



SD(M) pavadinimas	Course title
Didieji duomenys	Big Data

SD(M) priklausomybė studijų pakopai

Course subjection to study level

Studijos: Studies:	B – Pirmosios pakopos First cycle
-----------------------	--------------------------------------

SD(M) priklausomybė studijų programai

Course subjection to programme

SD(M) priklausomybė studijų krypčių ir krypčių grupei

The list of study fields and groups of fields

SD(M) priklausomybė dalykų grupei Course subjection to group	1 – studijų dalyko Course	Studijų krypčių grupės kodas Code of the group of study fields	Studijų krypties kodas Code of the study field
SD(M) priklausomybė programos daliai Course subjection to part of the programme	C – Specializacijos dalykų dalis Part of Specialization Subjects		
Struktūrinė SD priklausomybė Course structural subjection	K – katedros Department	A	A01

*) **Grupė:** *) 1 - studijų dalyko; 2 - praktikos; 3 - baigiamojo darbo ar projekto; 4 - baigiamojo egzamino; 5 - tiriamojo darbo; 6 - profesinio testavimo; 7 - kitas.

**) A - Bendrųjų universitetinių studijų; B - Studijų krypties; C - Specializacijos.

***) U - universiteto; F - fakulteto; K - katedros.

*) **Group:** *) 1 - Course; 2 - Practice; 3 - Final Work or Project; 4 - Final Examination; 5 - Research Work; 6 - Professional Testing; 7 - Other.

**) A - General; B - Field; C - Specialization.

***) U - University; F - Faculty; K - Department.

SD(M) kodas

Course number

Fakultetas Faculty	Katedra Department	Pakopa Study cycle	Modulio Nr. Number
F	M	B	16508

SD(M) kreditai

Course volume in credits

Iš viso: Total:	Iš jų: KD, KS, KP, PR There out:
3	0

SD(M) Atsiskaitymo forma

Course assessment

I, E1, E2, E, BE, BD, TD, A	KD, KS, KP, PR
E	-

*) B - pirmoji pakopa; A - vientisosios studijos; M - antroji pakopa.

*) B - first cycle studies; A - integrated studies; M - second cycle studies.

SD(M) valandų paskirstymas pagal studijų formas ir būdus

Distribution of course hours by study forms and ways

Studijų forma Study form	Valandos Hours								Kontaktinių Contact
	Kodas Code	Studijų būdas Study way	Paskaitoms Lectures	Lab. darbams Laboratory works	Pratyboms Practical works	Konsultacijoms Consultation	Sav. darbui Independent work	Iš viso Total	
Nuolatinės studijos Full-time studies	NL	S	30	15	0	2	33	80	47

*) Studijų būdas: S - semestrais; M - moduliais; C - ciklais; T - nuotolinis; NI - neakivaizdinis intensyvusis.

*) Study process forms: S - semesters; M - modules; C - periods; T - distance; NI - part-time.

SD(M) ANOTACIJA

Studijų dalyką sudaro didžiųjų duomenų (Big Data) pagrindai, supažindinama su pagrindinėmis duomenų apdorojimo programavimo modeliais ir technologijomis. Sudijų metu supažindinama su pagrindinėmis didžiųjų duomenų sąvokomis. Apžvelgiami su didžiųjų duomenų apdorojimu susieti iššūkiai bei sistemų reikalavimai reikalingi jiems spręsti. Kurso metu dirbama su didžiųjų duomenų apdorojimui įrankiais tokiais kaip Hadoop ir Spark. Taip pat nagrinėjami įvairūs realūs panaudojimo uždaviniai.

Studentai numatytu tvarkaraštyje metu privalo dalyvauti ne mažiau kaip 50 proc. teorinių paskaitų ir atlikti ne mažiau kaip

80 proc. laboratorinių darbų.

ANNOTATION OF COURSE

The subject of the studies consists of the basics of big data, introduction to fundamental data processing programming models and technologies. During the studies, the fundamental concepts of big data are introduced. Challenges related to big data processing and system requirements necessary to solve them are overviewed. The course involves working with big data processing tools such as Hadoop and Spark. Various real-world use cases are also examined.

Students must attend at least 80% of the time scheduled laboratory works and 50% of the lectures.

SD(M) TIKSLAS

Dalyko tikslas yra suteikti pagrindines žinias apie didžiuosius duomenis (Big Data) bei jų apdorojimo technologijas.

AIM OF COURSE

The aim of course is to provide the main knowledge about the big Data and the Big Data processing technologies.

Studento pasiekimų vertinimo formulė

Galutinio įvertinimo pažymio G formulė:

$$G = SE \times 0,5 + TE \times 0,2 + L \times 0,3$$

čia: SE - sesijos egzamino pažymys; L - laboratorinių darbų suminis pažymys; TE - tarpinio egzamino pažymys

Assessments methods of students formula

The final assessment mark E is calculated as:

$$E = SE \times 0,5 + TE \times 0,2 + L \times 0,3$$

where SE is the mark of the session exam; L is the total mark of laboratory works; TE is the mark of the intermediate exam.

Pagrindinė literatūra (ne daugiau kaip 5 šaltiniai):

Main references (not more than 5 references)

Eil. Nr. <i>No.</i>	Leidinio autoriai ir pavadinimas (elektroninių leidinių ir žiniatinklio adreso) <i>Authors and title (site address in case of e-publication)</i>
1.	Big data computing / edited by Rajendra Akerkar. Boca Raton, FL : CRC Press, Taylor & Francis Group, 2014
2.	R. Glass, S. Callahan. The Big Data-Driven Business: How to Use Big Data to Win Customers, Beat Competitors, and Boost Profits. John Wiley & Sons. Hoboken: New Jersey, 2015.
3.	J. Dean. Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners. John Wiley & Sons. Hoboken: New Jersey, 2014.
4.	J. Leskovec. Mining of massive datasets 2nd ed. Cambridge: Cambridge University Press, 2015.

*) Kortelės pildymo metu

*) *At the form filling moment*

Papildoma literatūra (ne daugiau kaip 10 šaltinių):

Additional references (not more than 10 references)

Eil. Nr. <i>No.</i>	Leidinio autoriai ir pavadinimas (elektroninių leidinių ir žiniatinklio adreso) <i>Authors and title (site address in case of e-publication)</i>
1.	EMC Education Services. Data Science and Big Data Analytics
2.	R. Smolan, J. Erwit. The Human Face of Big Data.
3.	V. Mayer-Schönberger. Big Data: A Revolution That Will Transform How We Live, Work, and Think. John Murray (Publishers), London, 2013.

*) Kortelės pildymo metu

*) *At the form filling moment*

Reikalingi IT resursai * (nurodyti 1-3 alternatyvas, pageidautina, kad bent 1 būtų nemokama)

Required IT Resources

Eil. Nr. <i>No.</i>	Programinės įrangos pavadinimas, gamintojas <i>Name of the software, manufacturer</i>	Licencijos tipas (pagal įsigijimo būdą) <i>License type</i>
1	Hortonworks	Nemokama Unpaid

*) Pildoma, jei tokie resursai reikalingi. Stulpelyje Licencijos tipas pasirenkamas iš sąrašo:

Mokama, akademinė

Mokama, komercinė

Nemokama

*) Should be completed if such reassures are needed. License type - select from the list:
Paid, academic
Paid, commercial

Savarankiško darbo turinys

Content of individual work

Užduoties pavadinimas <i>Assignment title</i>	Sav. darbo apimtis vienai užduočiai <i>Amount of hours of independent work for a single task</i>					Užduočių skaičius <i>Number of tasks</i>					Iš viso valandų <i>Total hours</i>					Įvertinimo dalis % <i>Part of Evaluation %</i>								
	Rekomenduojamos val. <i>Recommended hours</i>	Skirta val. <i>Separated hours</i>					NL (T)	NL (S)	NL (Sav.)	I(S)	I(T)	NL(T)	NL(S)	NL (Sav.)	I(S)	I(T)	NL(T)	NL(S)	NL (Sav.)	I(S)	I(T)			
		NL(T)	NL(S)	NL (Sav.)	I(S)	I(T)																		
Laboratorinis darbas Laboratory work	2-12		5					2					10						30					
Kolokviumas Intermediate examination	8-27		10					1					10						20					
Pasirengimas atsiskaitymui Preparation for evaluation	10-60		13					1					13											
Iš viso: <i>Total:</i>													33											

*) Papildomas laukas pildomas tik tada, kada taikomas SD(M) kortelėje nenurodytas studijų būdas: M - moduliai; C - ciklais; T - nuotoliniu

*) Must be used in case study way does not fall into standard category: M - modules; C - periods; T - distance

Savarankiško darbo grafikas

Individual work schedule

Užduoties tipas <i>Task type</i>	Užduoties pateikimo(*) ir atsiskaitymo(+) savaitė <i>Week of Assignment setting (*) and assessment(+)</i>																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Nuolatinės studijos (S) <i>Full-time studies</i>																				
Laboratorinis darbas <i>Laboratory work</i>	*			1			2													
	+				1				2											
Kolokviumas <i>Intermediate examination</i>	*				1															
	+							1												

Laboratorinių darbų sąrašas

List of the Course laboratory work

Temos pavadinimas <i>Topic title</i>	Valandų skaičius <i>Number of hours</i>				
	NL(T)	NL(S)	NL(Sav.)	I(S)	I(T)
1. Map-Reduce užduočių įgyvendinimas naudojant Hortonworks platformą <i>The implementation of Map-Reduce problems using Hortonworks platform.</i>		5			
2. Spark sintaksės demonstracija ir bandymai. <i>Demonstration and trials of the Spark syntax.</i>		5			
3. Spark SQL ir ML bibliotekų isisavinimas. <i>Assimilation of Spark SQL and ML libraries.</i>		5			
Iš viso: <i>Total:</i>		15			

Paskaitų temų sąrašas

List of the Course lecture topics

Temos pavadinimas <i>Topic title</i>	Valandų skaičius <i>Number of hours</i>				
	NL(T)	NL(S)	NL(Sav.)	I(S)	I(T)
1. Įvadas į didžiuosius duomenis (Big Data). <i>The introduction to Big Data.</i>		6			
2. Didžiųjų duomenų sistemų apžvalga ir palyginimas. <i>The overview of Big Data systems and their comparison.</i>		6			
3. Programavimo modeliai skirti didžiųjų duomenų apdorojimui. <i>The programming models for big data processing.</i>		6			
4. Supažindinimas su didžiųjų duomenų apdorojimo įrankiais: Hadoop MapReduce, Apache Spark . <i>Introduction to the tools for Big Data processing: Hadoop MapReduce, Apache Spark.</i>		6			

Paskaitų temų sąrašas
List of the Course lecture topics

Temos pavadinimas <i>Topic title</i>	Valandų skaičius <i>Number of hours</i>				
	NL(T)	NL(S)	NL(Sav.)	I(S)	I(T)
5. Apache Spark ekosistema: Spark Core, Spark SQL, Spark Streaming, Spark MLlib, Spark GraphX, SparkR. <i>Apache Spark ecosystem: Spark Core, Spark SQL, Spark Streaming, Spark MLlib, Spark GraphX, SparkR.</i>		6			
Iš viso: <i>Total:</i>		30			

*) Papildomas laukas pildomas tik tada, kada taikomas SD(M) kortelėje nenurodytas studijų būdas: M - moduliai; C - ciklais; T - nuotolinis

*) *Must be used in case study way does not fall into standard category: M - modules; C - periods; T - distance*

Fundamentinių mokslų fakulteto *Modernių technologijų matematikos (612G12001) 2016-07-01* programos studijų rezultatų sąsajos su SDM rezultatais bei studijų ir studentų pasiekimų vertinimo metodais

Links of the Mathematics of Modern Technologies (612G12001) of the Faculty of Fundamental Sciences with the course unit and evaluation methods of students achievements

Programos studijų rezultatai <i>Study programme outcomes</i>	SD(M) rezultatai <i>Course results</i>	Studijų metodai <i>Methods of studies</i>	Studento pasiekimų vertinimo metodai <i>Evaluation methods of student achievements</i>	Studentų pasiekimų vertinimo kriterijai pagal lygmenis <i>Assessments criteria of students achievements by Assessment levels</i>
GT2. Gebės surinkti duomenis ir juos analizuoti, taikys įvairių jų peržiūros algoritmus, siūlys naujus technologinius sprendimus, kurs naujus matematinius modelius ir tobulins jau sukurtus. GT2. Will be able to collect and analyze data, apply their various review algorithms, offer new technological solutions, develop new mathematical models and improve the existing ones.	Gebės praktinio uždavinio formuluotę susieti su tipiniais didžiųjų duomenų uždaviniais, išskirti ir išnaudoti savybes leidžiančias efektyviai išspręsti uždavinį. Will be able to connect the practical problem to typical Big Data problems, identify and utilize the properties which let to solve the problem efficiently.	Teorinės paskaitos, literatūros studijos, savarankiškas darbas. Theoretical lectures, study of literature, independent work.	Tarpinis egzaminas, egzaminas. Intermediate examination, final examination	Slenkstinis (5-6): studentas žino pagrindines dalyko sąvokas ir moka jomis naudotis sprendžiamas paprasčiausias užduotis. Tipinis (7-8): studentas žino pagrindines sąvokas, gali jas paaiškinti ir teorinius faktus geba tinkamai taikyti atlikdamas praktinių ir teorinių užduočių analizę, skaičiavimams taiko kompiuterines programas. Puikusias (9-10): studentas yra ne tik pasiekęs tipinį lygmenį, bet įgytas žinias geba taikyti naujose situacijose. Threshold (5-6): the student knows the basic definitions and using example can analyze a simple problems. Typical (7-8): the student knows the basic definitions, is able to explain it and to apply theoretical facts for the problem analysis and use the computer programs for calculations. Excellent (9-10): the student is not only achieving typical level, but also is able to adapt the knowledges for new situation.

SG2. Gebės valdyti ir apdoroti didelius duomenų srautus, taikyti tinkamus algoritmus ir technologijas. SG2. Will be able to manage and process large data streams, apply the right algorithms and technologies.	Gebės apdoroti didelius duomenų kiekius, taikyti tinkamus algoritmus ir technologijas. Will be able to process large amount of data, apply the right algorithms and technologies.	Teorinės paskaitos, literatūros studijos, savarankiškas darbas ir laboratoriniai darbai. Theoretical lectures, study of literature, individual work, laboratory works.	Kolokviumas, kontrolinis darbas, laboratoriniai darbai, egzaminas. Laboratory works, intermediate examination, final examination	Slenkstinis (5-6): studentas žino pagrindines dalyko sąvokas ir moka jomis naudotis sprendžiamas paprasčiausias užduotis. Tipinis (7-8): studentas žino pagrindines sąvokas, gali jas paaiškinti ir teorinius faktus geba tinkamai taikyti atlikdamas praktinių ir teorinių užduočių analizę, skaičiavimams taiko kompiuterines programas. Puikūs (9-10): studentas yra ne tik pasiekęs tipinį lygmenį, bet įgytas žinias geba taikyti naujose situacijose. Threshold (5-6): the student knows the basic definitions and using example can analyze a simple problems. Typical (7-8): the student knows the basic definitions, is able to explain it and to apply theoretical facts for the problem analysis and use the computer programs for calculations. Excellent (9-10): the student is not only achieving typical level, but also is able to adapt the knowledges for new situation.
---	---	--	--	--

SD(M) sudarytojas (-ai) (parašas, vardas ir pavardė)

Course compiled by (full name, signature)

Andrej Bugajev

Katedros vedėjas (parašas, vardas ir pavardė)

Head of Department (full name, signature)

Raimondas Čiegis

SD(M) atestuojamas <i>The Course is certified</i>		
SD(M), skirtas studijų programai: <i>The Course for the programme of studies:</i>	Modernių technologijų matematika Mathematics of Modern Technologies	
SD(M) atestacija galioja: <i>Course certification is valid:</i>	nuo <i>from</i>	iki <i>till</i>

SD(M) atestavo <i>the Course certified by</i>	Fundamentinių mokslų fakulteto studijų komitetas Faculty of Fundamental Sciences Study Committee		
Fakulteto studijų komiteto pirmininkas (vardas ir pavardė, parašas) <i>Chairman of the Studies committee (full name, signature)</i>		Data <i>Date</i>	