(*faculty stamp*)

COURSE DESCRIPTION

1. Course title: DEVELOPMENT TRENDS IN DATA ENGINEERING	**2. Course code**
3. Validity of course description: 2018/2019	
4. Level of studies: MSc programme	
5. Mode of studies: intramural studies	
6. Field of study: COMPUTER SCIENCE	(FACULTY SYMBOL)
7. Profile of studies:	
8. Programme: DATABASES AND SYSTEMS ENGINEERING	
9. Semester: 1, 2	
10. Faculty teaching the course: Institute of Informatics, RAu2	
11. Course instructor: Henryk Josiński, PhD	
12. Course classification:	
13. Course status: compulsory	
14. Language of instruction: English	
15. Pre-requisite qualifications: Fundamentals of Database Systems, Database Systems and Applications. Student understands relational database model and architecture of a database management system. He/she has skills to prepare the database schema with normalized relations and is capable of formulating SQL queries as well as creating database applications.	
16. Course objectives: The course includes lecture and lab exercises. purpose of the subject is to teach students how to develop and use modern database systems. The goal of the lecture is to teach / present students topics / issues from the area of data engineering related to processing of diverse, large and variable data sets: data models, data structures, areas of application, selected mechanisms of database management system, query languages. The goal of the lab exercises is to practice the processing of diverse data sets based on different data models: data storing, modifying, retrieving.	
17. Description of learning outcomes:	

Nr	Learning outcomes description	Method of assessment	Teaching methods	Learning outcomes reference code
W1	Student understands the concept of a non-relational database model and is capable of indicating areas of application for presented database solutions.	exam, reports, tests	lecture, laboratory	K2A_W07 K2A_W08 K2A_W09
W2	Student understands the differences between row- and column-oriented data structures and understands the consequences arising therefrom.	exam, reports, tests	lecture, laboratory	K2A_W08 K2A_W09 K2A_U11
W3	Student understands the necessity of elimination of certain ACID properties in NoSQL databases in favor of CAP compromises.	exam, reports, tests	lecture, laboratory	K2A_W05 K2A_W06 K2A_W07
U1	Student is capable of creating databases of different types as well as processing diverse data sets based on different data models: data storing, modifying, retrieving.	exam, reports, tests	laboratory	K2A_U01 K2A_U11 K2A_U13
U2	Student is capable of properly using XML.	exam, reports, tests	lecture, laboratory	K2A_U01 K2A_U10 K2A_U13
U3	Student is aware of influence of appropriate selection of database solutions on processing	exam, reports, tests	lecture, laboratory	K2A_U11 K2A_U12

	efficiency.			K2A_U13 K2A_K02

18. Teaching modes and hours

Lecture / BA / MA Seminar / Class / Project / Laboratory

Lecture 30 h., Laboratory 45h

19. Syllabus description:

Lectures:

Big Data. The *3/5V* model, data storing, programming and processing.

NoSQL databases (2 lectures). Data models used by NoSQL databases: „key-value" (*Amazon Dynamo*), wide column (*Google BigTable*, *Cassandra*), graph (*Neo4j* and *Cypher* query language as well as *Apache TinkerPop* tools), and document (*MongoDB*, *CouchDB*).

In-memory databases. Utilization of RAM for acceleration of data management in database management systems. Utilization of RAM disks. In-memory database management system (*IBM solidDB*). Suggested configurations and applications.

Distributed databases (2 lectures). Architecture of a distributed database management system. Types of data distribution. Semi-join operator. Distributed transaction. Two-phase commit protocol (2PC). Update propagation of replicated data.

Query optimization in a relational dbms (2 lectures). Logical and physical optimization. Rules for transformation of relational algebra expressions. Join algorithms. User's influence on query optimization. Examples and hints.

Fuzzy databases. Fuzzy operators. Membership function. Query formulation.

Full-Text Search (FTS) mechanism. Indexing binary data. Data search modes. Predicates and operators used in queries. FTS in selected database systems.

XML – properties of the language. XML in databases – classification of XML documents with regard to their structure. Storing data from XML documents vs storing XML documents. Mapping of an XML document structure to a schema of a relational database. Databases supporting XML. Native XML databases.

Query languages for XML – *XPath*, *XQuery*.

Architecture, administration and security of a database management system based on the case of *Oracle DBMS*.

The *Flashback Queries* (*Flashback Data*base, *Flashback Table*, *Flashback Drop*) mechanism in *Oracle DBMS*. Retrieving metadata and historical data for a specific time interval.

Laboratory:

12 lab exercises (3h/week):

Big Data – a model, data storing, programming and processing.

Document, graph and column-oriented data models in *NoSQL* databases (3 lab exercises).

IBM solidDB in-memory database.

Fuzzy databases.

Distributed database and data replication (2 lab exercises).

Architecture, administration and security in *Oracle DBMS*.

Flashback queries in *Oracle DBMS*.

XML – mapping between relational databases and XML, XML in database management systems (2 lab exercises).

The database management systems used in lab exercises: *Oracle*, *solidDB*, *Cassandra*, *MongoDB*, *CouchDB*,

Neo4j, *MS SQL Server*, *MySQL*, *PostgreSQL*.

20. Examination: after 2nd semester – written exam

21. Primary sources:

H.Garcia-Molina, J.D.Ullman, J.Widom: Database Systems: The Complete Book (2nd Edition)

R.Elmasri, S.Navathe: Fundamentals of Database Systems (7th Edition)

P.D. Sadalage, M. Fowler: NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence (1st Edition)

G. Harrison: Next Generation Databases: NoSQLand Big Data (1st Edition)

22. Secondary sources:

A. Ploetz, D. Kandhare, S. Kadambi, X. Wu: Seven NoSQL Databases in a Week: Get up and running with the fundamentals and functionalities of seven of the most popular NoSQL databases, 2018.

J. Carpenter, E. Hewitt: Cassandra: The Definitive Guide. Distributed Data at Web Scale. Helion, 2016. (ebook)

Internet sources presented during the lectures and laboratories

23. Total workload required to achieve learning outcomes

Lp.	Teaching mode :	Contact hours / Student workload hours
1	Lecture	30/30
2	Classes	/
3	Laboratory	45/45
4	Project	/
5	BA/ MA Seminar	/
6	Other	15/0
	Total number of hours	90/75

24. Total hours:165

25. Number of ECTS credits: 3

26. Number of ECTS credits allocated for contact hours: 3

27. Number of ECTS credits allocated for in-practice hours (laboratory classes, projects): 2

26. Comments:

Approved:

…….
(date, Instructor's signature)

……
(*date , the Director of the Faculty Unit signature)*