

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ЖЕЛЕЗНОДОРОЖНОГО ТРАНСПОРТА

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

УТВЕРЖДЕНА
приказом и.о. ректора
от «17» июня 2022 г. № 77

Б1.О.58 Обработка и анализ больших данных

рабочая программа дисциплины

Специальность/направление подготовки – 10.05.03 Информационная безопасность автоматизированных систем

Специализация/профиль – Безопасность открытых информационных систем

Квалификация выпускника – Специалист по защите информации

Форма и срок обучения – очная форма 5 лет, 6 месяцев

Кафедра-разработчик программы – Информационные системы и защита информации

Общая трудоемкость в з.е. – 3

Часов по учебному плану (УП) – 108

Формы промежуточной аттестации

очная форма обучения:

экзамен 10 семестр

Очная форма обучения

Распределение часов дисциплины по семестрам

Семестр	10	Итого
Вид занятий	Часов по УП	Часов по УП
Аудиторная контактная работа по видам учебных занятий/ в т.ч. в форме ПП*	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 от 26.11.2020 № 1457.

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 «17» июня 2022 г. № 12

Зав. кафедрой, к.э.н., доцент

Т.К. Кириллова

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 теоретических основ 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	
Профессионально-трудовое воспитание обучающихся	
Цель – формирование у обучающихся осознанной профессиональной ориентации, понимания общественного смысла труда и значимости его для себя лично, ответственного, сознательного и творческого отношения к будущей деятельности, профессиональной этики, способности предвидеть изменения, которые могут возникнуть в профессиональной деятельности, и умению работать в изменённых, вновь созданных условиях труда.	
Цель профессионально-трудового воспитания достигается по мере решения в единстве следующих задач:	
– формирование сознательного отношения к выбранной профессии;	
– воспитание чести, гордости, любви к профессии, сознательного отношения к профессиональному долгу, понимаемому как личная ответственность и обязанность;	
– формирование психологии профессионала;	
– формирование профессиональной культуры, этики профессионального общения;	
– формирование социальной компетентности и другие задачи, связанные с имиджем профессии и авторитетом транспортной отрасли	

2 МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП	
Блок/часть ОПОП	Блок 1. Дисциплины / Обязательная часть
2.1 Дисциплины и практики, на которых основывается изучение данной дисциплины	
1	Б1.О.07 Математический анализ
2	Б1.О.08 Алгебра и геометрия
3	Б1.О.09 Дискретная математика
4	Б1.О.10 Математическая логика и теория алгоритмов
5	Б1.О.11 Теория вероятностей и математическая статистика
6	Б1.О.12 Численные методы и теория оптимизации
7	Б1.О.25 Теория информации
8	Б1.О.26 Языки программирования
9	Б1.О.27 Основы кибернетики
10	Б1.О.28 Технологии и методы программирования
11	Б1.О.39 Программно-аппаратные средства защиты информации
12	Б1.О.53 Теория и практика защиты информации в автоматизированных системах железнодорожного транспорта
2.2 Дисциплины и практики, для которых изучение данной дисциплины необходимо как предшествующее	
1	БЗ.01(Д) Подготовка к процедуре защиты выпускной квалификационной работы
2	БЗ.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	ОПК-2.1 Применяет современные программные средства системного и прикладного назначений, в том числе отечественного производства, при решении	Знать: стандарты обработки и анализа больших данных, и требования, связанные с созданием и использованием SQL и NoSQL систем хранения и обработки данных
		Уметь: использовать современные инструментальные и вычислительные средства, осуществлять постановку задач анализа данных, визуализацию интерпретацию результатов

числе отечественного производства, для решения задач профессиональной деятельности	задач профессиональной деятельности	Владеть: технологиями хранения и обработки больших данных; современными программными средствами анализа больших объемов информации
	ОПК-2.2 Знает основы программных средств системного и прикладного значения, в том числе отечественного производства	Знать: понятие «большие данные»; характеристики больших данных; принципы работы с большими данными Уметь: применять современные программные платформы и системы для больших данных; выбирать инструментарий и технологии для работы с BigData Владеть: основами технологий хранения и обработки больших данных и современными программными средствами 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 автоматов и формальных языков	Знать: основы математического анализа, алгебры, теории вероятностей и математической статистики, дискретной математики, математической логики и теории алгоритмов, теории автоматов и формальных языков, необходимые при постановке задач анализа больших данных
		Уметь: Уметь: применять математические модели эффективных алгоритмов поиска и размещения данных в децентрализованной распределенной файловой системе
		Владеть: навыками применения основ математического анализа, алгебры, теории вероятностей и математической логики и теории алгоритмов, теории автоматов и формальных языков при постановке задач
	ОПК-3.2 Умеет использовать типовые математические методы и модели для решения задач профессиональной деятельности	Знать: типовые математические методы и модели для решения задач профессиональной деятельности; архитектуру системы и алгоритмы обработки больших данных; уровни в системах обработки больших данных
		Уметь: 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	Владеть: необходимыми элементами анализа и проектирования моделей данных для реляционных и не реляционных систем управления базами данных (СУБД)
Знать: методы решения задач обработки и анализа больших данных, возможности высокопроизводительных вычислительных систем, технологии распределенных вычислений, методы и модели Data Mining Уметь: 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 Владеть: способностью собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов; методами статистического и математического анализа больших данных		

4 СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Код	Наименование разделов, тем и видов работ	Очная форма				*Код индикатора достижения компетенции
		Семестр	Часы			
			Лек	Пр	Лаб	
1.0	Раздел 1. Большие данные. Основы систем больших данных. Программные платформы и системы для больших данных.					
1.1	Тема 1: Понятие «большие данные». Характеристики больших данных. Принципы работы с большими данными. Большие данные в научных исследованиях. Экосистема больших данных	10	2	4		ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3
1.2	Начало работы. Знакомство с NoSQL СУБД MongoDB	10			2	ОПК-2.1

4 СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Код	Наименование разделов, тем и видов работ	Очная форма				*Код индикатора достижения компетенции
		Семестр	Часы			
			Лек	Пр	Лаб	
						ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3
1.3	Тема 2: Технологии хранения больших данных. Распределенные файловые системы. Распределенные фреймворки. Хранилища, классификация и примеры	10	2			ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3
1.4	Форматы обмена данными в MongoDB. Моделирование данных	10			3	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3
1.5	Тема 3: 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. Семантический анализ текста онлайн, seo анализ текста	10	3	3		ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3
1.6	Работа с базой данных NoSQL Документы и индексы в MongoDB	10			2	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3
1.7	Тема 4: Параллельные алгоритмы для работы с данными. Операторы Map и Reduce. Обработка данных в реальном времени. Лямбда-архитектура. Каппа-архитектура. Разработка запросов на выборку информации на основе MapReduce	10	3	3		ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3
1.8	Запросы и запросы с условием в MongoDB	10			2	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3
1.9	Тема 5: Системы управления большими данными. Аналитические платформы. Состав и возможности программного комплекса Apache Hadoop. Изучение и конфигурирование программного комплекса Apache Hadoop. Размещения набора данных по заданной тематике	10	3	3		ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3
1.10	Распределенные Вычисления. MapReduce в MongoDB	10			2	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3
1.11	Проработка лекционного материала и подготовка к практическим занятиям раздела I	10				13 ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3
2.0	Раздел 2. Технологии анализа больших данных.					
2.1	Тема 6: Жизненный цикл анализа больших данных, стандарты. Технологии KDD и Data MiningОбщая схема анализа. Этапы моделирования. Процесс построения моделей	10	4	4		ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3
2.2	Ассоциативные правила	10			1	ОПК-2.1 ОПК-2.2 ОПК-3.1

4 СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ							
Код	Наименование разделов, тем и видов работ	Очная форма				*Код индикатора достижения компетенции	
		Семестр	Часы				
			Лек	Пр	Лаб		СР
						ОПК-3.2 ОПК-3.3	
2.3	Кластеризация	10			2	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	
2.4	Классификация. Прогноз	10			3	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	
2.5	Проработка лекционного материала и подготовка к практическим занятиям раздела 2	10				8 ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	
	Форма промежуточной аттестации – экзамен	10			36	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	
	Итого часов (без учёта часов на промежуточную аттестацию)		17	17	17	21	

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

	Библиографическое описание	Кол-во экз. в библиотеке/ онлайн
6.1.1.1	Дадян, Э. Г. Методы, модели, средства хранения и обработки данных : учебник / Э. Г. Дадян, Ю. А. Зеленков. Москва : Вузовский учебник : ИНФРА-М, 2022. - 168с.	Онлайн

6.1.2 Дополнительная литература

	Библиографическое описание	Кол-во экз. в библиотеке/ онлайн
6.1.2.1	Бродовская, Е. В. Большие данные в исследовании политических процессов : учебное пособие / Е. В. Бродовская, А. Ю. Домбровская. Москва : Московский педагогический государственный университет (МПГУ), 2018. - 88с. - Текст: электронный. - URL: https://biblioclub.ru/index.php?page=book&id=563578 (дата обращения: 14.09.2022)	Онлайн
6.1.2.2	Северо-Кавказский, Федеральные Распределенные базы данных : учебное пособие / авт.-сост. Н. Ю. Братченко. Ставрополь : Северо-Кавказский Федеральный университет (СКФУ), 2015. - 130с. - Текст: электронный. - URL: https://biblioclub.ru/index.php?page=book&id=457594 (дата обращения: 14.09.2022)	Онлайн

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

	Библиографическое описание	Кол-во экз. в библиотеке/

		онлайн
6.1.3.1	Молчанова Е.И. Методические указания по изучению дисциплины Б1.О.58 Обработка и анализ больших данных по специальности – 10.05.03 Информационная безопасность автоматизированных систем, специализация – N 5 Безопасность открытых информационных систем / Е.И. Молчанова ; ИрГУПС. – Иркутск : ИрГУПС, 2023. – 13 с. - Текст: электронный. - URL: https://www.irgups.ru/eis/for_site/umkd_files/mu_2136_1529_2022_1_signed.pdf	Онлайн
6.2 Ресурсы информационно-телекоммуникационной сети «Интернет»		
6.2.1	Научная электронная библиотека «КиберЛенинка» — https://cyberleninka.ru/	
6.2.2	Электронно-библиотечная система «Издательство Лань», https://e.lanbook.com/	
6.2.3	Научная электронная библиотека eLIBRARY.RU — https://elibrary.ru/	
6.2.4	Электронно-библиотечная система «Университетская библиотека онлайн», https://biblioclub.ru/	
6.2.5	Электронно-библиотечная система «Образовательная платформа ЮРАЙТ», https://urait.ru/	
6.3 Программное обеспечение и информационные справочные системы		
6.3.1 Базовое программное обеспечение		
6.3.1.1	Microsoft Windows Professional 10, государственный контракт от 20.07.2021 № 0334100010021000013-01	
6.3.1.2	Microsoft Office Russian 2010, государственный контракт от 20.07.2021 № 0334100010021000013-01	
6.3.1.3	FoxitReader, свободно распространяемое программное обеспечение http://free-software.com.ua/pdf-viewer/foxit-reader/	
6.3.1.4	Adobe Acrobat Reader DC свободно распространяемое программное обеспечение https://get.adobe.com/ru/reader/enterprise/	
6.3.1.5	Яндекс. Браузер. Прикладное программное обеспечение общего назначения, Офисные приложения, лицензия – свободно распространяемое программное обеспечение по лицензии BSD License	
6.3.2 Специализированное программное обеспечение		
6.3.2.1	MathCAD_student 15.0 Academic_License, Customer Number 434692, контракт от 03.12.2012 № 0334100010012000148-0000756-01	
6.3.2.2	Python 3.9, свободно распространяемое программное обеспечение https://docs.python.org/3/license.html	
6.3.2.3	Dev-C++, свободная интегрированная среда разработки приложений для языков программирования C/C++, https://code-live.ru/post/dev-cpp-free-cpp-ide-for-windows/	
6.3.2.4	MatLab Classroom, R2015a, R2015b, контракт от 09.07.2014 № 0334100010014000028-0000756-01.	
6.3.2.5	MatLab Classroom, R2010a, R2010b, лицензия от 16.03.2011 № 689810, ГК № 0334100010011000032-00000756-01.	
6.3.3 Информационные справочные системы		
6.3.3.1	Не предусмотрены	
6.4 Правовые и нормативные документы		
6.4.1	Не предусмотрены	

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 ИрГУПС находится – по адресу г. Иркутск, ул. Лермонтова, д.80	
2	Учебная аудитория Д-216 для проведения лекционных и практических занятий, лабораторных работ, групповых и индивидуальных консультаций, курсового проектирования (выполнения курсовых работ), текущего контроля и промежуточной аттестации. Основное оборудование: специализированная мебель, мультимедиапроектор, экран, (ноутбук переносной)	
3	Учебная аудитория Д-518 для проведения лекционных и практических занятий, лабораторных работ, групповых и индивидуальных консультаций, курсового проектирования (выполнения курсовых работ), текущего контроля и промежуточной аттестации. Основное оборудование: специализированная мебель, мультимедиапроектор, экран, (ноутбук переносной)	
4	Учебная аудитория Д-521 для проведения лекционных и практических занятий, лабораторных работ, групповых и индивидуальных консультаций, курсового проектирования (выполнения курсовых работ), текущего контроля и промежуточной аттестации. Основное оборудование: специализированная мебель, мультимедиапроектор, экран, (ноутбук переносной)	
5	Компьютерный класс Д-513 для проведения лекционных и практических занятий, лабораторных работ, групповых и индивидуальных консультаций, курсового проектирования (выполнения курсовых работ), текущего контроля и промежуточной аттестации. Основное оборудование: специализированная мебель, мультимедиапроектор (переносной), экран (переносной), компьютер	

6	Компьютерный класс Д-509 для проведения лекционных и практических занятий, лабораторных работ, групповых и индивидуальных консультаций, курсового проектирования (выполнения курсовых работ), текущего контроля и промежуточной аттестации. Основное оборудование: специализированная мебель, мультимедиапроектор (переносной), экран (переносной), компьютер
7	Компьютерный класс Д-516 для проведения лекционных и практических занятий, лабораторных работ, групповых и индивидуальных консультаций, курсового проектирования (выполнения курсовых работ), текущего контроля и промежуточной аттестации. Основное оборудование: специализированная мебель, мультимедиапроектор (переносной), экран (переносной), компьютер
8	Компьютерный класс «Информатика». «Информационные технологии» Д-505 для проведения лекционных и практических занятий, лабораторных работ, групповых и индивидуальных консультаций, курсового проектирования (выполнения курсовых работ), текущего контроля и промежуточной аттестации. Основное оборудование: специализированная мебель, мультимедиапроектор (переносной), экран (переносной), компьютер
9	Класс "Деловых игр" А-401 для проведения лекционных и практических занятий, лабораторных работ, групповых и индивидуальных консультаций, курсового проектирования (выполнения курсовых работ), текущего контроля и промежуточной аттестации. Основное оборудование: специализированная мебель, мультимедиапроектор, экран, компьютер
10	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; – помещения для хранения и профилактического обслуживания учебного оборудования – А-521

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. 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 лекции. К каждому занятию 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. 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 удается 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 данной сфере науки. Практические занятия развивают научное мышление и речь, 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</p>

	<p>следующей теме. Систематическое выполнение домашних заданий обязательно и является важным фактором, способствующим успешному усвоению дисциплины</p>
<p>Лабораторная работа</p>	<p>Основной целью лабораторных работ является теоретическое обоснование, наглядное и/или экспериментальное подтверждение и/или проверка существенных теоретических положений (законов, закономерностей) анализ существующих методик и методов их реализации и т.д. Они занимают преимущественное место при изучении дисциплин обязательной части и части, формируемой участниками образовательных отношений Блока 1.</p> <p>Исходя из цели, содержанием лабораторных работ могут быть:</p> <ul style="list-style-type: none"> - экспериментальная проверка формул, методик расчета; - проведение натурных измерений свойств, рабочих параметров, режимов работы при помощи лабораторного оборудования и/или стендов и макетов; - ознакомление, анализ и теоретические выкладки по устройству, принципу действия и способам обслуживания аппаратов, деталей машин, механизмов, процессов, протекающих в них при этом и т.д.; - наглядная графическая интерпретация чертежей, схем, объемных поверхностей и т.д., воспроизводимых с помощью специализированного программного обеспечения; - имитационное моделирование процессов, протекающих в сложных химических, физических, механических, электрических и пр. объектах; - наглядное представление о работе персонала конкретной организации или подразделения ОАО «РЖД» посредством моделирования штатных и внештатных ситуаций в виртуальных специализированных АРМ (автоматизированных рабочих мест); - установление и подтверждение закономерностей (путем сравнения проведенного эксперимента и рассчитанных значений) и т.д.; - ознакомление с методиками проведения экспериментов, наглядным устройством стенд-макетов и пр.; - установление свойств веществ, их качественных и количественных характеристик; - анализ различных характеристик процессов, в том числе производственных и иных процессов; - расчет параметров различных явлений и процессов, смоделировать которые не возможно в реальных условиях (например, чрезвычайные ситуации и пр.); - наблюдение развития явлений, процессов и др. <p>Допускается иное содержание лабораторных работ, если это будет способствовать реализации целей и задач дисциплины и формированию соответствующих компетенций.</p> <p>По характеру выполняемых лабораторных работ возможны:</p> <ul style="list-style-type: none"> - ознакомительные работы, используемые для закрепления изученного теоретического материалы; - аналитические работы, используемые для получения новой информации на основе формализованных методов; - творческие работы, ориентированные на самостоятельный выбор подходов решения задач. <p>Прежде, чем приступить к лабораторным занятиям, обучающимся необходимо повторить теоретический материал по теме работы. Каждая лабораторная работа оснащена методическими указаниями, разработанными преподавателями, ведущими дисциплину</p>
<p>Самостоятельная работа</p>	<p>Обучение по дисциплине «Обработка и анализ больших данных» предусматривает активную самостоятельную работу обучающегося. 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) и других видов работ, предусмотренных рабочей программой дисциплины. 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>Домашние задания, индивидуальные домашние задания и другие работы, предусмотренные рабочей программой дисциплины должны быть выполнены обучающимся в установленные преподавателем сроки в соответствии с требованиями к оформлению текстовой и графической документации, сформулированным в Положении «Требования к оформлению текстовой и графической документации. Нормоконтроль»</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

Приложение № 1 к рабочей программе

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

**для проведения текущего контроля успеваемости
и промежуточной аттестации**

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 ИрГУПС, а также 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;
- 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

Программа контрольно-оценочных мероприятий очная форма обучения

№	Наименование контрольно-оценочного мероприятия	Объект контроля	Код индикатора достижения компетенции	Наименование оценочного средства (форма проведения*)
10 семестр				
1.0	Раздел 1. Большие данные. Основы систем больших данных. Программные платформы и системы для больших данных			
1.1	Текущий контроль	Тема 1: Понятие «большие данные». Характеристики больших данных. Принципы работы с большими данными. Большие данные в научных исследованиях. Экосистема больших данных	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Собеседование (устно)
1.2	Текущий контроль	Начало работы. Знакомство с NoSQL СУБД MongoDB	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Лабораторная работа (письменно/устно)
1.3	Текущий контроль	Тема 2: Технологии хранения больших данных. Распределенные файловые системы. Распределенные фреймворки. Хранилища, классификация и примеры	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Собеседование (устно)
1.4	Текущий контроль	Форматы обмена данными в MongoDB. Моделирование данных	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Лабораторная работа (письменно/устно)
1.5	Текущий контроль	Тема 3: 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. Семантический анализ текста онлайн, seo анализ текста	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Собеседование (устно)
1.6	Текущий контроль	Работа с базой данных NoSQL. Документы и индексы в MongoDB	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2	Лабораторная работа (письменно/устно)

			ОПК-3.3	
1.7	Текущий контроль	Тема 4: Параллельные алгоритмы для работы с данными. Операторы Map и Reduce. Обработка данных в реальном времени. Лямбда-архитектура. Каппа-архитектура. Разработка запросов на выборку информации на основе MapReduce	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Собеседование (устно)
1.8	Текущий контроль	Запросы и запросы с условием в MongoDB	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Лабораторная работа (письменно/устно)
1.9	Текущий контроль	Тема 5: Системы управления большими данными. Аналитические платформы. Состав и возможности программного комплекса Apache Hadoop. 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	Собеседование (устно)
1.10	Текущий контроль	Распределенные Вычисления. MapReduce в MongoDB	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Лабораторная работа (письменно/устно)
1.11	Текущий контроль	Проработка лекционного материала и подготовка к практическим занятиям раздела 1	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Собеседование (устно)
2.0	Раздел 2. Технологии анализа больших данных			
2.1	Текущий контроль	Тема 6: Жизненный цикл анализа больших данных, стандарты. Технологии KDD и Data Mining Общая схема анализа. Этапы моделирования. Процесс построения моделей	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Собеседование (устно)
2.2	Текущий контроль	Ассоциативные правила	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Лабораторная работа (письменно/устно)
2.3	Текущий контроль	Кластеризация	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Лабораторная работа (письменно/устно)
2.4	Текущий контроль	Классификация. Прогноз	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Лабораторная работа (письменно/устно)
2.5	Текущий контроль	Проработка лекционного материала и подготовка к практическим занятиям раздела 2	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Собеседование (устно)

	Промежуточная аттестация	Раздел 1. Большие данные. Основы систем больших данных. Программные платформы и системы для больших данных. Раздел 2. Технологии анализа больших данных	ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Экзамен (собеседование) Экзамен - тестирование (компьютерные технологии)
--	--------------------------	---	---	---

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

Описание показателей и критериев оценивания компетенций.

Описание шкал оценивания

Контроль качества освоения дисциплины 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 двухбалльная шкала: «зачтено», «не зачтено».

Перечень оценочных средств, используемых для оценивания компетенций, а также краткая характеристика этих средств приведены в таблице.

Текущий контроль

№	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 для собеседования по темам/разделам дисциплины
2	Лабораторная работа	Средство, позволяющее оценить умение обучающегося письменно/устно 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	Образец задания для выполнения лабораторной работы и примерный перечень вопросов для ее защиты

Промежуточная аттестация

№	Наименование оценочного средства	Краткая характеристика оценочного средства	Представление оценочного средства в ФОС
1	Экзамен	Средство, позволяющее оценить знания, умения, навыков и (или) опыта деятельности обучающегося по дисциплине. Может быть использовано для оценки знаний, умений, навыков и (или) опыта деятельности обучающихся	Перечень теоретических вопросов и практических заданий (образец)

			экзаменационного билета) к экзамену
2	Тест – промежуточная аттестация в форме экзамена	Система автоматизированного контроля освоения компетенций (части компетенций) обучающимся по дисциплине (модулю) с использованием информационно-коммуникационных технологий. Может быть использовано для оценки знаний, умений, навыков и (или) опыта деятельности обучающихся	Фонд тестовых заданий

Критерии и шкалы оценивания компетенций в результате изучения дисциплины при проведении промежуточной аттестации в форме экзамена. Шкала оценивания уровня освоения компетенций

Шкала 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

Тест – промежуточная аттестация в форме экзамена

Критерии оценивания	Шкала оценивания
Обучающийся верно ответил на 90 – 100 % тестовых заданий при прохождении тестирования	«отлично»
Обучающийся верно ответил на 80 – 89 % тестовых заданий при прохождении тестирования	«хорошо»
Обучающийся верно ответил на 70 – 79 % тестовых заданий при прохождении тестирования	«удовлетворительно»
Обучающийся верно ответил на 69 % и менее тестовых заданий при прохождении тестирования	«неудовлетворительно»

Критерии и шкалы оценивания результатов обучения при проведении текущего контроля успеваемости

Собеседование

Шкалы оценивания		Критерии оценивания
«отлично»	«зачтено»	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 зачтено»	Не было попытки выполнить задание

Лабораторная работа

Шкалы оценивания		Критерии оценивания
«отлично»	«зачтено»	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»	«не зачтено»	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

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

3.1 Типовые контрольные задания для проведения собеседования

Контрольные варианты заданий выложены в электронной информационно-образовательной среде ИрГУПС, доступной обучающемуся через его личный кабинет.

Ниже приведен образец типовых вариантов заданий для проведения собеседований.

Образец типового варианта вопросов для проведения собеседования

«Тема 1: Понятие «большие данные». Характеристики больших данных. Принципы работы с большими данными. Большие данные в научных исследованиях. Экосистема больших данных»

1. Что такое данные?
2. Какие ГОСТы с определениями данных вам известны?
3. Какие определения даются в ФЗ-149?
4. Что такое жизненный цикл данных?
5. Что такое Большие данные?
6. Какие пять характеристик присущи Большим данным?
7. Какие существуют базовые принципы обработки Больших данных?

Образец типового варианта вопросов для проведения собеседования

«Тема 2: Технологии хранения больших данных. Распределенные файловые системы. Распределенные фреймворки. Хранилища, классификация и примеры»

1. Хранилища данных. Технологии OLTP, OLAP, ETL.
2. Что такое столбцовые базы данных?
3. Что такое хранилища документов?
4. Что такое потоковые данные?
5. Что такое хранилища для ключей?
6. Что такое SQL на Hadoop?
7. Что такое новый SQL?
8. Что такое графовые базы данных?

Образец типового варианта вопросов для проведения собеседования

«Тема 3: 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. Семантический анализ текста онлайн, seo анализ текста»

1. Какие уровни можно выделить в системах обработки Больших данных?
2. Зачем компьютеры объединяют в кластер?
3. В чём заключается приём данных от источников?
4. В чём может быть сложность обработки структурированных данных?
5. Приведите пример полуструктурированных данных.
6. Какие задачи может решать анализ Больших данных?
7. В каких целях используется слой/уровень компонентов системы “Прием данных”?
8. В каких целях используется слой/уровень компонентов системы “Сбор данных”?
9. В каких целях используется слой/уровень компонентов системы “Анализ данных”?
10. В каких целях используется слой/уровень компонентов системы “Представление результата”?
11. Как проверить текст на SEO?
12. Как провести SEO анализ?
13. Что такое SEO анализ текста?
14. Как проводится семантический анализ?

Образец типового варианта вопросов для проведения собеседования

«Тема 4: Параллельные алгоритмы для работы с данными. Операторы Map и Reduce. Обработка данных в реальном времени. Лямбда-архитектура. Каппа-архитектура. Разработка

запросов на выборку информации на основе MapReduce»

1. Какой тип данных может возвращать оператор свертки, примененный к начальному значению, функции и набору данных?
2. В каком случае свертка данных неэффективна?
3. Какой тип данных возвращает оператор Map, примененный к функции и набору данных?
4. Какая обработка данных обеспечивает более высокую достоверность результатов – в реальном времени или в пакетном режиме?
5. Что означает термин MapReduce?
6. Из каких шагов состоит работа MapReduce?
7. Какими преимуществами обладает MapReduce по сравнению с обычными вычислениями?
8. Опишите работу с MapReduce в MongoDB.

Образец типового варианта вопросов для проведения собеседования
«Тема 5: Системы управления большими данными. Аналитические платформы. Состав и возможности программного комплекса Apache Hadoop. Изучение и конфигурирование программного комплекса Apache Hadoop. Размещения набора данных по заданной тематике»

1. Где монтируется оборудование для обработки Больших данных?
2. Из каких компонентов состоит оборудование для обработки Больших данных?
3. Какие параметры существенны для вычислительных узлов кластера?
4. Объясните, что такое узлы хранения и вычисления?
5. Какие параметры существенны для системы хранения данных?
6. Какие параметры существенны для сетевой инфраструктуры?
7. Сколько максимально узлов может иметь кластер, использующий архитектуру подключения узлов гиперкуб, при наличии четырех быстродействующих сетевых портов в каждом узле?
8. Какое оборудование требуется для обработки Больших данных?
9. Центр обработки данных какого уровня обеспечивает максимальную надежность?
10. Центр обработки данных какого уровня обеспечивает резервирование?
11. Центр обработки данных какого уровня позволяет проводить обслуживание оборудования одновременно с обработкой данных?
12. Как много времени необходимо для создания центра обработки данных "под ключ"?
13. Какой файл управляет отчетностью в Hadoop?
14. Укажите, какой компонент хранения данных использует Hadoop?
15. Объясните, что такое Sqoop в Hadoop?
16. Перечислите три файла конфигурации Hadoop?

Образец типового варианта вопросов для проведения собеседования
«Проработка лекционного материала и подготовка к практическим занятиям раздела 1»
Осуществляется в контексте вопросов, предложенных по каждой теме раздела.

Образец типового варианта вопросов для проведения собеседования
«Тема 6: Жизненный цикл анализа больших данных, стандарты. Технологии KDD и Data Mining. Общая схема анализа. Этапы моделирования. Процесс построения моделей»

1. Понятие бизнес-аналитики. Классификация средств «бизнес-аналитики».
2. CRISP-DM
3. Этапы анализа данных. KDD.
4. Data Mining. Средства обработки Data Mining
5. Элементы математической статистики. Описательная статистика. Операции агрегирования данных.
6. Начальные этапы KDD. ETL. Средства очистки и трансформации данных.
7. Классификация методов предобработки и очистки данных.
8. Общая характеристика задач кластерного анализа.

9. Метрики кластерного анализа.
10. Ассоциативные правила. Свойства антимонотонности.
11. Метрики построения ассоциативных правил.
12. Алгоритм построения ассоциативных правил a priori.
13. Задачи классификации.
14. Оценка качества классификации.

Образец типового варианта вопросов для проведения собеседования
 «Проработка лекционного материала и подготовка к практическим занятиям раздела 2»
 Осуществляется в контексте вопросов, предложенных по каждой теме раздела.

3.2 Типовые задания для выполнения лабораторной работы и примерный перечень вопросов для ее защиты

Контрольные варианты заданий выложены в электронной информационно-образовательной среде ИрГУПС, доступной обучающемуся через его личный кабинет.

Ниже приведен образец типовых вариантов заданий для выполнения лабораторных работ и примерный перечень вопросов для их защиты.

Образец заданий для выполнения лабораторных работ и примерный перечень вопросов для их защиты

«Начало работы. Знакомство с NoSQL СУБД MongoDB»

1. Установите MongoDB одним из описанных выше способов.
2. Произведите подключение к тестовой базе данных.
3. Добавьте произвольные данные в БД с использованием командной строки.
4. Извлеките добавленные на предыдущем шаге данные с помощью командной строки.
5. Добавьте произвольные данные в базу данных с использованием Mongo Explorer.

Контрольные вопросы

1. Что означает термин NoSQL?
2. Какие преимущества предоставляют NoSQL базы данных в сравнении с реляционными базами данных?
3. Какими особенностями обладает MongoDB?
4. Сколькими способами можно произвести установку MongoDB? Кратко опишите эти способы.
5. На какие группы делятся приложения, входящие в состав MongoDB?
6. Создает ли MongoDB по умолчанию какую-либо базу данных? Если создает, назовите её имя.
7. Какие существуют способы взаимодействия с БД Mongo?
8. Существует ли возможность взаимодействовать с БД Mongo посредством WEB?
9. Какими особенностями обладает Mongo Explorer?
10. Опишите интерфейс приложения Mongo Explorer.
11. Каков формат команд командной строки?

Образец заданий для выполнения лабораторных работ и примерный перечень вопросов для их защиты

«Форматы обмена данными в MongoDB. Моделирование данных»

Индивидуальное задание

Составить схему РБД в соответствии с вашим вариантом. Создать БД по созданной вами схеме в любой СУБД. Заполнить базу данных произвольными данными. Представить созданную на предыдущем шаге БД в нереляционном виде, записанную в JSON.

Варианты индивидуального задания

1. Фирма, торгующая автомобилями.
2. Магазин, торгующий цифровыми фотоаппаратами.
3. Фирма, занимающаяся производством USB-гаджетов.

Контрольные вопросы

1. Дайте определения терминам JSON и BSON.
2. На каких структурах строится JSON?
3. Какие форма представления данных используются в JSON?
4. Возможно ли ссылаться из одних документов MongoDB на другие? Если возможно, то какие механизмы используются для этого?
5. Возможно ли в MongoDB использовать массив документов?
6. Существует ли возможность в MongoDB использовать вложенные документы?

Образец заданий для выполнения лабораторных работ и примерный перечень вопросов для их защиты

«Работа с базой данных NoSQL. Документы и индексы в MongoDB»

1. Создайте базу данных, с которой вы в дальнейшем будете работать
2. Создайте коллекцию, в которой у вас будут храниться документы.
3. Наполните коллекцию документами в соответствии с вашим индивидуальным вариантом. Создайте не менее 10 документов. Документы коллекции должны содержать, по крайней мере, один массив, и один вложенный документ.
4. Создайте составной и полнотекстовый индексы для документов в вашей коллекции.
5. Создайте запрос, производящий замену документа.
6. Создайте запрос, производящий обновление любого поля документа и запрос, добавляющий в него новые поля.
7. Создайте запрос для получения всех документов коллекции и определите, используется ли для него индекс.

Контрольные вопросы

1. Какие способы создания БД присутствуют в MongoDB?
2. Возможно ли неявное создание коллекции?
3. Каков синтаксис функции обновления документа?
4. Приведите пример запроса для обновления поля документа?
5. Какие модификаторы функции обновления вы знаете? Перечислите их назначение.
6. С помощью какой функции вы можете удалить документ?
7. Для чего используется индекс в БД mongo?
8. Как создается и удаляется индекс в MongoDB?
9. Перечислите опции создания индекса?
10. Какие типы индексов можно создать в MongoDB?

Образец заданий для выполнения лабораторных работ и примерный перечень вопросов для их защиты

«Запросы и запросы с условием в MongoDB»

1. Создайте пару простых запросов для выборки данных из БД.
2. Создайте сложные запросы с каждым из перечисленных модификаторов.
3. Создайте запросы с использованием методов сортировки, ограничения и пропуска данных.
4. Всего у вас должно получиться не менее десяти уникальных запросов.

Контрольные вопросы

1. Приведите синтаксис метода «Find».
2. Возможно, ли производить поиск по нескольким полям?
3. Каким образом можно произвести сортировку по нескольким полям? Приведите пример запроса.
4. Можно ли смешивать включения и исключения полей при составлении запроса?

Образец заданий для выполнения лабораторных работ и примерный перечень вопросов для их защиты

«Распределенные Вычисления. Mapreduce в MongoDB»

1. Создайте коллекцию документов для обработки её с помощью MapReduce.

2. Наполните коллекцию документами.
3. Произведите обработку коллекции с использованием модели распределенных вычислений MapReduce.

Контрольные вопросы

1. Что означает термин MapReduce?
2. Из каких шагов состоит работа MapReduce?
3. Какими преимуществами обладает MapReduce по сравнению с обычными вычислениями?
4. Опишите работу с MapReduce в MongoDB.

Образец заданий для выполнения лабораторных работ и примерный перечень вопросов для их защиты

«Ассоциативные правила»

Построить ассоциативные правила по имеемым транзакциям. Рассчитать характеристики для каждого правила.

Транзакционная база данных	
TID	Приобретенные покупки
100	ремень, женская сумка, портмоне
200	женская сумка, косметичка
300	женская сумка, ремень, ключница, портмоне
400	дамский зонт, ключница, косметичка
500	ремень, женская сумка, портмоне, ключница
600	косметичка, портмоне
700	ремень, портфель

Контрольные вопросы

1. Понятие ассоциативного правила и транзакции.
2. Определение поддержки и достоверности.
3. Определение значимости и полезности ассоциативных правил, показатели их характеризующие.

Образец заданий для выполнения лабораторных работ и примерный перечень вопросов для их защиты

«Кластеризация»

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.

Контрольные вопросы

1. Формальная постановка задачи кластеризации.
2. Цели кластеризации.
3. Основные шаги алгоритма k-means. Условие остановки алгоритма k-means.
4. Преимущества и недостатки алгоритма k-means.

Образец заданий для выполнения лабораторных работ и примерный перечень вопросов для их защиты

«Классификация. Прогноз»

Построить дерево решений по данным, приведенным в таблице.

Рейтинг	Возраст	Уровень Дохода	Образование
0	35	3000	0
0	25	5000	1
0	31	7000	1
1	56	1000	0
1	62	1100	1
1	49	1500	0

Контрольные вопросы

1. Определение дерева решений.

2. Структура дерева решений.
3. Выбор атрибута разбиения в узле.

3.3 Типовые контрольные задания для проведения тестирования

Фонд тестовых заданий по дисциплине содержит тестовые задания, распределенные по разделам и темам, с указанием их количества и типа.

Структура фонда тестовых заданий по дисциплине

Индикатор достижения компетенции	Тема в соответствии с РПД/РПП	Характеристика ТЗ	Количество тестовых заданий, типы ТЗ
ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Тема 1: Понятие «большие данные». Характеристики больших данных. Принципы работы с большими данными. Большие данные в научных исследованиях. Экосистема больших данных	Знание	4 – ОТЗ 2 – ЗТЗ
		Умение	2 – ОТЗ 3 – ЗТЗ
		Действие	1 – ОТЗ 1 – ЗТЗ
ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Тема 2: Технологии хранения больших данных. Распределенные файловые системы. Распределенные фреймворки. Хранилища, классификация и примеры	Знание	4 – ОТЗ 2 – ЗТЗ
		Умение	2 – ОТЗ 3 – ЗТЗ
		Действие	2 – ОТЗ 2 – ЗТЗ
ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Тема 3: 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. Семантический анализ текста онлайн, seo анализ текста	Знание	3 – ОТЗ 2 – ЗТЗ
		Умение	2 – ОТЗ 3 – ЗТЗ
		Действие	1 – ОТЗ 1 – ЗТЗ
ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Тема 4: Параллельные алгоритмы для работы с данными. Операторы Map и Reduce. Обработка данных в реальном времени. Лямбда-архитектура. Каппа-архитектура. Разработка запросов на выборку информации на основе mapreduce	Знание	3 – ОТЗ 2 – ЗТЗ
		Умение	2 – ОТЗ 3 – ЗТЗ
		Действие	2 – ОТЗ 2 – ЗТЗ
ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Тема 5: Системы управления большими данными. Аналитические платформы. Состав и возможности программного комплекса Apache Hadoop. Изучение и конфигурирование программного комплекса Apache Hadoop. Размещения набора данных по заданной тематике	Знание	3 – ОТЗ 2 – ЗТЗ
		Умение	2 – ОТЗ 4 – ЗТЗ
		Действие	1 – ОТЗ 1 – ЗТЗ
ОПК-2.1 ОПК-2.2 ОПК-3.1 ОПК-3.2 ОПК-3.3	Тема 6: Жизненный цикл анализа больших данных, стандарты. Технологии KDD и Data Mining. Общая схема анализа. Этапы моделирования. Процесс построения моделей	Знание	3 – ОТЗ 2 – ЗТЗ
		Умение	2 – ОТЗ 3 – ЗТЗ
		Действие	2 – ОТЗ 2 – ЗТЗ
		Итого	81

Полный комплект ФТЗ хранится в электронной информационно-образовательной среде ИрГУПС и обучающийся имеет возможность ознакомиться с демонстрационным вариантом ФТЗ.

Ниже приведен образец типового варианта итогового теста, предусмотренного рабочей программой дисциплины.

1. 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.

Ответ: а.

2.

Классификация — ...	а) некоторый набор операций над базой данных, который рассматривается как единственное
Транзакция — ...	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
	г) это установление зависимости дискретной выходной переменной от входных переменных

Ответ: г), а).

3. 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.

4. 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.**

5. К каким алгоритмам классификации относится метод ближайших соседей?

- а) **Метрическим**
- б) Логическим
- в) Линейным
- г) Нет верного ответа

6. 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**

7. Данные имеющие определенный тип, формат и структуру (например, транзакционные данные) являются:

- а) **Структурированными.**
- б) Полуструктурированными.
- в) Квазиструктурированными.
- г) Неструктурированными.

8. Данные текстовых файлов с определенными паттернами для их обработки (например, XML) являются:
- Структурированными.
 - Полуструктурированными.**
 - Квазиструктурированными.
 - Неструктурированными.
9. Выделение таблицы или группы таблиц на отдельный сервер это...
- Горизонтальное масштабирование.
 - Вертикальное масштабирование.
 - Горизонтальный шардинг.
 - Вертикальный шардинг.**
10. Комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке: аномалий, пропусков, дубликатов, противоречий, шумов и шумов и т.д. _____

Ответ: **Очистка данных**

11. Технология машинного обучения, когда нет ответов и требуется искать зависимости между объектами, называется _____

Ответ: **Обучение без учителя**

12. Ассоциативное правило можно найти с помощью алгоритма _____

Ответ: **apriori**

13. 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: _____

Ответ: **горизонтальное масштабирование**

14. Что означает термин NoSQL? _____

Ответ: **Не только SQL**

15. Для достижения какого свойства в БД типа NoSQL нет JOIN операций? _____

Ответ: **Consistency**

16. Сколько уровней имеет лямбда-архитектура? _____

Ответ: **3**

17. Какая компания создала технологию MapReduce? _____

Ответ: Google

18. Процессом создания и выбора модели для предсказания вероятности наступления некоторого события является: _____

Ответ: **Predictive Modelling**

3.4 Перечень теоретических вопросов к экзамену

(для оценки знаний)

- Понятие «большие данные». Характеристики больших данных. Принципы работы с большими данными.
- Большие данные в научных исследованиях. Экосистема больших данных.
- Технологии хранения больших данных.
- Распределенные файловые системы.
- Распределенные фреймворки.
- Хранилища, классификация и примеры.
- Распределенные базы данных NoSQL и новые SQL базы данных.
- 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).
- Параллельные алгоритмы для работы с данными.
- Операторы Map и Reduce. Оператор Map (предварительная обработка). Оператор Reduce (свертка).
- 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.

3.5 Перечень типовых простых практических заданий к экзамену (для оценки умений)

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.

3.6 Перечень типовых практических заданий к экзамену (для оценки навыков и (или) опыта деятельности)

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.

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.

Наименование оценочного средства	Описания процедуры проведения контрольно-оценочного мероприятия и процедуры оценивания результатов обучения
Собеседование	Собеседование, предусмотренное рабочей программой дисциплины, проводится на практическом занятии. Преподаватель на практическом занятии, предшествующем занятию проведения контроля, доводит до обучающихся тему, вопросы для подготовки к собеседованию. Результаты собеседования преподаватель доводит до обучающихся сразу после завершения собеседования
Лабораторная работа	Защита лабораторных работ проводится во время лабораторных занятий. Во время проведения защиты лабораторной работы пользоваться учебниками, справочниками, конспектами лекций, тетрадями не разрешено. Преподаватель на лабораторной работе, предшествующей занятию проведения защиты лабораторной работы, доводит до обучающихся: номер защищаемой лабораторной работы, время на защиту лабораторной работы. Преподаватель информирует обучающихся о результатах защиты лабораторной работы сразу после ее контрольно-оценочного мероприятия

Для организации и проведения промежуточной аттестации 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 и типовые практические задания разного уровня сложности для проведения промежуточной аттестации 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 или в форме компьютерного тестирования.

При проведении промежуточной аттестации в форме собеседования билеты составляются таким образом, чтобы 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; два практических задания: одно из них для оценки умений (выбирается из перечня типовых простых практических заданий к экзамену); другое практическое задание для оценки навыков и (или) опыта деятельности (выбираются из перечня типовых практических заданий к экзамену).

Распределение теоретических вопросов и практических заданий 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 фондов оценочных средств.


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

При проведении промежуточной аттестации в форме компьютерного тестирования вариант тестового задания формируется из фонда тестовых заданий по дисциплине случайным

образом, но с условием: 50 % заданий должны быть заданиями открытого типа и 50 % заданий – закрытого типа.

Образец экзаменационного билета

 <p>ИРГУПС 20__-20__ учебный год</p>	<p>Экзаменационный билет № 1 по дисциплине «<u>Обработка и анализ больших данных</u>»</p>	<p>Утверждаю: Заведующий кафедрой «_____» ИРГУПС _____</p>
<ol style="list-style-type: none">1. Распределенные файловые системы.2. Параллельные алгоритмы для работы с данными.3. Приведите последовательность шагов в алгоритме k-средних.4. 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.		