COURSE DESCRIPTION CARD - SYLLABUS

Course name
Distributed Databases [S2Inf1>RBD]

Course

Field of study
Computing

Year/Semester
1/2

Area of study (specialization)
Distributed and cloud systems

Profile of study
general academic

Level of study
second-cycle

Course offered in
Polish

Form of study
full-time

Requirements
compulsory

Number of hours

Lecture
30

Laboratory classes
30

Other
0

Tutorials
0

Projects/seminars
0

Number of credit points

4,00

Coordinators

prof. dr hab. inż. Tadeusz Morzy

dr inż. Juliusz Jezierski
juliusz.jezierski@put.poznan.pl

Lecturers

Prerequisites

A student starting this subject should have a basic knowledge of database systems. The ability to formulate queries in SQL and to program in PL/SQL is essential. He should have the ability to solve basic problems in the administration and operation of database systems, and the ability to obtain information from indicated sources. He should also understand the necessity of expanding his competencies. In addition, in terms of social competence, the student must present such attitudes as honesty, responsibility, perseverance, cognitive curiosity, creativity, personal culture, respect for other people.

Course objective

1. To provide students with basic knowledge of technologies and methods used in distributed database systems. 2. to develop in students the ability to solve problems related to the design and operation of distributed database systems. 3.To develop in students the skills of teamwork and integration of knowledge from different areas of computer science. 4.To develop in students the ability to formulate and test hypotheses related to engineering problems and simple research problems in the field of distributed database systems. 5. to provide students with practical knowledge of technology using leading vendors of distributed database systems.

Course-related learning outcomes

Knowledge:
has advanced and in-depth knowledge of distributed database systems, their theoretical foundations and the methods, tools and programming environments used for their implementation (k2st_w1)
has advanced and detailed knowledge of selected issues in the field of problems of distributed database systems (k2st_w3)
has detailed and advanced knowledge of the processes occurring in the life cycle of an information system using a distributed database system (k2st_w5)
knows advanced methods, techniques and tools used in solving engineering problems and research work in the field of building distributed database systems (k2st_w6)

Skills:
is able, when formulating and solving engineering tasks, to integrate knowledge from different areas of computer science (k2st_u5)
is able to assess the usefulness and applicability of new methods, techniques and it tools (k2st_u6)
is able to solve complex it tasks in the field of distributed systems, including non-standard tasks or tasks that include a research component (k2st_u10)
is able, according to given specifications, to design and implement a complex distributed database system using appropriate methods, techniques and tools (k2st_u11)
is able to identify and implement directions for further self-education (k2st_u16)

Social competences:
understands that in modern computer science, knowledge and skills are rapidly becoming obsolete and require constant updating (k2st_k1)
understands the importance of using the latest knowledge in computer science in solving research and practical problems (k2st_k2)

Methods for verifying learning outcomes and assessment criteria

Learning outcomes presented above are verified as follows:

Learning outcomes presented above are verified as follows:
The learning outcomes outlined above are verified as follows:
Formative assessment:
(a) in terms of lectures:
- on the basis of evaluations of the implemented exercises/tasks at the blackboard
b) in terms of laboratories:
- on the basis of the evaluation of the current progress of the tasks,
Summative assessment:
a) in terms of lectures, verification of the established learning outcomes is realized by:
- evaluation of knowledge and skills demonstrated on an open problem-based written colloquium (the student may use any teaching materials), The colloquium consists of 5-6 problem-based tasks, for which 10 points can be obtained. A total of 50-60 points can be obtained. A passing grade of 3.0 requires 50% of the maximum number of points.
- Discussion of the results of the exam,
b) in the field of laboratories verification of the established learning outcomes is carried out by:
- evaluation of the degree of assimilation of knowledge presented during the laboratory through written reports on the exercises performed
- multiple-choice test assessing the ability to synthesize the acquired knowledge
- discussion of results

Programme content

Introduction to distributed database systems.
Horizontal and vertical data fragmentation.
Data allocation.
Decomposition, localization and optimization of distributed queries.
Management of distributed transactions.
Synchronous and asynchronous data replication.

Course topics

Introduction to distributed database systems: distributed database systems, multi-base systems, parallel database systems. Architectures of distributed systems. Horizontal fragmentation schemes: desiderata, basic and deduced fragmentation, algorithm for defining a horizontal fragmentation scheme. Vertical fragmentation scheme: desiderata, BEA algorithm for defining a vertical fragmentation scheme. Data allocation scheme: problem formulation, information requirements for a data allocation scheme, basic heuristics for data allocation in a distributed database system. Distributed query processing: introduction, semantic analysis of a query, decomposition of a query with respect to fragments, localization of a query with respect to fragments, optimization of distributed queries, semi-join algorithm, SDD-1 algorithm. Management of distributed transactions: concepts and models of distributed transactions. Management of concurrent execution of distributed transactions: global serializability criterion, quasi-serializability criterion, local correctness conditions for concurrent execution of transactions. Algorithms for managing concurrent execution of distributed transactions: locking, algorithm for ordering transactions by time labels, the issue of jamming and methods for solving it. Data replication: problem formulation. Synchronous and asynchronous replication. Consistency criteria. Criterion of correctness of concurrent execution of transactions (1-copy serializability). Algorithm for verification of the criterion. Thomas" rule. Basic replication algorithms. Asynchronous muliti-master replication problem: problem formulation, CAP theorem, consistency criteria and models (eventual consistency, casual consistency, read committed, snapshot isolation), concept of "vector clocks". Reliability in distributed database systems: the 2PC algorithm. Distributed NoSQL databases: concepts, advantages and disadvantages, BASE model.
Laboratory classes are conducted in the form of 15 2-hour exercises, held in the laboratory, preceded by a 1-hour instructional session at the beginning of the semester. The exercises are carried out individually. The laboratory program includes the following topics:
Installing software and creating a distributed database. Transparent access to remote data. Distributed transactions. Simulating failures and restoring the 2PC protocol. Handling distributed jamming. Basic data replication. Horizontal fragmentation of data objects (sharding). Optimization of queries directed to horizontally fragmented tables. Replication of table fragments. Physical and logical streaming replication. Asynchronous bidirectional replication. Types of data update conflicts and their resolution methods. Replication of DDL operations. Materialized perspectives and methods for refreshing them. Configuring and testing the failover mechanism. Reconfiguration of replication after permanent loss of quorum. The class is taught using the software of two post-relational databases: Postgres and Oracle, and two noSQL databases: Cassandra and Mongo.

Teaching methods

Lecture: multimedia presentation, illustrated by examples given on the blackboard.
Laboratories: multimedia presentation illustrated by examples given on the blackboard and performance of tasks given by the instructor - practical exercises.

Bibliography

Basic
1. M. T. Ozsu, P. Valduriez, Principles of Distributed Database Systems, Springer, 3rd ed., 2011.
Additional
1. S. K. Rahimi, F. S. Haug, Distributed database management Systems: A Practical Approach, Wiley, 2010.
2. P.A. Bernstein, E. Newcomer, Principles of Transaction processing, Morgan Kaufmann, 2nd ed., 2009.

Breakdown of average student's workload

	Hours	ECTS
Total workload	100	4,00
Classes requiring direct contact with the teacher	60	2,00
Student's own work (literature studies, preparation for laboratory classes/ tutorials, preparation for tests/exam, project preparation)	40	2,00