

RTU studiju kurss "Paralēlās datu bāzu sistēmas"**33000 Datorzinātnes, informācijas tehnoloģijas un enerģētikas fakultāte****Vispārējā informācija**

Kods	DE0277
Nosaukums	Paralēlās datu bāzu sistēmas
Studiju kursa statuss programmā	Obligātais/Ierobežotās izvēles
Atbildīgais mācītbspēks	Jānis Eiduks - Doktors, Docents
Mācītbspēks	Māra Romanovska - Doktors, Docents Ainārs Auziņš - Lektors
Apjoms daļās un kredītpunktos	1 daļa, 3.0 kredītpunkti
Studiju kursa īstenošanas valodas	LV, EN
Anotācija	Ārējās un operatīvās atmiņas attīstības disbalanss bija galvenā motivācija izmantot paralēlās apstrādes iespējas datu bāzes sistēmās (DBS). Ir divas galvenās problēmas: paralēlā vaicājumu apstrāde un režģa (neviendabīga resursu sadarbība) datu bāzes (DB). Pirmā problēma apskata vaicājumu (galvenokārt lasāmu) apstrādes veikspējas palielināšanu, izpildot tos paralēli. To sauc par "ātruma palielināšanu" (speed up). Otrā apskata režģa transakciju vadību un mēroga jeb "paralēlisma pakāpes palielināšanu" (scale up). Studiju kursā tiek definēti un analizēti paralēlās apstrādes ieviešanas galvenie pamatjēdzieni un tiek vērtētas dažādas paralēlisma apstrādes formas: intervaicāšanas paralēlisms (interquery), iekšējais vaicājuma paralēlisms (intraquery), sadalītais paralēlisms (intraoperation), sadarbības paralēlisms (interoperation), konveijera režīms un neatkarīgais paralēlisms. Studiju kursā liela uzmanība tiek veltīta transakciju vadībai. Veicot paralēlas darbības gan vienā datorā, gan daudzos, transakciju vadība ir pareizas un drošas darbības pamatjautājums. Jaunākās paaudzes NoSQL un NewSQL DBS balstās uz intensīvās paralēlās apstrādes (massively parallel processing (MPP)) konceptiem. Tiek apskatītas galvenās MMP arhitektūras: režģa skaitļošana (grid computing), datoru klasteri (computer clustering), MapReduce algoritma izmantošana, datu plūsmas veidošanas modelis (dataflow MMP). NoSQL DBS netiek lietots relāciju modelis. Ir plaša klasteru lietošana, nav datu bāzes shēmas, ir orientācija uz datu agregātiem. Paralēlo komponentu izkliešanas modeļi ir nodalījumi (sharding) un replikācijas (mater-slave replication, peep-to-peer replication). Galvenie NoSQL datu modeļi ir atslēga – vērtība (key – value) datu modelis, dokumentu datu modelis, kolonu kopas un grafa datu modeļi. Studiju kursā tiek analizētas populārākās NoSQL sistēmas. NewSQL ir jaunākā DBS klase, kas mēģina nodrošināt NoSQL sistēmu veikspēju tiešsaistes transakciju apstrādes sistēmām (online transaction processing systems (OLTPs)) datu lasīšanas un rakstīšanas darba slodzei, saglabājot relāciju datu bāzes sistēmu ACID īpašību izpildes garantijas. NewSQL ir veidotas izmantojot relāciju modeli un SQL valodu. Tās saglabā ACID principa ievērošanu transakciju izpildē un spēj nodrošināt NoSQL datu bāzes veikspēju pie datu ierakstīšanas un nolasīšanas ar horizontālās mērogojamības palīdzību. Mezgli darbojas ar augstu veikspēju, kas tiek panākta ar datu glabāšanu operatīvajā atmiņā. NewSQL izmanto "nekā kopēja" arhitektūru, kurā mezgliem nav nekādu kopēji izmantojamu resursu. Lieto laiksakritības (concurrency) vadību, kura neizmanto bloķēšanas mehānismu konkurences pretrunu risināšanai (non-locking concurrency control).
Mērķis un uzdevumi, izteikti kompetencēs un prasmēs	Studiju kursa mērķis ir iepazīstināt studentus ar paralēlās apstrādes (paralelizācijas) iespēju realizēšanu DBS. Studiju kursa uzdevumi ir: - sniegt zināšanas par jaunās paaudzes DBS NoSQL un NewSQL iespējām; - sniegt padziļinātas zināšanas par režģa datu bāzes sistēmu arhitektūru; - pilnveidot praktiskās iemaņas, lai students spēj izstrādāt datu bāzes piemērus ar NoSQL un NewSQL sistēmām.
Patstāvīgais darbs, tā organizācija un uzdevumi	Studiju kursa laikā studentiem jāizpilda četri praktiskie darbi. 1. Jāizveido datu bāze un jāizpilda vaicājumi, kas orientēti uz datu šķirošanu, datu grupēšanu un tabulu savienošanu izmantojot paralēlo apstrādi un bez tās. Jāveic iegūto rezultātu izvērtējums. 2. Jāizveido datu bāze un vairākas datu izgūšanas transakcijas (salikti vaicājumi). Jāuzzīmē transakciju apakšvaicājumu izpildes shēma, jāpārbauda tās realizācija praksē un jāizvērtē transakcijas izpildes problēmas. 3. Jāinstalē kāda no NoSQL DBS (MongoDB, Cassandra, ElasticSearch) un jārealizē datu glabāšanas struktūras definēšana, datu ievade un vaicājumu izpilde. Jāanalizē un jāpamato iegūtie rezultāti. 4. Jāinstalē kāda no NewSQL DBS (NuoDB, ClustrixDB, Altibase) un jārealizē datu glabāšanas struktūras definēšana, datu ievade un vaicājumu izpilde. Jāanalizē un jāpamato iegūtie rezultāti. Izmantojot iegūtos rezultātus, jāveic NewSQL un NoSQL DBS salīdzinājums.

Literatūra	Obligātā/Obligatory: 1.Avi Silberschatz, Henry F. Korth, S. Sudarshan. Database System Concepts (Seventh Edition), McGraw-Hill, 2019, 1100 p., ISBN 978-0-078-02215-9. Papildu/Additional: 1.David Taniar, Clement H. C. Leung, Wenny Rahayu, Sushant Goel. HighPerformance Parallel Database Processing and Grid Databases, John Wiley & Sons, 2008, 551 p., ISBN 978-0-470-10762-1. 2.Shivnath Babu, Herodotos Herodotou. Massively Parallel Databases and MapReduce Systems. Foundations and Trends in Databases. Vol.5, No. 1, 2012, 1 – 104 p. DOI: 10.1561/19000000036. 3.Guy Harrison. Next Generation Databases. NoSQL, NewSQL and Big Data. Apress, 2015, 235 p., ISBN-13: 978-1-48-42-1330-8. 4.Andreas Meier, Michael Kaufmann. SQL & NoSQL Databases: Models, Languages, Consistency Options and Architectures for Big Data Management. Springer, 2019, 248 p., ISBN-10 3658245484. 5.M. Tamer Özsu, Patrick Valduriez. Principles of Distributed Database Systems. Springer, 2020, 691 p., ISBN-10: 3030262553. 6.Darl Kuhn, Thomas Kyte. Expert Oracle Database Architecture: Techniques and Solutions for High Performance and Productivity. Apress, 2021, 1179 p., ISBN-10:1484274989. 7.Tushar Mahapatra, Sanjay Mishra. Oracle Parallel Processing. O'Reilly, 2000, 288 p., ISBN-10: 156592701X.
Nepieciešamās priekšzināšanas	Relāciju algebra, relāciju datu bāzes sistēmas, datora uzbūve, objektorientētā programmēšana, datu struktūras, SQL valoda, datu bāzes servera iekšējā programmēšanas valoda.

Studiju kursa saturs

Saturs	Pilna un nepilna laika klātienē studijas		Nepilna laika neklātienē studijas	
	Kontakt stundas	Patstāv. darbs	Kontakt stundas	Patstāv. darbs
Paralēlisma ieviešanas galvenie pamatjēdzieni.	4	4	0	0
Transakciju vadība.	2	2	0	0
Daudzlīmeņu transakcijas.	2	2	0	0
Intensīvā paralēlā apstrāde (massively parallel processing (MPP)).	2	2	0	0
Klasiskā paralēlo DBS kategorija (partīciju izmantošana un	2	2	0	0
MMP DBS ar MapReduce algoritma izmantošanu.	2	2	0	0
NoSQL datu bāzes sistēmu pamatkonceptija un pamatiezīmes.	2	2	0	0
Lasīšanas, rakstīšanas un datu modificēšanas konsekvence (consistency).	2	2	0	0
Semestra vidus kontroldarbs par paralēlisma idejām un transakcijām.	2	2	0	0
NoSQL DBS daudzveidība.	2	2	0	0
NoSQL datu bāzes projektēšana un realizēšana.	6	6	0	0
NewSQL DBS pamatiezīmes.	2	2	0	0
NewSQL datu bāzes sistēmu klases.	2	2	0	0
Atšķirības NoSQL and NewSQL datu bāzes sistēmās.	4	4	0	0
NewSQL datu bāzes sistēmu izpēte.	2	2	0	0
Gala eksāmens par NoSQL un NewSQL datu bāzu sistēmām.	2	2	0	0
Kopā:	40	40	0	0

Sasniedzamie studiju rezultāti un to vērtēšana

Sasniedzamie studiju rezultāti	Rezultātu vērtēšanas metodes
Izprot dažādas datu bāzes sistēmu arhitektūras ātrdarbības uzlabošanai un spēj tās realizēt praktiski.	Praktiskais darbs, attiecīgie jautājumi semestra vidus kontroldarbā.
Zina atšķirības starp transakciju apstrādi dažādās arhitektūrās un saprot no tā izrietošās priekšrocības un trūkumus.	Praktiskais darbs, attiecīgie jautājumi semestra vidus kontroldarbā.
Prot praktiski izmantot NoSQL DB sistēmas, zina to priekšrocības un trūkumus.	Praktiskais darbs, attiecīgie jautājumi gala eksāmenā.
Prot praktiski izmantot NewSQL datu bāzes sistēmas un zina to potenciālās priekšrocības un trūkumus.	Praktiskais darbs, attiecīgie jautājumi gala eksāmenā.

Studiju rezultātu vērtēšanas kritēriji

Kritērijs	% no kopējā vērtējuma
Praktisko darbu vidējā atzīme. Praktiskajā darbā students iegūst 4, ja ir ieguvis 50% no darbā paredzētajiem punktiem, izpildot minimālās prasības	50
Semestra vidus kontroldarbs paredz praktisku un teorētisku zināšanu pārbaudi. Students iegūst sekmīgu vērtējumu, ja ir saņēmis 50% punktu	25
Gala eksāmens paredz praktisku un teorētisku zināšanu pārbaudi. Students iegūst sekmīgu vērtējumu, ja ir saņēmis 50% punktu	25
Kopā:	100

Studiju kursa plānojums

Daļa	KP	Stundas			Pārbaudījumi		
		Lekcijas	Prakt d.	Laborat	Ieskaite	Eksām.	Darbs
1.	3.0	40.0	0.0	0.0		*	