

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ЖЕЛЕЗНОДОРОЖНОГО ТРАНСПОРТА
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Иркутский государственный университет путей сообщения»
(ФГБОУ ВО ИрГУПС)

УТВЕРЖДЕНА
приказ ректора
от «07» июня 2021 г. № 78

Б1.О.58 Обработка и анализ больших данных

рабочая программа дисциплины

Специальность – 10.05.03 Информационная безопасность автоматизированных систем

Специализация – № 5 "Безопасность открытых информационных систем"

Квалификация выпускника – Специалист по защите информации

Форма и срок обучения – 5л 6м очная форма

Кафедра-разработчик программы – Информационные системы и защита информации

Общая трудоемкость в з.е. – 3
Часов по учебному плану (УП) – 108

Формы промежуточной аттестации в семестрах/на курсах
очная форма обучения:
экзамен А

Очная форма обучения

Распределение часов дисциплины по семестрам

Семестр	А	Итого
Число недель в семестре	16	
Вид занятий	Часов по УП	Часов по УП
Аудиторная контактная работа по видам учебных занятий/ *	51	51
– лекции	17	17
– практические (семинарские)	17	17
– лабораторные	17	17
Самостоятельная работа	21	21
Экзамен	36	36
Итого	108	108

ИРКУТСК

Электронный документ выгружен из ЕИС ФГБОУ ВО ИрГУПС и соответствует оригиналу

Подписант ФГБОУ ВО ИрГУПС Трофимов Ю.А.

00a73c5b7b623a969ccad43a81ab346d50 с 08.12.2022 14:32 по 02.03.2024 14:32 GMT+03:00

Подпись соответствует файлу документа



Рабочая программа дисциплины разработана в соответствии с федеральным государственным образовательным стандартом высшего образования – специалитет по специальности 10.05.03 Информационная безопасность автоматизированных систем, утвержденным приказом Минобрнауки России от 07.06.2021 №78.

Программу составил(и):

д.т.н., профессор, профессор,

_____ Е.И. Молчанова

Рабочая программа рассмотрена и одобрена для использования в учебном процессе на заседании кафедры «Информационные системы и защита информации», протокол от «04» 06 2021 г. № 11/2

Заведующий кафедрой, к.э.н., доцент

_____ Т.К. Кирилова

1 ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ	
1.1 Цели дисциплины	
1	Формирование профессиональных компетенций, необходимых для получения фундаментальных знаний в области систем и технологий обработки больших данных.
2	Овладение умениями и способами деятельности в процессе освоения программных средств, используемых при обработке огромных объемов данных.
1.2 Задачи дисциплины	
1	Изучить технологии хранения, обработки и анализа больших данных.
2	Обеспечить сознательное и прочное овладение обучающимися теоретическими основами современных информационных технологий получения, хранения, обработки, анализа и визуализации больших объемов данных, систематизировать знания в данной области.
3	Создать условия для формирования умений и овладения обучающимися способами деятельности сознательного и рационального использования систем и технологий обработки больших данных.
1.3 Цель воспитания и задачи воспитательной работы в рамках дисциплины	
<p>Цель воспитания обучающихся – разностороннее развитие личности будущего конкурентоспособного специалиста с высшим образованием, обладающего высокой культурой, интеллигентностью, социальной активностью, качествами гражданина-патриота.</p> <p>Задачи воспитательной работы с обучающимися:</p> <ul style="list-style-type: none"> – развитие мировоззрения и актуализация системы базовых ценностей личности; – приобщение студенчества к общечеловеческим нормам морали, национальным устоям и академическим традициям; – воспитание уважения к закону, нормам коллективной жизни, развитие гражданской и социальной ответственности как важнейшей черты личности, проявляющейся в заботе о своей стране, сохранении человеческой цивилизации; – воспитание положительного отношения к труду, развитие потребности к творческому труду, воспитание социально значимой целеустремленности и ответственности в деловых отношениях; – обеспечение развития личности и ее социально-психологической поддержки, формирование личностных качеств, необходимых для эффективной профессиональной деятельности; – выявление и поддержка талантливых обучающихся, формирование организаторских навыков, творческого потенциала, вовлечение обучающихся в процессы саморазвития и самореализации 	

2 МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП	
2.1 Требования к предварительной подготовке обучающегося	
Б1.О.07 Математический анализ Б1.О.08 Алгебра и геометрия Б1.О.09 Дискретная математика Б1.О.10 Математическая логика и теория алгоритмов Б1.О.11 Теория вероятностей и математическая статистика Б1.О.12 Численные методы и теория оптимизации Б1.О.27 Основы кибернетики Б1.О.40 Разработка и эксплуатация автоматизированных систем в защищенном исполнении Б1.О.39 Программно-аппаратные средства защиты информации	
2.2 Дисциплины и практики, для которых изучение данной дисциплины необходимо как предшествующее	
1	Б3.01(Д) Подготовка к процедуре защиты выпускной квалификационной работы
2	Б3.02(Д) Защита выпускной квалификационной работы

3 ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫЕ С ТРЕБОВАНИЯМИ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ		
Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Планируемые результаты обучения

ОПК-2 Способен применять программные средства системного и прикладного назначений, в том числе отечественного производства, для решения задач профессиональной деятельности	ОПК-2.1 Применяет современные программные средства системного и прикладного назначений, в том числе отечественного производства, при решении задач профессиональной деятельности	<p>Знать:</p> <p>–стандарты обработки и анализа больших данных, и требования, связанные с созданием и использованием SQL и NoSQL систем хранения и обработки данных.</p> <p>Уметь:</p> <p>–использовать современные инструментальные и вычислительные средства, осуществлять постановку задач анализа данных, визуализацию интерпретацию результатов.</p> <p>Владеть:</p> <p>–технологиями хранения и обработки больших данных; –современными программными средствами анализа больших объемов информации;</p>
	ОПК-2.2 Знает основы программных средств системного и прикладного значения, в том числе отечественного производства	
ОПК-3 Способен использовать математические методы, необходимые для решения задач профессиональной деятельности	ОПК-3.1 Знает и имеет навыки применения основ математического анализа, алгебры, теории вероятностей и математической статистики, дискретной математики, математической логики и теории алгоритмов, теории автоматов и формальных языков	<p>Знать:</p> <p>–методы решения задач обработки и анализа больших данных, возможности высокопроизводительных вычислительных систем, технологии распределенных вычислений, методы и модели Data Mining.</p> <p>Уметь:</p> <p>–разрабатывать и анализировать концептуальные и теоретические модели прикладных задач анализа больших данных; – использовать и применять углубленные знания в области обработки и анализа больших данных; – создавать алгоритмы анализа и обработки большого объема данных с применением моделей Data Mining.</p> <p>Владеть:</p> <p>–способностью собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов; –методами статистического и математического анализа больших данных.</p>
	ОПК-3.2 Умеет использовать типовые математические методы и модели для решения задач профессиональной деятельности	
	ОПК-3.3 Владеет подходами к решению стандартных математических задач, выполнению расчетов математических величин, применению математических методов обработки экспериментальных данных для решения задач профессиональной деятельности	

4 СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Код	Наименование разделов, тем и видов работы	Семестр	Часы				*Код индикатора достижения компетенции
			Лек	Пр	Лаб	СР	
1.	Раздел 1. Большие данные. Основы систем больших данных. Программные платформы и системы для больших данных	А	13	13	11	13	ОПК-2.1; ОПК-2.2; ОПК-3.1; ОПК-3.2; ОПК-3.3
1.1.	Понятие «большие данные». Характеристики больших данных. Принципы работы с большими	А	2				

	данными. Большие данные в научных исследованиях. Экосистема больших данных (лекция)						
	Выбор инструментария для работы с BigData (практика)	A		4			
	Начало работы. Консолидация данных. Понятие сценария и узла обработки. (лабораторная работа)	A			2		
1.2.	Технологии хранения больших данных. Распределенные файловые системы. Распределенные фреймворки. Хранилища, классификация и примеры. (лекция)	A	2				
1.3.	Многомерное представление данных (лабораторная работа)	A			3		
1.4.	Архитектура системы и алгоритмы обработки больших данных. Уровни в системах обработки больших данных. Прием данных (Data Ingestion). Сбор данных (Data Staging). Анализ данных (Analysis Layer). Представление результатов (Consumption Layer). Методы преобразования данных. Семантические анализаторы. (лекция)	A	3				
1.5.	Семантический анализ текста онлайн, seo анализ текста (практика)	A		3			
1.6.	Работа с базой данных NoSQL (лабораторная работа)	A			2		
1.7.	Параллельные алгоритмы для работы с данными. Операторы Map и Reduce. Обработка данных в реальном времени. Лямбда-архитектура. Каппа-архитектура. (лекция)	A	3				
1.8.	Разработка запросов на выборку информации на основе mapreduce. (практика)	A		3			
1.9.	Извлечение набора данных по заданной тематике (лабораторная работа)	A			2		
1.10.	Системы управления большими данными. Аналитические платформы. Состав и возможности программного комплекса Apache Hadoop. (лекция)	A	3				
1.11.	Изучение и конфигурирование программного комплекса Apache Hadoop. Размещения набора данных по заданной тематике (практика)	A		3			
1.12.	Языки поисковых запросов для Hadoop. (лабораторная работа)	A			2		
3.	Раздел №2 Технологии анализа больших данных	A	4	4	6	8	ОПК-2.1; ОПК-2.2; ОПК-3.1; ОПК-3.2; ОПК-3.3
3.1.	Жизненный цикл анализа больших данных, стандарты. Технологии KDD и Data Mining. (лекция)		4				
3.2.	Общая схема анализа. Этапы моделирования. Процесс построения моделей. (практика)			4			
3.3.	Ассоциативные правила (лабораторная работа)				1		
3.4.	Кластеризация (лабораторная работа)				2		

3.5.	Классификация. Прогноз. (лабораторная работа)				3		
------	--	--	--	--	---	--	--

* Код индикатора достижения компетенции проставляется или для всего раздела, или для каждой темы или для каждого вида работы.

**5 ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ
ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ
АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ**

Фонд оценочных средств для проведения текущего контроля успеваемости и промежуточной аттестации по дисциплине: оформлен в виде приложения № 1 к рабочей программе дисциплины и размещен в электронной информационно-образовательной среде Университета, доступной обучающемуся через его личный кабинет

**6 УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ
ДИСЦИПЛИНЫ**

6.1 Учебная литература

6.1.1 Основная литература

	Авторы, составители	Заглавие	Издательство, год издания	Кол-во экз. в библиотеке/ 100% онлайн
6.1.1.1	Дадян, Э. Г.	Методы, модели, средства хранения и обработки данных : учебник	М. : Вузовский учебник : ИНФРА-М, 2018 — URL: https://znanium.com/catalog/product/961470 (дата обращения: 29.06.2021)	100% онлайн

6.1.2 Дополнительная литература

	Авторы, составители	Заглавие	Издательство, год издания	Кол-во экз. в библиотеке/ 100% онлайн
6.1.2.1	авт.-сост. Н. Ю. Братченко	Распределенные базы данных : учебное пособие	Северо-Кавказский федеральный университет. – Ставрополь : Северо-Кавказский Федеральный университет (СКФУ), 2015.	100% онлайн
6.1.2.2	Бродовская, Е. В.	Большие данные в исследовании политических процессов : учебное пособие: [16+] / Е. В. Бродовская, А. Ю. Домбровская	Министерство науки и высшего образования Российской Федерации, Московский педагогический государственный университет. – Москва : Московский педагогический государственный университет (МПГУ), 2018. – 88 с. : схем., табл., ил. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=563578	100% онлайн

6.1.3 Учебно-методические разработки (в т. ч. для самостоятельной работы обучающихся)

	Авторы, составители	Заглавие	Издательство, год издания/ Личный кабинет обучающегося	Кол-во экз. в библиотеке/ 100% онлайн
6.1.3.1	Siegel E.	Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die. Predictive	Online Digital Library "Books24x7", 2013[Electronic Resource] / Eric Siege – Hoboken, N.J: Wiley, 2013. – 320 p. - Authorized access: http://library.books24x7.com/toc.aspx?bookid=52840	100% онлай

		Analytics		
6.1.3.2	Cavanillas, J	Cavanillas, J. New Horizons for a Data- Driven Economy.	Online Digital Library "Springer eBooks", 2016 Authorized access: http://link.springer.com/10.1007/978-3-319-21569-3	100% онлайн
6.2 Ресурсы информационно-телекоммуникационной сети «Интернет»				
6.2.1	https://openedu.ru/course/misis/ABD/ - Открытое образование, Каталог курсов, MOOK: Введение в инженерию больших данных			
6.2..2	Электронная база данных «ZBMATH – The database Zentralblatt MATH» https://zbmath.org/			
6.3 Программное обеспечение и информационные справочные системы				
6.3.1 Базовое программное обеспечение				
6.3.1.1	ОС Microsoft Windows 7 Professional, лицензия № 49379844, обновление - контракт №0334100010018000027-0000756-02 от 28.05.2018 АО СофтЛайн Трейд, обновление - контракт № 0334100010019000029-0000756-02 от 17.09.2019г. АО СофтЛайн Трейд, контракт № 0334100010020000010-0000756-02 от 16.06.2020 АО СофтЛайн Трейд Windows Edu Per Device 10 Education, Соглашение № V6760694, обновление - контракт № 0334100010020000010-0000756-02 от 16.06.2020 АО СофтЛайн Трейд			
6.3.1.2	Офисный пакет Microsoft Office 2010, Лицензия № 48288083, обновление - контракт №0334100010018000027-0000756-02 от 28.05.2018 АО СофтЛайн Трейд, обновление - контракт № 0334100010019000029-0000756-02 от 17.09.2019г. АО СофтЛайн Трейд, обновление - контракт № 0334100010020000010-0000756-02 от 16.06.2020 АО СофтЛайн Трейд; Office Professional 2019 - Соглашение № V0709762, контракт № 0334100010020000010-0000756-02 от 16.06.2020 АО СофтЛайн Трейд; LibreOffice v. 5.2, свободно распространяемое ПО, https://ru.libreoffice.org			
6.3.2 Специализированное программное обеспечение				
6.3.2.1	Apache Hadoop - свободно распространяемый набор утилит, библиотек и фреймворк для разработки и выполнения распределённых программ, работающих на кластерах из сотен и тысяч узлов https://hadoop.apache.org/releases.html			
6.3.2.2	Apache Spark - фреймворк с открытым исходным кодом для реализации распределённой обработки неструктурированных и слабоструктурированных данных, входящий в экосистему проектов Hadoop https://spark.apache.org/			
6.3.2.3	Платформа Deductor Academic			
6.3.3 Информационные справочные системы				
6.3.3.1	Nature Publishing Group [Электронный ресурс] : реферативная база данных. - Режим доступа: http://www.nature.com/siteindex/index.htm			
6.3.3.2	SCOPUS [Электронный ресурс] : реферативная база данных / компания Elsevier. – Режим доступа: https://www.scopus.com/			
6.4 Правовые и нормативные документы				
6.4.1	Информационная справочная система: СПС КонсультантПлюс http://www.consultant.ru/			

7 ОПИСАНИЕ МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ОСУЩЕСТВЛЕНИЯ УЧЕБНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ	
1	Корпуса А, Б, В, Г, Д, Е ИрГУПС находятся по адресу г. Иркутск, ул. Чернышевского, д. 15; корпус Л – по адресу г. Иркутск, ул. Лермонтова, д.80.
2	Учебные аудитории для проведения занятий лекционного типа, занятий семинарского типа, курсового проектирования (выполнения курсовых проектов, работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, укомплектованные специализированной мебелью и техническими средствами обучения (ноутбук, проектор, экран), служащими для представления учебной информации большой аудитории. Для проведения занятий лекционного типа имеются учебно-наглядные пособия (презентации, плакаты, таблицы), обеспечивающие тематические иллюстрации содержания дисциплины. Помещения для хранения и профилактического обслуживания учебного оборудования – А-521.
3	Учебные залы вычислительной техники для занятий лабораторного типа, выполнения практических работ, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенные компьютерной техникой, подключенной к информационно-телекоммуникационной сети

	«Интернет», и обеспеченные доступом в электронную информационно-образовательную среду ИрГУПС.
4	Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой, подключенной к информационно-телекоммуникационной сети «Интернет», и обеспечены доступом в электронную информационно-образовательную среду ИрГУПС. Помещения для самостоятельной работы обучающихся: – читальные залы; – учебные залы вычислительной техники А-401, А-509, А-513, А-516, Д-501, Д-503, Д-505, Д-507.

8 МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Вид учебной деятельности	Организация учебной деятельности обучающегося
Лекция	<p>Лекция (от латинского «lection» – чтение) – вид аудиторных учебных занятий. Лекция: закладывает основы научных знаний в систематизированной, последовательной, обобщенной форме; раскрывает состояние и перспективы развития соответствующей области науки и техники; концентрирует внимание обучающихся на наиболее сложных, узловых вопросах; стимулирует познавательную активность обучающихся.</p> <p>Во время лекционных занятий обучающийся должен уметь сконцентрировать внимание на изучаемых проблемах и включить в работу все виды памяти: словесную, образную и моторно-двигательную. Для этого весь материал, излагаемый преподавателем, обучающемуся необходимо конспектировать. В конспект рекомендуется выписывать определения, формулировки и доказательства теорем, формулы и т.п. На полях конспекта следует пометить вопросы, выделенные обучающимся для консультации с преподавателем. Выводы, полученные в виде формул, рекомендуется в конспекте подчеркивать или обводить рамкой, чтобы лучше запоминались. Полезно составить краткий справочник, содержащий определения важнейших понятий и наиболее часто употребляемые формулы дисциплины. К каждой лекции следует разобрать материал предыдущей лекции. Изучая материал по учебнику или конспекту лекций, следует переходить к следующему вопросу только в том случае, когда хорошо усвоен предыдущий вопрос. При этом необходимо воспроизводить на бумаге все рассуждения, как имеющиеся в учебнике или конспекте, так и пропущенные в силу их простоты. Ряд вопросов дисциплины может быть вынесен на самостоятельное изучение. Такое задание требует оперативного выполнения. В конспекте лекций необходимо оставить место для освещения упомянутых вопросов. Обозначить вопросы, термины, материал, который вызывает трудности, пометить и попытаться найти ответ в рекомендуемой литературе. Если самостоятельно не удастся разобраться в материале, то необходимо сформулировать вопрос и задать преподавателю на консультации, на практическом занятии</p>
Практическое занятие	<p>Практическое занятие – вид аудиторных учебных занятий, целенаправленная форма организации учебного процесса, при реализации которой обучающиеся по заданию и под руководством преподавателя выполняют практические задания. Практические задания направлены на углубление научно-теоретических знаний и овладение определенными методами работы, в процессе которых вырабатываются умения и навыки выполнения тех или иных учебных действий в работе с большими данными, позволяют проверить знания обучающихся, выступают как средства оперативной обратной связи; цель практических занятий – углублять, расширять, детализировать знания, полученные на лекции, в обобщенной форме и содействовать выработке навыков профессиональной деятельности.</p> <p>На практических занятиях подробно рассматриваются основные вопросы дисциплины, разбираются основные типы задач. К каждому практическому занятию следует заранее самостоятельно выполнить домашнее задание и выучить лекционный материал к следующей теме. Систематическое выполнение домашних заданий обязательно и является важным фактором, способствующим успешному усвоению дисциплины.</p> <p>Особое внимание следует обращать на определение основных понятий дисциплины. Обучающийся должен подробно разбирать примеры, которые поясняют понятия</p>
Лабораторная работа	<p>Систематизация и обобщение знаний по большим базам данных, формирование умений работать с инструментами хранения, обработки и анализа баз данных Big Data. При подготовке к лабораторным занятиям следует использовать основную литературу из представленного списка, а также руководствоваться заданиями к лабораторным работам, размещенными в электронной информационно-образовательной среде ИрГУПС.</p>

<p>Самостоятельная работа</p>	<p>Обучение по дисциплине «Обработка и анализ больших данных» предусматривает активную самостоятельную работу обучающегося. На самостоятельную работу отводится <u>21</u> часа(ов) по очной форме обучения. В разделе 4 рабочей программы, который называется «Структура и содержание дисциплины», все часы самостоятельной работы расписаны по темам и вопросам, а также указана необходимая учебная литература: обучающийся изучает учебный материал, разбирает примеры и решает разноуровневые задачи в рамках выполнения как общих домашних заданий, так и индивидуальных домашних заданий (ИДЗ). При выполнении домашних заданий обучающемуся следует обратиться к задачам, решенным на предыдущих практических занятиях, решенным домашним работам, а также к примерам, приводимым лектором. Если этого будет недостаточно для выполнения всей работы можно дополнительно воспользоваться учебными пособиями, приведенными в разделе 6.1 «Учебная литература». Если, несмотря на изученный материал, задание выполнить не удастся, то в обязательном порядке необходимо посетить консультацию преподавателя, ведущего практические занятия, и/или консультацию лектора.</p> <p>ИДЗ должны быть выполнены обучающимся в установленные преподавателем сроки в соответствии с требованиями к оформлению КР (текстовой и графической частей), сформулированным в Положении «Требования к оформлению текстовой и графической документации. Нормоконтроль» № П.420700.05.4.092-2017.</p> <p>Обучающийся очной формы обучения выполняет:</p> <p>I семестр</p> <p>ИДЗ № 1 «Извлечение набора данных по заданной тематике.» Задания размещены в электронной информационно-образовательной среде ИрГУПС, доступной обучающемуся через его личный кабинет</p>
<p>Комплекс учебно-методических материалов по всем видам учебной деятельности, предусмотренным рабочей программой дисциплины (модуля), размещен в электронной информационно-образовательной среде ИрГУПС, доступной обучающемуся через его личный кабинет.</p>	

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ЖЕЛЕЗНОДОРОЖНОГО ТРАНСПОРТА

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Иркутский государственный университет путей сообщения»
(ФГБОУ ВО ИрГУПС)

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

**для проведения текущего контроля успеваемости
и промежуточной аттестации по дисциплине
(модулю)/практике**

Б1.О.58 Обработка и анализ больших данных

Приложение № 1 к рабочей программе

Специальность – 10.05.03 Информационная безопасность автоматизированных систем

Специализация– специализация N 5 "Безопасность открытых информационных систем"

ИРКУТСК

1. Общие положения

Фонд оценочных средств (ФОС) является составной частью нормативно-методического обеспечения системы оценки качества освоения обучающимися образовательной программы.

Фонд оценочных средств предназначен для использования обучающимися, преподавателями, администрацией Университета, а так же сторонними образовательными организациями для оценивания качества освоения образовательной программы и уровня сформированности компетенций у обучающихся.

Задачами ФОС являются:

– оценка достижений обучающихся в процессе *изучения дисциплины (модуля) или прохождения практики*;

– обеспечение соответствия результатов обучения задачам будущей профессиональной деятельности через совершенствование традиционных и внедрение инновационных методов обучения в образовательный процесс;

– самоподготовка и самоконтроль обучающихся в процессе обучения.

Фонд оценочных средств сформирован на основе ключевых принципов оценивания: валидность, надежность, объективность, эффективность.

Для оценки уровня сформированности компетенций используется трехуровневая система:

– минимальный уровень освоения, обязательный для всех обучающихся по завершению освоения образовательной программы; дает общее представление о виде деятельности, основных закономерностях функционирования объектов профессиональной деятельности, методов и алгоритмов решения практических задач;

– базовый уровень освоения, превышение минимальных характеристик сформированности компетенций; позволяет решать типовые задачи, принимать профессиональные и управленческие решения по известным алгоритмам, правилам и методикам;

– высокий уровень освоения, максимально возможная выраженность характеристик компетенций; предполагает готовность решать практические задачи повышенной сложности, нетиповые задачи, принимать профессиональные и управленческие решения в условиях неполной определенности, при недостаточном документальном, нормативном и методическом обеспечении.

2. Перечень компетенций, в формировании которых участвует дисциплина.

Программа контрольно-оценочных мероприятий.

Показатели оценивания компетенций, критерии оценки

Дисциплина «Обработка и анализ больших данных» участвует в формировании компетенций:

ОПК-2 Способен применять программные средства системного и прикладного назначений, в том числе отечественного производства, для решения задач профессиональной деятельности;

ОПК-3 Способен использовать математические методы, необходимые для решения задач профессиональной деятельности ПК-N.

Программа контрольно-оценочных мероприятий

очная форма обучения

№	Неделя	Наименование контрольно-	Объект контроля (понятие/тем/раздел и т.д.	Код индикатора достижения	Наименование оценочного средства
---	--------	--------------------------	--	---------------------------	----------------------------------

		оценочного мероприятия	дисциплины)	компетенции	(форма проведения*)
А семестр					
1.	3	Защита лабораторной работы	Начало работы. Консолидация данных. Понятие сценария и узла обработки.	ОПК-2.1; ОПК-2.2; ОПК-3.1; ОПК-3.2; ОПК-3.3	Письменно, компьютерные технологии
2.	5	Практическая работа	Выбор инструментария для работы с BigData	ОПК-2.1; ОПК-2.2; ОПК-3.1; ОПК-3.2; ОПК-3.3	Устно
3.	6	Защита лабораторной работы	Многомерное представление данных	ОПК-2.1; ОПК-2.2; ОПК-3.1; ОПК-3.2; ОПК-3.3	Письменно, компьютерные технологии
4.	7	Защита лабораторной работы	Работа с базой данных NoSQL	ОПК-2.1; ОПК-2.2; ОПК-3.1; ОПК-3.2; ОПК-3.3	Письменно, компьютерные технологии
5.	8	Практическая работа	Семантический анализ текста онлайн, seo анализ текста	ОПК-2.1; ОПК-2.2; ОПК-3.1; ОПК-3.2; ОПК-3.3	Устно
6.	9	Защита лабораторной работы	Извлечение набора данных по заданной тематике	ОПК-2.1; ОПК-2.2; ОПК-3.1; ОПК-3.2; ОПК-3.3	Письменно, компьютерные технологии
7.	11	Практическая работа	Разработка запросов на выборку информации на основе mapreduce.	ОПК-2.1; ОПК-2.2; ОПК-3.1; ОПК-3.2; ОПК-3.3	Устно
8.	11	Защита лабораторной работы	Языки поисковых запросов для Hadoop.	ОПК-2.1; ОПК-2.2; ОПК-3.1; ОПК-3.2; ОПК-3.3	Письменно, компьютерные технологии
9.	12	Защита лабораторной работы	Ассоциативные правила	ОПК-2.1; ОПК-2.2; ОПК-3.1; ОПК-3.2; ОПК-3.3	Письменно, компьютерные технологии
10.	13	Практическая работа	Изучение и конфигурирование программного комплекса Apache Hadoop. Размещения набора данных по заданной тематике	ОПК-2.1; ОПК-2.2; ОПК-3.1; ОПК-3.2; ОПК-3.3	Устно
11.	13	Защита лабораторной работы	Ассоциативные правила	ОПК-2.1; ОПК-2.2; ОПК-3.1; ОПК-3.2; ОПК-3.3	Письменно, компьютерные технологии
12.	15	Защита лабораторной работы	Кластеризация	ОПК-2.1; ОПК-2.2; ОПК-3.1;	Письменно, компьютерные технологии

				ОПК-3.2; ОПК-3.3	
13.	17	Защита лабораторной работы	Классификация. Прогноз.	ОПК-2.1; ОПК-2.2; ОПК-3.1; ОПК-3.2; ОПК-3.3	Письменно, компьютерные технологии
14.	17	Практическая работа	Общая схема анализа. Этапы моделирования. Процесс построения моделей.	ОПК-2.1; ОПК-2.2; ОПК-3.1; ОПК-3.2; ОПК-3.3	Устно

*Форма проведения контрольно-оценочного мероприятия: устно, письменно, компьютерные технологии.

Описание показателей и критериев оценивания компетенций.

Описание шкал оценивания

Контроль качества освоения дисциплины/прохождения практики включает в себя текущий контроль успеваемости и промежуточную аттестацию. Текущий контроль успеваемости и промежуточная аттестация обучающихся проводятся в целях установления соответствия достижений обучающихся поэтапным требованиям образовательной программы к результатам обучения и формирования компетенций.

Текущий контроль успеваемости – основной вид систематической проверки знаний, умений, навыков обучающихся. Задача текущего контроля – оперативное и регулярное управление учебной деятельностью обучающихся на основе обратной связи и корректировки. Результаты оценивания учитываются в виде средней оценки при проведении промежуточной аттестации.

Для оценивания результатов обучения используется четырехбалльная шкала: «отлично», «хорошо», «удовлетворительно», «неудовлетворительно» и/или двухбалльная шкала: «зачтено», «не зачтено».

Перечень оценочных средств, используемых для оценивания компетенций, а так же краткая характеристика этих средств приведены в таблице

№	Наименование оценочного средства	Краткая характеристика оценочного средства	Представление оценочного средства в ФОС
1.	Собеседование	Средство контроля на практическом занятии, организованное как специальная беседа преподавателя с обучающимся на темы, связанные с изучаемой дисциплиной, и рассчитанное на выяснение объема знаний обучающегося по определенному разделу, теме, проблеме и т.п. Может быть использовано для оценки знаний обучающихся	Вопросы по темам/разделам дисциплины
2.	Защита лабораторной работы	Средство, позволяющее оценить умение обучающегося письменно излагать суть поставленной задачи, самостоятельно применять стандартные методы решения поставленной задачи с использованием имеющейся лабораторной базы, проводить анализ полученного результата работы. Может быть использовано для оценки умений, навыков и (или) опыта деятельности обучающихся	Темы лабораторных работ и требования к их защите
3.	Тест	Система стандартизированных заданий, позволяющая автоматизировать процедуру измерения уровня знаний и умений обучающегося. Может быть использовано для оценки знаний, умений, навыков и (или) опыта деятельности обучающихся	Фонд тестовых заданий
4.	Экзамен	Средство, позволяющее оценить знания, умения, навыки и (или) опыта деятельности обучающегося по дисциплине.	Перечень теоретических

	Может быть использовано для оценки знаний, умений, навыков и (или) опыта деятельности обучающихся	вопросов и практических заданий (билетов) к экзамену
--	---	--

Критерии и шкалы оценивания компетенций в результате *изучения дисциплины/ прохождения практики при проведении промежуточной аттестации* в форме зачета и/или экзамена. Шкала оценивания уровня освоения компетенций

Шкалы оценивания	Критерии оценивания	Уровень освоения компетенции
«отлично»	Обучающийся правильно ответил на теоретические вопросы. Показал отличные знания в рамках учебного материала. Правильно выполнил практические задания. Показал отличные умения и владения навыками применения полученных знаний и умений при решении задач в рамках учебного материала. Ответил на все дополнительные вопросы	Высокий
«хорошо»	Обучающийся с небольшими неточностями ответил на теоретические вопросы. Показал хорошие знания в рамках учебного материала. С небольшими неточностями выполнил практические задания. Показал хорошие умения и владения навыками применения полученных знаний и умений при решении задач в рамках учебного материала. Ответил на большинство дополнительных вопросов	Базовый
«удовлетворительно»	Обучающийся с существенными неточностями ответил на теоретические вопросы. Показал удовлетворительные знания в рамках учебного материала. С существенными неточностями выполнил практические задания. Показал удовлетворительные умения и владения навыками применения полученных знаний и умений при решении задач в рамках учебного материала. Допустил много неточностей при ответе на дополнительные вопросы	Минимальный
«неудовлетворительно»	Обучающийся при ответе на теоретические вопросы и при выполнении практических заданий продемонстрировал недостаточный уровень знаний и умений при решении задач в рамках учебного материала. При ответах на дополнительные вопросы было допущено множество неправильных ответов	Компетенция не сформирована

Критерии и шкалы оценивания результатов обучения при проведении текущего контроля успеваемости

Коллоквиумов, собеседования – пример 1

Шкала оценивания	Критерии оценивания
«отлично»	Глубокое и прочное усвоение программного материала. Полные, последовательные, грамотные и логически излагаемые ответы при видоизменении задания. Обучающийся свободно справляется с поставленными задачами, может обосновать принятые решения, демонстрирует владение разносторонними навыками и приемами выполнения практических работ
«хорошо»	Знание программного материала, грамотное изложение, без существенных неточностей в ответе на вопрос, правильное применение теоретических знаний, владение необходимыми навыками при выполнении практических задач
«удовлетворительно»	Обучающийся демонстрирует усвоение основного материала, при ответе допускаются

	неточности, при ответе недостаточно правильные формулировки, нарушение последовательности в изложении программного материала, затруднения в выполнении практических заданий
	Слабое знание программного материала, при ответе возникают ошибки, затруднения при выполнении практических работ
«неудовлетворительно»	Не было попытки выполнить задание

Защита лабораторной работы

Шкала оценивания	Критерии оценивания
«отлично»	Лабораторная работа выполнена в обозначенный преподавателем срок, письменный отчет без замечаний. Лабораторная работа выполнена обучающимся в полном объеме с соблюдением необходимой последовательности. Обучающийся работал полностью самостоятельно; показал необходимые для проведения работы теоретические знания, практические умения и навыки. Работа (отчет) оформлена аккуратно, в наиболее оптимальной для фиксации результатов форме
«хорошо»	Лабораторная работа выполнена в обозначенный преподавателем срок, письменный отчет с небольшими недочетами. Лабораторная работа выполнена обучающимся в полном объеме и самостоятельно. Допущены отклонения от необходимой последовательности выполнения, не влияющие на правильность конечного результата. Работа показывает знание обучающимся основного теоретического материала и овладение умениями, необходимыми для самостоятельного выполнения работы. Допущены неточности и небрежность в оформлении результатов работы (отчета)
«удовлетворительно»	Лабораторная работа выполнена с задержкой, письменный отчет с недочетами. Лабораторная работа выполняется и оформляется обучающимся при посторонней помощи. На выполнение работы затрачивается много времени. Обучающийся показывает знания теоретического материала, но испытывает затруднение при самостоятельной работе с источниками знаний или приборами
«неудовлетворительно»	Лабораторная работа не выполнена, письменный отчет не представлен. Результаты, полученные обучающимся не позволяют сделать правильных выводов и полностью расходятся с поставленной целью. Показывается плохое знание теоретического материала и отсутствие необходимых умений. Лабораторная работа не выполнена, у учащегося отсутствуют необходимые для проведения работы теоретические знания, практические умения и навыки

Оценочное средство «Тест»

Тестирование с применением компьютерных технологий проводится по окончании каждого семестра (если дисциплина не является односеместровой) и по окончании изучения дисциплины и (или) в течение года по завершению изучения дисциплины (контроль/проверка остаточных знаний, умений, навыков и (или) опыта деятельности).

Тесты формируются из фонда тестовых заданий по дисциплине. Структура фонда тестовых заданий по дисциплине, структуры тестов по итогам каждого семестра (если дисциплина не является односеместровой) и итогового теста по дисциплине и типовые примеры тестов приведены в разделе 3 данного документа.

Результаты тестирования могут быть использованы при проведении промежуточной аттестации, как в форме зачета, так и в форме экзамена.

Промежуточная аттестация в форме зачета:

Критерии оценивания	Шкала оценивания
Обучающийся набрал при тестировании более 69 баллов	«зачтено»
Обучающийся набрал при тестировании	«не зачтено»

менее 69 баллов	
-----------------	--

Промежуточная аттестация в форме экзамена – результаты тестирования могут являться допуском к экзамену:

Критерии оценивания	Шкала оценивания
Обучающийся набрал при тестировании более 69 баллов	Обучающийся к экзамену допущен
Обучающийся набрал при тестировании менее 69 баллов	Обучающийся к экзамену не допущен

Преподаватель вправе предусмотреть тесты для самоконтроля обучающихся по разделам дисциплины, сформировав их из материалов фонда тестовых заданий дисциплины. Требования к тестам для самоконтроля аналогичны требованиям к итоговым тестам по семестрам и дисциплине в целом.

3. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности

3.9 Перечень теоретических вопросов к экзамену (для оценки знаний)

1. Понятие «большие данные». Характеристики больших данных. Принципы работы с большими данными.
2. Большие данные в научных исследованиях. Экосистема больших данных.
3. Технологии хранения больших данных.
4. Распределенные файловые системы.
5. Распределенные фреймворки.
6. Хранилища, классификация и примеры.
7. Распределенные базы данных NoSQL и новые SQL базы данных.
8. Архитектура системы и алгоритмы обработки больших данных.
9. Уровни в системах обработки больших данных. Прием данных (Data Ingestion). Сбор данных (Data Staging). Анализ данных (Analysis Layer). Представление результатов (Consumption Layer).
10. Параллельные алгоритмы для работы с данными.
11. Операторы Map и Reduce. Оператор Map (предварительная обработка). Оператор Reduce (свертка).
12. Лямбда-архитектура.
13. Каппа-архитектура.
14. Системы управления потоками данных.
15. Системы хранения больших данных. Платформы больших данных.
16. Обработка данных в реальном времени. Системы управления большими данными. Аналитические платформы.
17. Состав и возможности программного комплекса Apache Hadoop.
18. Языки поисковых запросов для Hadoop. Принципы работы Apache Spark. Прочие компоненты экосистемы Hadoop.
19. Понятие сценария и узла обработки. Консолидация данных. Трансформация данных. Визуализация данных.
20. Ассоциативные правила. Поиск ассоциативных правил.
21. Кластеризация. Алгоритм кластеризации k-means.
22. Прогнозирование с помощью линейной регрессии.
23. Классификация с помощью нейросети.

24. Классификация с помощью деревьев решений.

3.10 Перечень типовых простых практических заданий к экзамену

1. Назовите входные данные алгоритма k-средних.
2. Назовите выходные данные алгоритма k-средних.
3. Приведите последовательность шагов в алгоритме k-средних.
4. Приведите порядок кластерного анализа с помощью ПО k-средних.
5. Найдите в сети Интернет два сайта, на которых используются системы прогнозирования.
6. Найдите в сети Интернет два сайта, на которых используются рекомендательные системы.
7. Пользуясь системой SCOPUS, проанализируйте динамику количества публикаций за пять лет по направлениям Deep Learning, Big Data, Recommender Systems, Social Network Analysis.
8. Пользуясь системой SCOPUS, найдите пять публикаций с наибольшей цитируемостью публикаций за последние десять лет по направлениям Deep Learning, Big Data, Recommender Systems, Social Network Analysis.
9. Пользуясь системами SCOPUS, Web of Science, E-library (РИНЦ), выявите нескольких ведущих ученых в сфере анализа данных.

3.11 Перечень типовых комплексных практических заданий к экзамену

1. Ознакомьтесь с доступными способами обработки данных. Для предложенных преподавателем данных выполните консолидацию, трансформацию, визуализацию данных.
2. Выполните построение ассоциативных правил для предложенных преподавателем данных, используя различные параметры построения ассоциативных правил. Сравните полученные результаты. Опишите 4-5 ассоциативных правил, полученных в ходе выполнения работы.
3. Используя механизм кластеризации реализованный на алгоритме k-means, основываясь на данных предложенных преподавателем, решите задачу распределения данных на кластеры и выявления скрытых закономерностей. Проанализируйте получившуюся картину распределения.
4. Постройте прогноз для предложенных преподавателем данных с помощью линейной регрессии. Проанализируйте построенную с помощью линейной регрессии модель прогноза.
5. Постройте карты Кохонена для предложенных преподавателем данных. Проанализируйте результаты. Используя различные отображения карты Кохонена, постройте 3-4 правила.
6. Постройте дерево решения для предложенных преподавателем данных. Попробуйте использовать различные значения параметров обучения дерева решения и сравните полученные деревья. Выведите 5 правил из построенного дерева решений. Приведите 4-5 примеров, для которых можно использовать метод обработки дерева решений.

3.12 Тестирование по дисциплине

3.12.1 Структура фонда тестовых заданий по дисциплине «Обработка и анализ больших данных»

Раздел дисциплины	Тема раздела	Количество тестовых заданий (ТЗ), типы ТЗ
Раздел 1. Большие данные. Основы систем	1.1. Понятие «большие данные».	21-тип А
		8-тип В

больших данных. Программные платформы и системы для больших данных		5-тип С
	1.2. Технологии хранения больших данных.	2-тип D
	1.3. Архитектура системы и алгоритмы обработки больших данных. Уровни в системах обработки больших данных.	
	1.4. Параллельные алгоритмы для работы с данными.	
	1.5. Системы управления большими данными.	
Итого по разделу		Σ 36 21– тип А 8– тип В 5 – тип С 2– тип D
Раздел №2 Технологии анализа больших данных	2.1. Жизненный цикл анализа больших данных, стандарты. Технологии KDD и Data Mining.	21-тип А
		8-тип В
		5-тип С
		2-тип D
		Σ 36 21– тип А 8– тип В 5 – тип С 2– тип D

**3.5.2 Структура и образец типового теста
за А семестр/итогового теста по дисциплине «Обработка и анализ больших данных»
за весь период ее освоения**

Раздел дисциплины	Тема раздела	Количество тестовых заданий (ТЗ), типы ТЗ	
Раздел 1. Большие данные. Основы систем больших данных. Программные платформы и системы для больших данных	1.6. Понятие «большие данные».	11-тип А	
		2-тип В	
		3-тип С	
	1.7. Технологии хранения больших данных. 1.8. Архитектура системы и алгоритмы обработки больших данных. Уровни в системах обработки больших данных. 1.9. Параллельные алгоритмы для работы с данными. 1.10. Системы управления большими данными.	2-тип D	
		Итого по разделу	
		Σ 18 11– тип А 2– тип В 3 – тип С 2– тип D	
Раздел №2 Технологии анализа больших данных	2.2. Жизненный цикл анализа больших данных, стандарты. Технологии KDD и Data Mining.	11-тип А	
		2-тип В	
		3-тип С	
		2-тип D	
		Σ 18 11– тип А 2– тип В 3 – тип С 2– тип D	

Образец типового теста за А семестр

Образец типового теста содержит задания для оценки знаний, для оценки умений, для оценки навыков и (или) опыта деятельности.

1. Кто ввел термин Большие данные?
 - А) Клиффорд Линч
 - Б) Алан Тьюринг
 - В) Бьерн Страуструп
 - Г) Дональд Кнут
2. Какие данные занимают больше мировой памяти относительно остальных?
 - А) Structured Data
 - Б) Unstructured Data
 - В) Semi-Structured Data
 - Г) Quasi-Structured Data
3. BigData – это ...
 - А) Представление фактов, понятий или инструкций в форме, приемлемой для интерпретации, или обработки.
 - Б) Комплексный набор методов обработки структурированных и неструктурированных данных колоссальных объемов.
 - В) Колоссальный объем данных, собранных человечеством.
 - Г) Класс в Java, предназначенный для хранения данных от 100 Гб
4. Какая компания создала технологию MapReduce?
 - А) Google
 - Б) Yahoo
 - В) EMC
 - Г) Oracle
5. Данные текстовых файлов с определенными паттернами для их обработки (например, XML) являются:
 - А) Структурированными
 - Б) Полуструктурированными
 - В) Квазиструктурированными
 - Г) Неструктурированными
6. Что означает термин Big Data в информационных технологиях?
 - А) Комплексный набор методов для создания файлов большого объема
 - Б) Комплексный набор методов обработки структурированных и неструктурированных данных колоссальных объемов.
 - В) Файлы с большим количеством данных.
 - Г) Представление времени, дня, месяца и года в качестве значения количества миллисекунд, прошедших с начала нашей эры.
7. Данные имеющие определенный тип, формат и структуру (например, транзакционные данные) являются:
 - А) Структурированными
 - Б) Полуструктурированными
 - В) Квазиструктурированными
 - Г) Неструктурированными
8. Что является средством анализа в BI?
 - А) Карты показателей;
 - Б) Совместная работа и управление рабочими процессами;
 - В) Информационные панели;
 - Г) BI инфраструктура.
9. Основное умение исследователя данных?
 - А) Умение находить наиболее важные элементы в хранимой информации
 - Б) Уметь прогнозировать исход работы системы
 - В) Находить скрытые логические связи в системе собранной информации
 - Г) Отличать неструктурированные данные от структурированных
10. Что означает термин Business Intelligence в информационных технологиях?
 - А) Комплексный набор методов для создания бизнес планов.
 - Б) Методы и инструменты для перевода необработанной информации в осмысленную, удобную для восприятия форму.
 - В) Файлы, содержащие информацию о бизнес плане.
 - Г) Технологии, направленные на развитие бизнеса.
11. Языком, на котором был разработан RabbitMQ, является: _____

12. Что является главным результатом процесса Business Intelligence?
- А) Возможность принятия решений для бизнеса
 - Б) Результаты интеллектуального анализа данных
 - В) Возможность использования искусственного интеллекта
 - Г) Получение структуризации данных после выполнения всех шагов процесса
13. Что из перечисленного не является средством анализа?
- А) Продвинутая визуализация
 - Б) Reporting
 - В) Predictive Modelling
 - Г) Data Mining
14. Что относится к средствам предоставления информации в Business Intelligence?
- А) Генератор нерегламентированных запросов
 - Б) Совместная работа и управление рабочими процессами
 - В) Предиктивное моделирование и Data Mining
 - Г) Карты показателей
15. Процессом создания и выбора модели для предсказания вероятности наступления некоторого события является:
- А) OLAP
 - Б) Data Mining
 - В) Predictive Modelling
 - Г) Data Science
16. Что не является целью процесса Business Intelligence?
- А) Интерпретация большого количества данных;
 - Б) Моделирование исходов различных вариантов действий;
 - В) Модификация существующего программного обеспечения;
 - Г) Отслеживание результатов решений.
17. Что из этого не является реализацией Hadoop?
- А) Google MapReduce
 - Б) Phoenix
 - В) GreenMint
 - Г) Qizmt
18. Какие из перечисленных пунктов являются достоинствами MapReduce?
- А) Оптимальная производительность
 - Б) Эффективное применение в маленьких кластерах с небольшим объемом данных
 - В) Масштабируемость
 - Г) Отказоустойчивость
19. Что такое Oozie?
- А) Распределенный координационный сервис
 - Б) Нереляционная распределенная база данных
 - В) Язык управления потоком данных и исполнительная среда для анализа больших объемов данных
 - Г) Сервис для записи и планировки заданий Hadoop
20. Сколько уровней имеет лямбда-архитектура? _____
21. Какие компоненты являются частями MapReduce?
- А) Task Tracker
 - Б) Name Node и Data Node
 - В) Job Tracker и Task Tracker
 - Г) Job Tracker, Task Tracker, Name Node и Data Node
22. Что такое Spark?
- А) Инструмент для кластерных вычислений
 - Б) Графический движок
 - В) Библиотека для работы с графами
 - Г) Технология распределенных вычислений
23. Дайте определение Map Reduce...
- А) Модель распределенных вычислений, предназначенная для параллельных вычислений над очень большими (до нескольких петабайт)

- объемами данных
- Б) Набор компонентов и интерфейсов для распределенных файловых систем и общего ввода-вывода
- В) Распределенная файловая система, работающая на больших кластерах типовых машин
- Г) Распределенный сервис для коллекционирования, сбора, и перемещения больших массивов данных
24. Что из этого является недостатком MapReduce?
- А) Фиксированный алгоритм обработки данных
- Б) Масштабируемость
- В) Отказоустойчивость
- Г) Возможность автоматического распараллеливания
25. Какое API было добавлено в Hadoop v2.0? _____
26. Какая цель у NameNode в HDFS?
- А) Хранить индекс того, какая часть данных находится в каком узле
- Б) Хранить имя файла, хранящегося в конкретном узле
- В) Хранить индекс узла, в котором хранится имя файла
- Г) Хранить имена узлов
27. Вертикальное масштабирование...
- А) Требуется изменений в прикладных программах, работающих на таких системах
- Б) Не требует никаких изменений в прикладных программах, работающих на таких системах
- В) Уменьшает производительность каждого компонента БД
- Г) Увеличивает скорость загрузки данных
28. Для достижения какого свойства в БД типа NoSQL нет JOIN операций?
- А) Intercepting
- Б) Concurrency
- В) Consistency
- Г) Capacity
29. Что, согласно теореме CAP (теореме Брюера), возможно обеспечить в любой реализации распределённых вычислений?
- А) Только согласованность данных
- Б) Только доступность данных
- В) Согласованность данных, доступность данных, устойчивость к разделению
- Г) Не более двух свойств из трёх вышеприведённых
30. Выберите верное определение понятия AP-система:
- А) Система, во всех узлах которой данные согласованы и обеспечена доступность, жертвует устойчивостью к распаду на секции
- Б) Распределённая система, в каждый момент обеспечивающая целостный результат и способная функционировать в условиях распада
- В) Распределённая система, отказывающаяся от целостности результата Г) Система, автоматически изменяющая данные алгоритма своего с целью сохранения оптимального состояния
31. Что означает термин NoSQL? _____
32. Разбиение системы на более мелкие структурные компоненты и разнесение их по отдельным физическим машинам (или их группам), и (или) увеличение количества серверов, параллельно выполняющих одну и ту же функцию, это:
- А) Горизонтальное масштабирование
- Б) Вертикальное масштабирование
- В) Master- slave репликация
- Г) Peer-to-peer репликация
33. Что из перечисленного относится к графо-ориентированным хранилищам (Graph Store)?
- А) Neo4j
- Б) BaseX
- В) Elasticsearch
- Г) Ничего
34. Что поддерживает NoSQL?

- A) Операцию Insert
 - Б) Полностью стандарт SQL
 - В) Операцию Join
 - Г) Операцию Group by
35. Какие три свойства фигурируют в определении теоремы CAP?
- A) Согласованность данных
 - Б) Сложность
 - В) Доступность
 - Г) Устойчивость к разделению
36. Выделение таблицы или группы таблиц на отдельный сервер это...
- A) Горизонтальное масштабирование
 - Б) Вертикальное масштабирование
 - В) Горизонтальный шардинг
 - Г) Вертикальный шардинг
37. Что из этого не является типом визуализации?
- A) График
 - Б) Текст
 - В) Круговая диаграмма
 - Г) Гистограмма
38. Отображение зависимости значений одной величины от другой - это..._____
39. К каким алгоритмам классификации относится метод ближайших соседей?
- A) Метрическим
 - Б) Логическим
 - В) Линейным
 - Г) Нет верного ответа
40. Преимуществом метода ближайшего соседа является:
- A) Устойчивость к погрешностям
 - Б) Наличие настраиваемых параметров
 - В) Высокое качество классификации
 - Г) Простота реализации
41. С помощью какого алгоритма можно найти ассоциативное правило?
- A) Алгоритм apriori
 - Б) Алгоритм k-means
 - В) Алгоритм c-means
 - Г) Иерархический алгоритм
42. Технология машинного обучения, когда нет ответов и требуется искать зависимости между объектами, называется ...
- A) Самостоятельное обучение
 - Б) Обучение без учителя
 - В) Обучение с учителем
 - Г) Обучение по зависимостям
43. Критерий Пирсона является:
- A) Критерием значимости
 - Б) Параметрическим критерием
 - В) Критерием согласия
 - Г) Непараметрических критерием
44. Чем отличаются ошибки первого и второго рода при принятии решений?
- A) Ошибка первого рода значительнее, нежели второго
 - Б) Ошибка второго рода не обнаруживает различия, которые есть, а первого обнаруживает, которых нет
 - В) Ошибка второго рода значительнее, нежели первого
 - Г) Ошибка первого рода не обнаруживает различия, которых нет, а второго обнаруживает (положительных классификаций)
45. Аналитик это ...
- а) специалист в области анализа и моделирования
 - б) специалист в предметной области;

- в) человек, решающий определенные задачи;
 - г) человек, который имеет опыт в программировании.
46. Эксперт это ...
- а) специалист в области анализа и моделирование;
 - б) специалист в предметной области;
 - в) человек, решать определенные задачи;
 - г) человек, который имеет опыт в программировании.
47. Задача классификации сводится к ...
- а) нахождению частых зависимостей между объектами или событиями;
 - б) определению класса объекта по его характеристикам;
 - в) определению по известным характеристикам объекта значения некоторого его параметра;
 - г) поиску независимых групп и их характеристик во всем множестве анализируемых данных.
48. Задача регрессии сводится к ...
- а) нахождению частых зависимостей между объектами или событиями;
 - б) определению класса объекта по его характеристикам;
 - в) определению по известным характеристикам объекта значения некоторого его параметра;
 - г) поиску независимых групп и их характеристик во всем множестве анализируемых данных.
49. Задача кластеризации заключается в ...
- а) нахождении частых зависимостей между объектами или событиями;
 - б) определении класса объекта по его характеристикам;
 - в) определении по известным характеристикам объекта значения некоторого его параметра;
 - г) поиске независимых групп и их характеристик во всем множестве анализируемых данных.
50. Целью поиска ассоциативных правил является ...
- а) нахождение частых зависимостей между объектами или событиями;
 - б) определение класса объекта по его характеристикам;
 - в) определение по известным характеристикам объекта значение некоторого его параметра;
 - г) поиск независимых групп и их характеристик в всем множестве анализируемых данных.
51. Модели классификации описывают ...
- а) правила или набор правил, в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
 - б) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
 - в) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
 - г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.
52. Виды лингвистической неопределенности:
- а) неточность измерений значений определенной величины, выполняемых физическими приборами;
 - б) неопределенность значений слов (Многозначность, размытость, непонятность, нечеткость); неоднозначность смысла фраз (Синтаксическая и семантическая);
 - в) случайность (или наличие в внешней среде нескольких возможностей, каждая из которых случайным образом может стать действительностью); неопределенность значений слов (многозначность, размытость, неясность, нечеткость)
 - г) неоднозначность смысла фраз (Синтаксическая и семантическая).
53. Модели ассоциации проявляют ...
- а) исключительные ситуации в записях, которые резко отличаются по произвольному признаку от основного множества записей;
 - б) ограничения на данные анализируемого массива;
 - в) закономерности между связанными событиями;
 - г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

54. Виды физической неопределенности данных:

- а) неточность измерений значений определенной величины, выполняемых физическими приборами; случайность (или наличие в внешней среде нескольких возможностей, каждая из которых случайным образом может стать действительностью)
- б) неопределенность значений слов (Многозначность, размытость, непонятность, нечеткость); неоднозначность смысла фраз (Синтаксическая и семантическая);
- в) случайность (или наличие в внешней среде нескольких возможностей, каждая из которых случайным образом может стать действительностью); неопределенность значений слов (многозначность, размытость, неясность, нечеткость);
- г) неоднозначность смысла фраз (Синтаксическая и семантическая).

55. Комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке: аномалий, пропусков, дубликатов, противоречий, шумов и шумов и т.д.

56.

Консолидация — ...	а) комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке: аномалий, пропусков, дубликатов, противоречий, шумов и т.д. б) процесс дополнения данных некоторой информацией, позволяющей повысить эффективность развязки аналитических задач в) объект, содержащий структурированные данные, которые могут оказаться полезными для развязки аналитической задачи г) комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразования в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему
Обогащение — ...	

57.

Классификация — ...	а) некоторый набор операций над базой данных, который рассматривается как единственное завершённое, с точки зрения пользователя, действие над некоторой информацией, обычно связано с обращением к базе данных б) разновидность систем хранения, ориентирована на поддержку целостности процесса анализа данных, обеспечивает, непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов в) высокоуровневые средства отражения информационной модели и описания структуры данных
Транзакция — ...	

	г) это установление зависимости дискретной выходной переменной от входных переменных
--	--

58. Аналитическая платформа — ...

- а) специализированное программное решение (или набор решений), который включает в себя все инструменты для извлечения закономерностей из сырых данных
- б) это группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов
- в) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, и соответствующий ему правильный выходной результат

59. Обучающая выборка — ...

- а) это группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов
- б) набор данных, каждая запись которого представляет собой учебный пример, содержащий заданное входное влияние, и соответствующий ему правильный выходной результат
- в) выявление в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности

60.

Ошибка обучения — ...	а) это ошибка, допущенная моделью на учебном множестве.
Ошибка обобщения — ...	б) это ошибка, полученная на тестовых примерах, то есть, вычисляется по тем же формулам, но для тестового множества
	в) имена, типы, метки и назначения полей исходной выборки данных
	г) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданное входное влияние, и соответствующий ему правильный выходной результат

4. Методические материалы, определяющие процедуру оценивания знаний, умений, навыков и (или) опыта деятельности

В таблице приведены описания процедур проведения контрольно-оценочных мероприятий и процедур оценивания результатов обучения с помощью оценочных средств в соответствии с рабочей программой дисциплины/практики.

Наименование оценочного средства	Описания процедуры проведения контрольно-оценочного мероприятия и процедуры оценивания результатов обучения
Собеседование	Описания процедуры проведения контрольно-оценочного мероприятия и процедуры оценивания результатов обучения
Защита лабораторной работы	Описания процедуры проведения контрольно-оценочного мероприятия и процедуры оценивания результатов обучения
Тест	Тестирование с применением компьютерных технологий проводится по окончании

	<p>каждого семестра и по окончании изучения дисциплины и (или) в течение года по завершению изучения дисциплины (контроль/проверка остаточных знаний, умений, навыков и (или) опыта деятельности). Тесты формируются из фонда тестовых заданий по дисциплине. Структура фонда тестовых заданий по дисциплине, структуры тестов по итогам каждого семестра и итогового теста по дисциплине и типовые примеры тестов приведены в разделе 3 данного документа.</p> <p>Результаты тестирования могут быть использованы при проведении промежуточной аттестации, как в форме зачета, так и в форме экзамена.</p> <p>Описание требований, выполнение которых необходимо для успешного выполнения теста: тематика теста; перечень знать, уметь, владеть; виды и количество предъявляемых обучающемуся тестовых заданий; проходной балл; критерии оценки; норма времени; дополнительные требования, включая необходимость использования справочных таблиц и проч.</p> <p>Тесты для самоконтроля обучающихся по разделам дисциплины, сформированы их из материалов фонда тестовых заданий дисциплины. Требования к тестам для самоконтроля аналогичны требованиям к итоговым тестам по семестрам и дисциплине в целом</p>
--	--

Для организации и проведения промежуточной аттестации (в форме зачета/экзамена) составляются типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы:

- перечень теоретических вопросов к зачету/экзамену для оценки знаний;
- перечень типовых простых практических заданий к зачету/экзамену для оценки умений;
- перечень типовых практических заданий к зачету/экзамену для оценки навыков и (или) опыта деятельности.

Перечень теоретических вопросов и перечни типовых практических заданий разного уровня сложности к зачету/экзамену обучающиеся получают в начале семестра через электронную информационно-образовательную среду ИрГУПС (личный кабинет обучающегося).

Описание процедур проведения промежуточной аттестации в форме экзамена и оценивания результатов обучения

Промежуточная аттестация в форме экзамена проводится путем устного собеседования по билетам. Билеты составлены таким образом, что в каждый из них включал в себя теоретические вопросы и практические задания.


Билет содержит: два теоретических вопроса для оценки знаний. Теоретические вопросы выбираются из перечня вопросов к экзамену; три практических задания: два из них для оценки умений (выбираются из перечня типовых простых практических заданий к экзамену); третье практическое задание для оценки навыков и (или) опыта деятельности (выбираются из перечня типовых практических заданий к экзамену).

Распределение теоретических вопросов и одно практическое задание по экзаменационным билетам находится в закрытом для обучающихся доступе. Разработанный комплект билетов (25-30 билетов) не выставляется в электронную информационно-образовательную среду ИрГУПС, а хранится на кафедре-разработчике ФОС на бумажном носителе в составе ФОС по дисциплине.

На экзамене обучающийся берет билет, для подготовки ответа на экзаменационный билет обучающемуся отводится время в пределах 45 минут. В процессе ответа обучающегося на вопросы и задания билета, преподаватель может задавать дополнительные вопросы.

Каждый вопрос/задание билета оценивается по четырехбалльной системе, а далее вычисляется среднее арифметическое оценок, полученных за каждый вопрос/задание. Среднее арифметическое оценок округляется до целого по правилам округления.

Образец экзаменационного билета

 <p>ИРГУПС 2020-2021 учебный год</p>	<p>Экзаменационный билет № 1 по дисциплине «Обработка и анализ больших данных» Специализация/профиль <u> N 5 </u> <u> A </u> семестр</p>	<p>Утверждаю: Заведующий кафедрой «ИСиЗИ» ИРГУПС _____</p>
<p>1. Распределенные файловые системы.</p> <p>2. Параллельные алгоритмы для работы с данными.</p> <p>3. Кластерный анализ. Наблюдается n объектов, каждый из которых характеризуется двумя числовыми признаками: $\{x_i\}$, $\{y_i\}$, $i=1, \dots, n$, а также номером класса. $\{c_i\}$, $i=1, \dots, n$. Требуется исследовать работу алгоритмов кластеризации объектов наблюдения по двум признакам.</p>		