

RTU studiju kurss "Ievads datu inženierijā"**33000 Datorzinātnes, informācijas tehnoloģijas un enerģētikas fakultāte****Vispārējā informācija**

Kods	DE1061
Nosaukums	Ievads datu inženierijā
Studiju kursa statuss programmā	Obligātais/Ierobežotās izvēles
Atbildīgais mācītspēks	Māra Romanovska - Doktors, Docents
Apjoms daļās un kredītpunktos	1 daļa, 6.0 kredītpunkti
Studiju kursa īstenošanas valodas	LV, EN
Anotācija	Dati, datu apstrāde un ar tiem cieši saistītā datu zinātne pēdējos gados ir ieguvusi arvien augstāku popularitāti. Tipiski datu zinātnes iesācēji saskaras ar situāciju, kad ir iegūta datu kopa un apzināti tās analīzes aspekti, tomēr trūkst zināšanu, lai datu kopu formalizētu un glabātu tālākai izmantošanai. It īpaši tas attiecas uz heterogēnām, jauktām un abstraktām datu kopām. Datu inženierija ir joma, kas pēta, apraksta un praksē demonstrē darbību kopumu, kas ļauj izveidot saskarnes un mehānismus, lai piekļūtu datiem laicīgi un efektīvi. Studiju kurss ir paredzēts studējošajiem, kuri strādā ar datiem, iegūst datu kopas un tālāk tās analizē dažādos griezumos. Studiju kurss koncentrējas uz trīs būtiskiem un savstarpēji saistītiem konceptiem: datu modelēšanu, datu izgūšanu no izveidotā modeļa un datu arhitektūras izveidi. Visiem trīs konceptiem tiek apskatīti tipiskākie risinājumi, uzsverot to piemērotību atkarībā no konkrētajiem datiem, un konkrētās datu kopas analīzes vajadzībām. Datu modelēšanai tiek apskatīta gan klasiskā entītijas relāciju pieeja kontekstā ar strukturētās vaicājumu valodas (SQL) datu bāzēm, gan arī pieejas, kas ļauj modelēt citas (sauktas par NoSQL) datu bāzes. Datu arhitektūrās tiek apskatītas centralizētās un necentralizētās pieejas ar uzsvaru uz datu noliktavām.
Mērķis un uzdevumi, izteikti kompetencēs un prasmēs	Studiju kursa mērķis ir attīstīt studējošā pamata zināšanas un prasmes datu inženierijas jomā tiktāl, lai students, iegūstot jaunu datu kopu vai darbojoties ar jaunu domēnu, ir spējīgs izstrādāt tai atbilstošu vieglu līdz vidēji sarežģītu datu inženierijas risinājumu. Studiju kursa uzdevumi: - radīt izpratni par datu inženierijas pamatjēdzieniem un to savstarpējo saistību, kā arī saistību ar kopējo datu zinātnes kontekstu; - radīt izpratni par dažādiem datu modeļiem un tiem atbilstošo datu sistēmu priekšrocībām un trūkumiem; - iemācīt praktiski modelēt datus, izmantojot realitāšu saišu pieeju vienkāršās līdz vidēji sarežģītās problēmjomās; - veidot praktiskās iemaņas darbā ar SQL tiktāl, ka students prot izgūt nepieciešamos datus; - radīt izpratni par dažādām datu arhitektūrām; - iemācīt praktiski strādāt ar datu noliktavām tiktāl, ka students prot izgūt nepieciešamos datus; - attīstīt spēju dokumentēt datu inženierijas risinājumus; - attīstīt spēju rakstiski izteikt savu viedokli un argumentus par labu dažādiem datu inženierijas risinājumiem.
Patstāvīgais darbs, tā organizācija un uzdevumi	Studiju kursa ietvaros studējošie patstāvīgi pilda trīs mājasdarbus. Pirmais mājasdarbs attiecas uz datu modelēšanu, tā ietvaros tiek izstrādāta realitāšu saišu diagramma un atbilstošie apraksti, balstoties uz mācībspēka definēto datu kopu. Studējošajam ir arī jāformulē dažādu modelēšanas pieeju piemērotība paša izvēlētai datu kopai. Otrais mājasdarbs ir saistīts ar SQL vaicājumu valodu un tās izmantošanu; tā ietvaros tiek praktiski realizēti vaicājumi mācībspēka iedotajai datu kopai un studenta izvēlētai datu kopai. Trešajā mājasdarbā tiek projektēta datu noliktava studenta izvēlētai datu kopai un tiek demonstrēta datu izguve no šīs datu arhitektūras. Studenti gatavoja testiem un eksāmenam.
Literatūra	Obligātā/ Obligatory: 1. Reis, Joe, and Matt Housley. Fundamentals of Data Engineering: Plan and Build Robust Data Systems. O'Reilly Media, 2022, pp. 450, O'Reilly Media, Inc. Papildu/ Additional: 1. Serra, James. Deciphering Data Architectures: Choosing Between a Modern Data Warehouse, Data Fabric, Data Lakehouse, and Data Mesh. O'Reilly Media, 2024, pp. 278. 2. Kleppmann, Martin. Designing data-intensive applications : the big ideas behind reliable, scalable, and maintainable systems. O'Reilly Media, , 2017, pp. 590.2. 3. Silberschatz, Abraham, Henry F. Korth, and S. Sudarshan. Database System Concepts. 7th ed., McGraw-Hill Education, 2020, pp. 1344.
Nepieciešamās priekšzināšanas	Datorprātības pamatzināšanas, programmēšanas pamatu izpratne.

Studiju kursa saturs

Saturs	Pilna un nepilna laika klātienes studijas		Nepilna laika neklātienes studijas	
	Kontakt stundas	Patstāv. darbs	Kontakt stundas	Patstāv. darbs
Ievads datu inženierijā un datu apstrāde kā datu inženierijas konteksts, datu krātuves jēdziens.	4	0	0	0
Datu inženierijas dzīvescikls. Datu inženierijas risinājuma sarežģītības novērtēšana.	4	0	0	0
Dažādi datu avoti, to priekšrocības un trūkumi.	2	2	0	0
Dažādas datu modelēšanas pieejas un dažādu aspektu nozīme, modelējot datus.	4	2	0	0
ER modelis kā dominējošā pieeja. ER modeļa komponentes.	4	6	0	0

Datu modeļa kvalitātes novērtēšana. Normalizācija.	4	8	0	0
Alternatīvas ER modelim - NoSQL pieejas.	4	6	0	0
Datu modeļa saistība ar tā realizāciju sistēmā.	2	2	0	0
Vaicājuma jēdziens. Vaicājumi dažādos datu modeļos.	2	2	0	0
SQL vaicājumi un to veidi. SQL vaicājumu apstrāde.	4	6	0	0
SQL vaicājumi. Vaicājumi ar vienu un vairākām tabulām.	4	8	0	0
SQL vaicājumi. Funkcijas un paplašinājumi, padziļināta sarežģītības līmeņa konstrukcijas.	4	8	0	0
Vaicājumi modeļiem, kas nebalstās uz entitiju relāciju modelēšanas pieeju. Piekārtošanas-reducēšanas pieeja.	4	4	0	0
Datu arhitektūru pārskats, balstoties uz dažādiem datu modeļiem. Centralizētas un decentralizētas datu arhitektūras.	4	2	0	0
Datu ezers, datu tīkls un citas aktuālās datu arhitektūras. To priekšrocības un trūkumi.	4	2	0	0
Datu noliktavas jēdziens. Datu noliktavas priekšrocības un trūkumi.	4	6	0	0
Darbs ar datu noliktavu. Dažādi datu noliktavas implementācijas varianti.	4	8	0	0
Vaicājumi ar datu noliktavu.	4	8	0	0
Decentralizētas datu arhitektūras un datu izgūšana no tām.	4	2	0	0
Datu arhitektūras tālākā izmantošana - datu pasniegšana.	4	2	0	0
Studiju kursa kopsavilkums, nākotnes virzieni.	4	0	0	0
Kopā:	78	84	0	0

Sasniedzamie studiju rezultāti un to vērtēšana

Sasniedzamie studiju rezultāti	Rezultātu vērtēšanas metodes
Zina dažādu datu inženierijas pamatjēdzienu savstarpējo saistību.	Vērtēšanas metodes: testi un eksāmens. Kritēriji: students spēj izskaidrot dažādus datu inženierijas jēdzienus, tos saistīt savā starpā, kā arī ar dažādām datu inženierijas risinājumu īpašībām, saprot, kuri datu inženierijas risinājumi ir piemēroti dažādos gadījumos.
Saprot, kuri datu modeļi un kuras datu arhitektūras ir attiecināmas dažādu datu kopu analīzei.	Vērtēšanas metodes: testi un eksāmens. Kritēriji: students zina dažādus datu modeļus un var tos saistīt ar dažādām datu arhitektūrām. Praktiski redzot datu kopu, students spēj to analizēt un pamatoti noteikt, kura datu arhitektūra un datu modelis ir piemērots tās realizācijai.
Prot modelēt problēmjomu ar realitāšu saišu pieeju vieglos līdz vidējo sarežģītos gadījumos. Sarežģītos gadījumos spēj atrast nepieciešamos resursus risinājuma izstrādei.	Vērtēšanas metodes: pirmais mājasdarbs. Kritēriji: students problēmsfēras aprakstā spēj identificēt realitāšu kopas, tās saistās saites, realitāšu kopu atribūtus. Students spēj tās saistīt ER diagrammā. Students zina par rīku vadītu problēmjomas modelēšanu un prot to izmantot. Students spēj izvērtēt un izskaidrot dažādas modelēšanas alternatīvas.
Spēj praktiski prot izmantot SQL vaicājumus, lai izgūtu nepieciešamos datus.	Vērtēšanas metodes: otrais mājasdarbs. Kritēriji: izmantojot esošu datu kopu, students prot veikt dažāda sarežģītības līmeņa vaicājumus. Students izprot, kā ar SQL izteiksmju palīdzību tiek izveidota datu bāze. Students spēj izvērtēt un izskaidrot dažādas SQL vaicājuma pieraksta alternatīvas.
Prot projektēt datu noliktavu un izgūt no tās datus.	Vērtēšanas metodes: trešais mājasdarbs. Kritēriji: students izprot, kas ir daudzdimensionāls datu modelis un kam to pielieto. Students spēj pārveidot datu modeli uz daudzdimensiju modeli jēgpilni un loģiski. Spēj veikt datu izgūšanu no modeļa, saprot, kas ir nepieciešams tā realizācijai. Vienkāršā gadījumā spēj izveidot datu noliktavas shēmu.

Studiju rezultātu vērtēšanas kritēriji

Kritērijs	% no kopējā vērtējuma
Pirmais tests: Datu modelēšana un datu inženierijas pamatjēdzieni.	5
Otrais tests: Datu izgūšana no datu modeļa.SQL.	5
Trešais tests: Datu arhitektūras.	5
Pirmais mājasdarbs: Datu modelēšanas ar entitiju relāciju pieeju. Datu modelēšanas ar entitiju relāciju pieeju. Students iegūst apmierinošu vērtējumu (4), ja mājas darbā iegūti vismaz 50% punktu. Lai nokārtotu kursu, vismaz 2 no 3 mājas darbiem jānoliek, tajos iegūstot apmierinošu vērtējumu.	20
Otrais mājasdarbs: Datu izgūšana ar SQL. Datu modelēšanas ar entitiju relāciju pieeju. Students iegūst apmierinošu vērtējumu (4), ja mājas darbā iegūti vismaz 50% punktu. Lai nokārtotu kursu, vismaz 2 no 3 mājas darbiem jānoliek, tajos iegūstot apmierinošu vērtējumu.	20
Trešais mājasdarbs: Datu noliktavas. Students iegūst apmierinošu vērtējumu (4), ja mājas darbā iegūti vismaz 50% punktu. Lai nokārtotu kursu, vismaz 2 no 3 mājas darbiem jānoliek, tajos iegūstot apmierinošu vērtējumu.	15
Gala eksāmens: Students iegūst apmierinošu vērtējumu (4), ja eksāmenā iegūti vismaz 50%. Lai nokārtotu kursu, eksāmens jānoliek, tajā iegūstot apmierinošu vērtējumu.	30

Kopā:	100
-------	-----

Studiju kursa plānojums

Daļa	KP	Stundas			Pārbaudījumi		
		Lekcijas	Prakt d.	Laborat	Ieskaite	Eksām.	Darbs
1.	6.0	40.0	38.0	0.0		*	