



STUDIJŲ DALYKO (MODULIO) APRAŠAS

Dalyko (modulio) pavadinimas	Kodas
Nereliacinės duomenų bazės	

Dėstytojas (-ai)	Padalinys (-iai)
Koordinuojantis: Giedrius Graževičius	Matematikos ir informatikos fakultetas Duomenų mokslo ir skaitmeninių technologijų institutas

Studijų pakopa	Dalyko (modulio) tipas
Pirmoji	Individualiosios studijos

Igyvendinimo forma	Vykdymo laikotarpis	Vykdymo kalba (-os)
Auditorinė	Rudens 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 geriausią duomenų bazių sistemos sprendimą modeliuojamos dalykinės srities duomenims saugoti.		
Dalyko (modulio) studijų siekiniai	Studijų metodai	Vertinimo metodai
Gebės konstruoti problemine srity atitinkančias nereliacinius 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 apsiėtimui 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.	4						4	6	Savarankiškas studijavimas, pasiruošimas egzaminui. Praktinių ir laboratorinių užduočių atlikimas.
2. Raktas-reikšmė duomenų modelis. Duomenų struktūrų modelis. Sudėtingesnių struktūrų modeliavimas raktas-reikšmė modelyje. Duomenų bazė <i>Redis</i> .	2				6		8	6	
3. Stulpelių šeimos duomenų modelis. <i>Apache Cassandra</i> analizė. CQL kalba.	2				6		8	6	
4. Dokumentų duomenų modelis. JSON ir XML formatai. <i>MongoDB</i> duomenų bazė.	4				6		10	6	
5. Grafų duomenų bazės. <i>Neo4J</i> duomenų bazė. Cypher užklausų kalba.	2				6		8	4	
6. Išskirstytosios duomenų bazės. Maišos funkcijos, pastovios maišos funkcijos. Išskirstymo modeliai: šeimininkas-vergas, replikavimas, suskaidymas. Lygiagretumo kontrolė: užraktai, versijavimas.	6						6	5	
7. CAP, PACELC teoremos. Pasirenkama darna.	2						2	4	
8. Išskirstytasis konsensusas. Dviejų, trijų fazių patvirtinimai, Paxos ir Raft algoritmai. Žurnalai.	4						4	5	

9. NewSQL, analitinės ir paieškos duomenų bazės.	2					2	5	
10. Duomenų bazių disko duomenų struktūros.	2					2	5	
11. Duomenų bazės pasirinkimas.	2				8	10	5	
12. Egzaminas							12	Literatūros kartojimas
Iš viso	32				32	64	69	

Vertinimo strategija	Svoris proc.	Atsiskaitymo laikas	Vertinimo kriterijai
Laboratoriniai darbai	50	12 savaitė	Atliekamos 5 užduotys, galutinio balo svoriai paskirstyti lygiomis dalimis po 10%. Vėluojant atsiskaityti užduoties įvertinimas mažinamas. Atsiskaitant eksternu, studentai gali pasirinkti, ar naudoti anksčiau surinktus balus, ar atsiskaityti už laboratorinius darbus iš naujo. Pasirinkus atsiskaitymą iš naujo, anksčiau surinkti balai anuliuojami.
Egzaminas	50	Egzaminų sesijos metu	Egzaminą sudaro dvi dalys: teorinių žinių testas (20%) ir praktinių situacijų sprendimas (30%). Eksternu kursą laikantys studentai gali laikyti egzaminą.

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
Lamport, Leslie	1998	The part-time parliament		ACM Transactions on Computer Systems (TOCS) https://lamport.azurewebsites.net/pubs/lamport-paxos.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
Daniel J. Abadi	2012	Consistency Tradeoffs in Modern Distributed Database System Design		IEEE Computer Society: Computer. Volume: 45, Issue: 2 https://www.cs.umd.edu/~abadi/papers/abadi-pacelc.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	Faculty of Mathematics and Informatics Institute of Data Science and Digital Technologies

Study cycle	Type of the course unit (module)
First	Individual studies

Mode of delivery	Period when the course unit (module) is delivered	Language(s) of instruction
face-to-face	Fall 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 the abilities of 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.	4						4	6	Self-study, laboratory and practical work, preparation for exam
2. Key-value data model. Data structure model. Modelling complex domains in key-value. Redis database.	2				6		8	6	
3. Column family data model. Apache Cassandra overview. CQL query language.	2				6		8	6	
4. Document store overview. JSON and XML document formats. MongoDB overview. Map-reduce queries.	4				6		10	6	
5. Graph databases. Neo4J overview. Cypher and Gremlin query languages.	2				6		8	4	
6. Distributed databases. Hash functions, consistent hashing. Data distribution: master-slave, replication, partitioning. Data concurrency: locking, multi-version concurrency control.	6						6	5	
7. CAP theorem. Tunable consistency.	2						2	4	
8. Distributed consensus. Two-phase and three-phase commit. Paxos and Raft algorithms. Logs.	4						4	5	
9. NewSQL, analytical and search databases.	2						2	5	
10. Database disk data structures.	2						2	5	
11. Choosing the right database.	2				8		10	5	
12. Exam								12	Literature review
Total	32				32		64	69	

Assessment strategy	Weight,%	Deadline	Assessment criteria
Practical assignments	50	12 th week	5 assignments, each worth up to 10% of the final grade. If the exam is taken externally, students can choose to use grades received in the previous attempt at the course or submit assignments anew. If assignments are submitted anew, the results of the previous attempt at the course are effectively annulled.
Exam	50	Exam session	Exam is composed of a theoretical knowledge test (20%) and practical situation solving (30%). Exam can be taken externally.

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
Lamport, Leslie	1998	The part-time parliament		ACM Transactions on Computer Systems (TOCS) https://lamport.azurewebsites.net/pubs/lamport-paxos.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
Daniel J. Abadi	2012	Consistency Tradeoffs in Modern Distributed Database System Design		IEEE Computer Society: Computer. Volume: 45, Issue: 2 https://www.cs.umd.edu/~abadi/papers/abadi-pacelc.pdf