COURSE DESCRIPTION CARD - SYLLABUS

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

Course

Field of study
Computing

Year/Semester
1/1

Area of study (specialization)
Advanced Internet Technologies

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
12

Laboratory classes
16

Other
0

Tutorials
0

Projects/seminars
0

Number of credit points

3,00

Coordinators

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

Lecturers

Prerequisites

Learning objectives of the first cycle studies defined in the resolution of the PUT Academic Senate, especially K_W1-2, K_W4, K_W6-15 that are verified in the admission process to the second cycle studies – the learning objectives are available at the website of the faculty. Learning objectives of the first cycle studies defined in the resolution of the PUT Academic Senate, especially K_U1-2, K_U4, K_U7-8, K_U14-20, K_U22-23, K_U26 that are verified in the admission process to the second cycle studies – the learning objectives are available at the website of the faculty. Learning objectives of the first cycle studies defined in the resolution of the PUT Academic Senate, especially K_K1-9 that are verified in the admission process to the second cycle studies – the learning objectives are available at the website of the faculty. In addition, in respect to the social skills the student should show attitudes as honesty, responsibility, perseverance, curiosity, creativity, manners, and respect for other people.

Course objective

Providing students with knowledge on the design and implementation of data warehousing and BI class applications, in terms of: architectures, data modeling, design of integration layer - ETL, physical and index structures of data warehouses, development trends of data warehouses, BI and Big Data systems. To provide students with basic knowledge of NoSQL databases. Develioping students' problem-solving skills in: designing and implementing a data warehouse system, evaluating the suitability of data warehouse and BI technologies and data mining for a specific application. Forming students' ability to implement data warehouse, BI and data mining projects. Forming in students the ability to use 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)
3
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:

Learning outcomes presented above are verified as follows:
In terms of lectures, the verification of the established learning outcomes is implemented by assessing the knowledge and skills demonstrated on a credit colloquium with test, problem and open questions. The colloquium is considered to be passed after obtaining more than 50% of the possible points. The following grading and scoring scale is adopted:
<0;50%): 2.0, <50%;60%): 3.0, <60%;70%): 3.5, <70%;80%): 4.0, <80%;90%): 4.5, <90%;100%>: 5.0.
In terms of laboratories, verification of the established learning outcomes is realized by:
- evaluation of the implementation of the tasks assigned in each class,
- evaluation of knowledge and skills related to the implementation of laboratory tasks by solving a test (with possible open questions) at the end of the semester.
- obtaining additional points for activity during classes, especially for:
- discussion of additional aspects of the issue,
- remarks related to the improvement of teaching materials.
In terms of laboratory, the following grading scale is adopted depending on the number of points obtained: <0;50%>: 2.0, (50%;60%>: 3.0, (60%;70%>: 3.5, (70%;80%>: 4.0, (80%;90%>: 4.5, (90%;100%>: 5.0

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 architectures (basic ETL, with ODS layer, with data mart layer, ELT architecture, BigData),
- data modeling (conceptual multidimensional model, relational implementation - star, snowflake, fact constellation diagrams 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, document databases, graph databases, ...).

The laboratory program is divided into the following parts:
1. introduction to the exercise environment
- case study,

4
- data sources, data warehouse schema,
- basics of Agile BI methodology.

2. introduction to the operation of the Pentaho Data Integration tool
- basic concepts,
- repository,
- single data source based transformation,
- subtransformation.

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

4. additional transformations
- methods for eliminating duplicates,
- automatic generation of data for dimensions,
- loading fact table.
- fundamentals of Agile BI methodology.

5. advanced transformations
- data sources based on CSV files, detection of changes in data sources,
- operational data store, data warehouse refresh.

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

7. data profiling and cleaning, historical data
- incorrect data detection (reference data, data patterns),
- automatic error correction, fixing errors in data sources,
- storing historical data for changing dimensions.

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

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

Teaching methods

1. lecture: multimedia presentation, the presentation is supplemented by short examples presented in a traditional manner with the use of a blackboard.
2. laboratories: multimedia presentation, the presentation is supplemented with short examples presented in a traditional manner with the use of the blackboard, performing exercises in the data warehouse, discussing more difficult exercises at the blackboard, answering questions as they arise,

solving problems as they arise

Bibliography

Basic
1. Z.Królikowski, Hurtownie danych – Struktury logiczne i fizyczne, Wydawnictwo Naukowe PP, 2008
2. A.Chodkowska-Gyurics, Hurtownie danych, Teoria i praktyka, PWN 2014
3. Guy Harrison - NoSQL, NewSQL i BigData. Bazy danych następnej generacji, Helion, 2019
3. P.J.Sadalage, M.Flower, „NoSQL - Kompendium wiedzy”, Helion, 2015
Additional
1. Jiang B.: Constructing Data Warehouses with Metadata-driven Generic Operators, and more: Architecture, Methodology, and Paradigm; Concepts, Algorithms, and Operators; Principles, Recommendations, and Exercises. DBJ Publishing, 2011, ISBN-13: 978-3033029200
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, Hurtownie danych – od przetwarzania analitycznego do raportowania, Wyd. Helion, 2011

Breakdown of average student's workload

	Hours	ECTS
Total workload	75	3,00
Classes requiring direct contact with the teacher	28	1,00
Student's own work (literature studies, preparation for laboratory classes/ tutorials, preparation for tests/exam, project preparation)	47	2,00