



STUDIJŲ DALYKO (MODULIO) APRAŠAS

Dalyko (modulio) pavadinimas	Kodas
Nereliacinės duomenų bazės	

Dėstytojas (-ai)	Padalinys (-iai)
Koordinuojantis: Giedrius Graževičius Kitas(-i): dr. Virginijus Marcinkevičius	Matematikos ir informatikos fakultetas Duomenų mokslo ir skaitmeninių technologijų institutas

Studijų pakopa	Dalyko (modulio) tipas
Pirmoji	Pasirenkamas

Igyvendinimo forma	Vykdymo laikotarpis	Vykdymo kalba (-os)
Auditorinė	3 semestras	Lietuvių / Anglų

Reikalavimai studijuojančiajam	
Išankstiniai reikalavimai: Algoritmai ir duomenų struktūros, Duomenų bazių valdymo sistemos.	Gretutiniai reikalavimai (jei yra):

Dalyko (modulio) apimtis kreditais	Visas studento darbo krūvis	Kontaktinio darbo valandos	Savarankiško darbo valandos
5	133	64	69

Dalyko (modulio) tikslas: studijų programos ugdomos kompetencijos
Dalyko tikslas – siekiama, kad studentai susipažintų su nereliacinėmis duomenų bazėmis, ugdytų gebėjimus įvertinti, pasirinkti ir pritaikyti geriausia duomenų bazių sistemos sprendimą modeliuojamos dalykinės srities duomenims saugoti.

Dalyko (modulio) studijų siekiniai	Studijų metodai	Vertinimo metodai
Gebės konstruoti probleminę sritį atitinkančias nereliacines duomenų bazes.	Probleminis dėstymas, laboratorinių darbų atlikimas, praktinės užduotys.	Laboratorinių darbų, praktinės užduoties ataskaitos vertinimas, egzaminas.
Gebės taikyti nereliacinių duomenų struktūrų modeliavimo principus, aprašyti ir kurti nereliaciniams duomenims saugoti skirtas duomenų struktūras, parinkti tinkamas nereliacinių duomenų bazių valdymo sistemas.		
Gebės kurti debesų kompiuterijos sprendimus, naudojančius nereliacines duomenų bazes.		
Gebės kurti nereliacinių duomenų bazių sprendimus efektyviam didžiųjų duomenų apsikeitimui tarp duomenų bazių valdymo sistemos ir analitikos algoritmų.		

Temos	Kontaktinio darbo valandos							Savarankiškų studijų laikas ir užduotys	
	Paskaitos	Konsultacijos	Seminarai	Pratybos	Laboratoriniai darbai	Praktika	Visas kontaktinis darbas	Savarankiškas darbas	Užduotys
1. Duomenų bazės. Pagrindiniai reliacinių ir nereliacinių duomenų bazių skirtumai. ACID garantijos. Izoliacijos lygiai. Žymų rikiavimo protokolas. Kelių versijų lygiagretumo valdymas. BASE garantijos.	6				6		12	11	
2. Raktas-reikšmė duomenų modelis. Duomenų struktūrų modelis. Sudėtingesnių struktūrų modeliavimas raktas-reikšmė modelyje. Duomenų bazės <i>Redis</i> analizė.	2				2		4	6	
3. Stulpelių šeimos duomenų modelis. <i>Apache Cassandra</i> analizė. CQL kalba.	2				2		4	6	
4. Dokumentų duomenų modelis. JSON ir XML formatai. <i>MongoDB</i> analizė. Map-reduce tipo užklausa.	4				4		8	8	
5. Grafių duomenų bazės. <i>Neo4J</i> analizė. Cypher užklausa kalba.	2				2		4	4	
6. Išskirtosios duomenų bazės. Maišos funkcijos, pastovios maišos funkcijos. Išskirtymo modeliai: šeiminkas-vergas, replikavimas, suskaidymas. Lygiagretumo kontrolė: užraktai, versijavimas.	6				6		12	5	

7. CAP teorema. Pasirenkama darna.	2				2		4	4	
8. Išskirstytasis konsensusas. Dviejų, trijų fazių patvirtinimai, Paxos ir Raft algoritmai. Žurnalai.	6				6		12	5	
9. Duomenų bazės pasirinkimas.	2				2		4	8	
10. Egzaminas.								12	
Iš viso	32				32		64	69	

Vertinimo strategija	Svoris proc.	Atsiskaitymo laikas	Vertinimo kriterijai
Laboratoriniai darbai	30	12 savaitė	Vertinamos balais nuo 0 iki 10. Skiriama po 2 balus už dalykinės srities duomenų bazės realizavimą kiekvieno tipo duomenų bazėje. 2 balai skiriami už gebėjimą palyginti duomenų bazes tarpusavyje. Užduotis gali būti atsiskaitoma dalimis. Vėluojant atsiskaityti užduoties įvertinimas mažinamas 50%.
Praktinės užduoties vertinimas	10	15 savaitė	Paruošti techninę ataskaitą, kurioje pagal pasirinktus kriterijus palyginamos kelios duomenų bazės.
Egzaminas	60	Egzaminų sesijos metu	Egzamine pateikiama 10 klausimų / užduočių. Vertinant atsakymo į klausimą išsamumą už kiekvieną atsakymą skiriama iki 6% galutinio įvertinimo.

Autorius	Leidimo metai	Pavadinimas	Periodinio leidinio Nr. ar leidinio tomas	Leidimo vieta ir leidykla ar internetinė nuoroda
Privaloma literatūra				
Martin Kleppmann	2017	Designing Data-Intensive Applications		O'Reilly Media Inc.
Pramod J. Sadalage; Martin Fowler	2012	NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence		Addison-Wesley
Papildoma literatūra				
Ongaro, Diego, and John Ousterhout	2014	In search of an understandable consensus algorithm.		USENIX Annual Technical Conference https://www.usenix.org/system/files/conference/atc14/atc14-paper-ongaro.pdf
Lampert, Leslie	1998	The part-time parliament		ACM Transactions on Computer Systems (TOCS) https://computerarchive.org/files/mirror/www.bitsavers.org/pdf/dec/tech_reports/SRC-RR-49.pdf
Bernstein, Philip A., and Nathan Goodman	1983	Multiversion concurrency control – theory and algorithms		ACM Transactions on Database Systems (TODS) https://www.researchgate.net/profile/Philip_Bernstein/publication/220225682_Multiversion_Concurrency_Control_-_Theory_and_Algorithms/links/0912f50a2ba4e5f1d6000000.pdf



COURSE UNIT (MODULE) DESCRIPTION

Course unit (module) title	Code
NoSQL Databases	

Lecturer(s)	Department(s) where the course unit (module) is delivered
Coordinator: Giedrius Graževičius Other(s): Dr. V. Marcinkevičius	Faculty of Mathematics and Informatics Institute of Data Science and Digital Technologies

Study cycle	Type of the course unit (module)
First	Optional

Mode of delivery	Period when the course unit (module) is delivered	Language(s) of instruction
face-to-face	3 rd semester	Lithuanian / English

Requirements for students	
Prerequisites: Algorithms and Data Structures, Database Management Systems	Additional requirements (if any):

Course (module) volume in credits	Total student's workload	Contact hours	Self-study hours
5	133	64	69

Purpose of the course unit (module): programme competences to be developed

The purpose of the course unit is to familiarize students with non-relational (NoSQL) databases. To acquaint and develop abilities students with different data models and basics of inner-workings of these systems in order to enable them to choose the right database for a task at hand.

Learning outcomes of the course unit (module)	Teaching and learning methods	Assessment methods
Ability to create data stored in NoSQL database processing algorithms.	Lectures, laboratory and practical assignments, problem oriented teaching.	Lab assignments, practical assignment, exam.
Ability to model unstructured data and implement them, select suitable data management solutions.		
Ability to apply non-structured data processing knowledge in high performance environment.		
Ability to do unstructured data analysis, gather statistics, solve unstructured data mining tasks.		

Content: breakdown of the topics	Contact hours							Self-study work: time and assignments	
	Lectures	Tutorials	Seminars	Exercises	Laboratory work	Internship/work placement	Contact hours	Self-study hours	Assignments
1. Databases. Main relational and non-relational database differences. ACID versus BASE. Transactions. Isolation levels. Time stamp ordering, Multi-version concurrency control.	6				6		12	11	
2. Key-value data model. Data structure model. Modelling complex domains in key-value. Redis database.	2				2		4	6	
3. Column family data model. Apache Cassandra overview. CQL query language.	2				2		4	6	
4. Document store overview. JSON and XML document formats. MongoDB overview. Map-reduce queries.	4				4		8	8	
5. Graph databases. Neo4J overview. Cypher and Gremlin query languages.	2				2		4	4	
6. Distributed databases. Hash functions, consistent hashing. Data distribution: master-slave, replication, partitioning. Data concurrency: locking, multi-version concurrency control.	6				6		12	5	
7. CAP theorem. Tunable consistency.	2				2		4	4	
8. Distributed consensus. Two-phase and three-phase commit. Paxos and Raft algorithms. Logs.	6				6		12	5	
9. Choosing the right database.	2				2		4	8	
10. Exam.								12	Literature review
Total	32				32		64	69	

Assessment strategy	Weight, %	Deadline	Assessment criteria
Practical assignment	30	12 th week	20% for each type of the database solution. 20% for comparing the database solutions. Students late to submit the solution will be penalized for 50% of the final evaluation.
Laboratory work	10	15 th week	Technical report comparing several databases.
Exam	60	Exam session	

Author	Year of publication	Title	Issue of a periodical or volume of a publication	Publishing place and house or web link
Compulsory reading				
Martin Kleppmann	2017	Designing Data-Intensive Applications		O'Reilly Media Inc.
Pramod J. Sadalage; Martin Fowler	2012	NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence		Addison-Wesley
Optional reading				
Ongaro, Diego, and John Ousterhout	2014	In search of an understandable consensus algorithm.		USENIX Annual Technical Conference https://www.usenix.org/system/files/conference/atc14/atc14-paper-ongaro.pdf
Lampert, Leslie	1998	The part-time parliament		ACM Transactions on Computer Systems (TOCS) https://computerarchive.org/files/mirror/www.bitsavers.org/pdf/dec/tech_reports/SRC-RR-49.pdf
Bernstein, Philip A., and Nathan Goodman	1983	Multiversion concurrency control – theory and algorithms		ACM Transactions on Database Systems (TODS) https://www.researchgate.net/profile/Philip_Bernstein/publication/220225682_Multiversion_Concurrency_Control_-_Theory_and_Algorithms/links/0912f50a2ba4e5f1d6000000.pdf
Ongaro, Diego, and John Ousterhout	2014	In search of an understandable consensus algorithm		USENIX Annual Technical Conference https://www.usenix.org/system/files/conference/atc14/atc14-paper-ongaro.pdf