

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИНЖЕНЕРНЫХ ТЕХНОЛОГИЙ»**

УТВЕРЖДАЮ
Проректор по учебной работе

_____ Василенко В.Н.
(подпись) (ф.и.о.)

«25» мая 2023 г.

РАБОЧАЯ ПРОГРАММА

дисциплины

Большие данные

Направление подготовки
09.03.02 Информационные системы и технологии

Направленность (профиль) подготовки
Моделирование и проектирование информационных технологий и систем

Квалификация выпускника
Бакалавр

Воронеж

1. Цели и задачи дисциплины

Целью освоения дисциплины является формирование компетенций обучающегося в области профессиональной деятельности и сфере профессиональной деятельности: 06 Связь, информационные и коммуникационные технологии

06 Связь, информационные и коммуникационные технологии

Дисциплина направлена на решение задач профессиональной деятельности следующих типов:

производственно-технологический;

организационно-управленческий;

проектный

Программа составлена в соответствии с требованиями Федерального государственного образовательного стандарта высшего образования по направлению подготовки 09.03.02 Информационные системы и технологии

2. Перечень планируемых результатов обучения, соотнесенных с планируемыми результатами освоения образовательной программы

№п /п	Код компетенции	Наименование компетенции	Код и наименование индикатора достижения компетенции
1	УК-2	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	ИД2 _{УК-2} – Проектирует и выбирает оптимальные способы решения определенных задач, исходя из действующих правовых норм, имеющихся ресурсов и ограничений и публично представляет результаты решения конкретной задачи проекта
2	УК-6	Способен управлять своим временем, выстраивать и реализовывать траекторию саморазвития на основе принципов образования в течение всей жизни	ИД1 _{УК-6} – Применяет знание о своих ресурсах и их пределах для успешного выполнения порученной работы и критически оценивает эффективность использования личного времени при решении поставленных задач в целях достижения планируемого результата
3	ОПК-2	Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности	ИД3 _{ОПК-2} – Решает стандартные задачи с применением современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности.

Код и наименование индикатора достижения компетенции	Результаты обучения (показатели оценивания)
ИД2 _{УК-2} – Проектирует и выбирает оптимальные способы решения определенных задач, исходя из действующих правовых норм, имеющихся ресурсов и ограничений и публично представляет результаты решения конкретной задачи проекта	Знает методы формирования и интеграции исходных данных в рамках поставленной задачи
	Умеет обеспечивать основной результат проекта, формировать перечни работ по проекту, определять и согласовывать критерии успешности реализации проекта, осуществлять планирование проекта
ИД1 _{УК-6} – Применяет знание о своих ресурсах и их пределах для успешного выполнения порученной работы и критически оценивает эффективность использования личного времени при решении поставленных задач в целях достижения планируемого результата	Владеет навыками применения нормативно-правовую документацию в сфере профессиональной деятельности
	Знает требования к программному обеспечению, применяемому при решении задач в сфере профессиональной деятельности
ИД3 _{ОПК-2} – Решает стандартные задачи с применением современных информационных технологий и программных средств, в том числе отечественного производства, при	Умеет выявлять взаимосвязи между входной и выходной информацией; проводить систематизацию требований к программному обеспечению.
	Владеет навыками документирования результатов сбора и систематизации требований к программному обеспечению.
ИД3 _{ОПК-2} – Решает стандартные задачи с применением современных информационных технологий и программных средств, в том числе отечественного производства, при	Знает современные информационные технологии и программные средства, применяемые при решении задач профессиональной деятельности.
	Умеет применять современные информационные технологии и программные средства, используемые при решении задач

решении задач профессиональной деятельности.деятельности.	профессиональной деятельности. Владеет навыками применения современных информационных технологий и программных средств.
---	--

3. Место дисциплины в структуре ООП ВО

Дисциплина «Большие данные» относится к блоку1 ОП и ееобязательной части.

Изучение дисциплины основано на знаниях, умениях и навыках, полученных при изучении обучающимися дисциплин «Математика», «Статистическое моделирование информационных процессов», «Программирование на языке высокого уровня».

Дисциплина является предшествующей для изучения: «Теория принятия решений в информационных системах», «Производственная практика, преддипломная практика» и при работе над ВКР.

4. Объем дисциплины и виды учебных занятий

Общая трудоемкость дисциплины составляет 5 зачетных единицы.

Виды учебной работы	Всего часов (акад.)	Семестр 6 (акад.)
Общая трудоемкость дисциплины	180	180
Контактная работа, в т.ч. ауд. занятия:		
Лекции	36	36
<i>в том числе в форме практической подготовки</i>	-	-
Практические занятия (ПЗ)	36	36
<i>в том числе в форме практической подготовки</i>	-	-
Консультации текущие	3,8	3,8
Виды аттестации: экзамен	0,2	0,2
Самостоятельная работа:	70,2	70,2
Проработка материала по конспекту лекций (подготовка к тестированию, собеседование)	20	20
Проработка материала по учебнику (подготовка к тестированию, собеседование)	22,2	22,2
Домашняя работа	10	10
Подготовка к защите работ, выполняемых на практических работах	18	18
Контроль	33,8	33,8

5 Содержание дисциплины

5.1 Содержание разделов дисциплины

№ п/п	Наименование раздела дисциплины	Содержание раздела
1	Большие данные	Понятие о BigData. Причины появления больших данных. Примеры возможностей для промышленности. Понятие жизненного цикла аналитики данных. Когнитивный анализ данных. Парадигма MapReduce. Проекты ApacheHadoop, ApacheSpark.
2	Машинное обучение	Классификация задач обработки данных. Информативность и выбор признаков. Применение машинного обучения в аналитике. K-means и C-means кластеризация, классификация.
3	Работа с данными	Основы языка R: основные функции, синтаксис, типы данных. DataMininng: инструменты, основные библиотеки. Технологии хранения больших данных. Парадигма NoSQL.
	Консультации текущие	3,8
	экзамен	0,2

5.2 Разделы дисциплины виды занятий

№ п/п	Наименование раздела дисциплины	Лекции, час	ПЗ, час	СРО, час			
				Проработка материала по конспекту лекций (подготовка к тестированию, собеседование)	Проработка материала по учебнику (подготовка к тестированию, собеседование)	Домашняя работа	Подготовка к защите работ, выполняемых на практических работах
1	Большие данные	12	12	7	6,2	2	6

2	Машинное обучение	12	12	7	8	4	6
3	Работа с данными	12	12	6	8	4	6
Консультации текущие		3,8					
экзамен		0,2					

5.2.1 Лекции

№ п/п	Наименование раздела дисциплины	Тематика лекционных занятий	Трудоемкость, час
1	Большие данные	Понятие о BigData. Причины появления больших данных. Примеры возможностей для промышленности. Понятие жизненного цикла аналитики данных. Когнитивный анализ данных. Парадигма MapReduce. Проекты ApacheHadoop, ApacheSpark.	12
2	Машинное обучение	Классификация задач обработки данных. Информативность и выбор признаков. Применение машинного обучения в аналитике. K-means и C-means кластеризация, классификация.	12
3	Работа с данными	Основы языка R: основные функции, синтаксис, типы данных. DataMininng: инструменты, основные библиотеки. Технологии хранения больших данных. Парадигма NoSQL.	12

5.2.2 Практические занятия

№ п/п	Наименование раздела дисциплины	Тематика практических занятий	Трудоемкость, час
1	Большие данные	Жизненный цикл аналитики данных. Когнитивный анализ данных. MapReduce. ApacheHadoop, ApacheSpark.	12
2	Машинное обучение	Задачи классификации и кластеризации. K-means и C-means кластеризация, классификация.	12
3	Работа с данными	Основы языка R: основные функции, синтаксис, типы данных языка R. Инструменты, основные библиотеки для работы с данными.	12

5.3.3 Лабораторный практикум не предусмотрен.

5.2.4 Самостоятельная работа обучающихся (СРО)

№ п/п	Наименование раздела дисциплины	Вид СРО	Трудоемкость, час
Четвертый семестр			
1.	Большие данные	Проработка материала по конспекту лекций	7
		Проработка материала по учебнику	6,2
		Домашняя работа	2
		Подготовка к защите работ, выполняемых на практических работах	6
2.	Машинное обучение	Проработка материала по конспекту лекций (подготовка к тестированию, собеседование)	7
		Проработка материала по учебнику (подготовка к тестированию, собеседование)	8
		Домашняя работа	4
		Подготовка к защите работ, выполняемых на практических работах	6
3.	Работа с данными	Проработка материала по конспекту лекций (подготовка к тестированию, собеседование)	6
		Проработка материала по учебнику (подготовка к тестированию, собеседование)	8
		Домашняя работа	2
		Подготовка к защите работ, выполняемых на практических работах	6

6 Учебно-методическое обеспечение и информационное обеспечение дисциплины

6.1 Основная литература

1. Колемаев, В.А. Теория вероятностей и математическая статистика : учебник / В.А. Колемаев, В.Н. Калинина. - Москва :Юнити-Дана, 2015. - 352 с. : табл. - ISBN 5-238-00560-1 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=436721>

2. Гусева, Е.Н. Теория вероятностей и математическая статистика : учебное пособие / Е.Н. Гусева. - Москва : Флинта, 2011. - 220 с. - ISBN 978-5-9765-1192-7 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=83543>

6.2 Дополнительная литература

1. Лисьев, В.П. Теория вероятностей и математическая статистика : учебное пособие / В.П. Лисьев. - Москва : Евразийский открытый институт, 2010. - 200 с. - ISBN 5-374-00005-5 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=90420>

6.3 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся.

Никитин Б. Е., Методические указания к самостоятельной работе обучающихся по дисциплине [Электронный ресурс] : для студентов, обучающихся по направлению 09.03.02 очной и заочной формы обучения / Б. Е. Никитин; ВГУИТ, Кафедра информационных технологий, моделирования и управления. - Воронеж : ВГУИТ, 2015. - Режим доступа <https://education.vsu.ru/mod/glossary/view.php?id=48526>

6.4 Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

Наименование ресурса сети «Интернет»	Электронный адрес ресурса
«Российское образование» - федеральный портал	http://www.edu.ru/index.php
Научная электронная библиотека	http://www.elibrary.ru/defaulttx.asp?
Федеральная университетская компьютерная сеть России	http://www.runnet.ru/
Информационная система «Единое окно доступа к образовательным ресурсам»	http://www.window.edu.ru/
Электронная библиотека ВГУИТ	http://biblos.vsu.ru/megapro/web
Сайт Министерства науки и высшего образования РФ	http://minobrnauki.gov.ru
Портал открытого on-line образования	http://npoed.ru
Информационно-коммуникационные технологии в образовании. Система федеральных образовательных порталов	http://www.ict.edu.ru/
Электронная образовательная среда ФГБОУ ВО «ВГУИТ	http://education.vsu.ru

6.5 Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем. При освоении дисциплины используется лицензионное и открытое программное обеспечение.

При изучении дисциплины используется программное обеспечение и информационные справочные системы: информационная среда для дистанционного обучения «Moodle», автоматизированная информационная база «Интернет-тренажеры».

При освоении дисциплины используется лицензионное и открытое программное обеспечение

Программы	Лицензии, реквизиты подтверждающего документа
Microsoft WindowsXP	Microsoft Open License Microsoft WindowsXP Academic OPEN No Level #44822753 от 17.11.2008 http://eopen.microsoft.com
Microsoft Windows 8.1 (64 - bit)	Microsoft Open License Microsoft Windows Professional 8 Russian Upgrade Academic OPEN 1 License No Level#61280574 от 06.12.2012 г. http://eopen.microsoft.com
MicrosoftOffice 2007	Microsoft Office 2007 Russian Academic OPEN No Level #44822753 от 17.11.2008 http://eopen.microsoft.com
MicrosoftOffice 2010	Microsoft Office 2010 Russian Academic OPEN 1 License No Level #47881748 от 24.12.2010 г. http://eopen.microsoft.com
AdobeReaderXI	(бесплатноеПО) https://acrobat.adobe.com/ru/ru/acrobat/pdf-reader/volumedistribution.htm
Альт Образование 8.2 + LibreOffice 6.2+Maxima	Лицензия № ААА.0217.00 с 21.12.2017 г. по «Бессрочно»

7 Материально-техническое обеспечение дисциплины

Аудитории для проведения лекционных занятий (а.334)	Комплекты мебели для учебного процесса- 30 шт.	
Читальные залы библиотеки	Компьютеры со свободным доступом в сеть Интернет и Электронными библиотечными и информационно справочными системами.	
Аудитории для проведения практических занятий (а.339 - компьютерный класс каф.ИТМУ)	Количество ПЭВМ -12 (Coreis 540), проектор – 1 (ViewSonikPJD5255)	Microsoft Office 2007, Microsoft Office Professional Plus 2007 (Access, Visio, Project), Microsoft Share Point Designer 2013 Russian Academic OPEN No Level #44822753 от 17.11.2008 http://eopen.microsoft.com
Помещение для самостоятельной работы (а.339)	Количество ПЭВМ -12 (Coreis 540), проектор – 1 (ViewSonikPJD5255)	Microsoft Office 2007, Microsoft Office Professional Plus 2007 (Access, Visio, Project), Microsoft Share Point Designer 2013 Russian Academic OPEN No Level #44822753 от 17.11.2008 http://eopen.microsoft.com

8. Оценочные материалы для промежуточной аттестации обучающихся по дисциплине (модулю)

Оценочные материалы (ОМ) для дисциплины включают в себя:

- перечень компетенций с указанием индикаторов достижения компетенций, этапов их формирования в процессе освоения образовательной программы;
- описание шкал оценивания;
- типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков;
- методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности.

ОМ представляются отдельным комплектом и **входят в состав рабочей программы дисциплины.**

Оценочные материалы формируются в соответствии с П ВГУИТ «Положение об оценочных материалах».

ПРИЛОЖЕНИЕ
к рабочей программе
Направление подготовки
09.03.02.- Информационные системы

Квалификация (степень) выпускника
Бакалавр

Форма обучения **заочная**

1. Организационно-методические данные дисциплины для заочной формы обучения

1.1 Объемы различных форм учебной работы и виды контроля в соответствии с учебным планом

Виды учебной работы	Всего часов (акад.)	Семестр 8 (акад.)
<i>Контактная работа, в т.ч. аудит. зан.:</i>		
Лекции	4	4
Практические занятия (ПЗ)	4	4
Текущие консультации	3,4	3,4
Виды аттестации: экзамен	0,2	0,2
<i>Самостоятельная работа:</i>	161,6	161,6
Проработка материала по конспекту лекций (подготовка к тестированию, собеседование)	55	55
Проработка материала по учебнику (подготовка к тестированию, собеседование)	80,6	80,6
Контрольная работа	15	15
Подготовка к защите работ, выполняемых на практических работах	11	11
Контроль	6,8	6,8
Общая трудоемкость дисциплины	180	180

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ
ДЛЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ**

по дисциплине

Большие данные

1 Перечень компетенций с указанием этапов их формирования

№п/п	Код компетенции	Наименование компетенции	Код и наименование индикатора достижения компетенции
1	2	3	4
1	УК-2	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	ИД2 _{УК-2} – Проектирует и выбирает оптимальные способы решения определенных задач, исходя из действующих правовых норм, имеющихся ресурсов и ограничений и публично представляет результаты решения конкретной задачи проекта
2	УК-6	Способен управлять своим временем, выстраивать и реализовывать траекторию саморазвития на основе принципов образования в течение всей жизни	ИД1 _{УК-6} – Применяет знание о своих ресурсах и их пределах для успешного выполнения порученной работы и критически оценивает эффективность использования личного времени при решении поставленных задач в целях достижения планируемого результата
3	ОПК-2	Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности	ИД3 _{ОПК-2} – Решает стандартные задачи с применением современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности

Код и наименование индикатора достижения компетенции	Результаты обучения (показатели оценивания)
ИД2 _{УК-2} – Проектирует и выбирает оптимальные способы решения определенных задач, исходя из действующих правовых норм, имеющихся ресурсов и ограничений и публично представляет результаты решения конкретной задачи проекта	Знает технологии хранения и обработки больших данных
	Умеет формулировать задачи в сфере анализа больших данных
	Владеет навыками использования современных информационных технологий для решения прикладных задач
ИД1 _{УК-6} – Применяет знание о своих ресурсах и их пределах для успешного выполнения порученной работы и критически оценивает эффективность использования личного времени при решении поставленных задач в целях достижения планируемого результата	Знает основные тенденции развития цифровых технологий в области анализа больших объемов данных
	Умеет использовать результаты анализа данных для принятия эффективных решений при согласовании требований к ИС с заказчиком
	Владеет навыками использования современных методов анализа больших данных
ИД3 _{ОПК-2} – Решает стандартные задачи с применением современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности	Знает специфику анализа больших данных
	Умеет проводить сравнительный анализ и выбор цифровых технологий для решения прикладных задач
	Владеть инструментальными средствами обработки статистической информации, а также навыками применения вероятностно - статистического подхода к постановке и решению прикладных задач

2 Паспорт оценочных материалов по дисциплине

№ п/п	Разделы дисциплины	Индекс контролируемой компетенции (или ее части)	Оценочные средства		Технология/процедура оценивания (способ контроля)
			наименование	№№ заданий	
1	Большие данные	УК-6	Тест	15-19	Компьютерное тестирование Процентная шкала. 0-100 %; 0-59,99% - неудовлетворительно; 60-74,99% - удовлетворительно; 75- 84,99% -хорошо; 85-100% - отлично.
			Собеседование (вопросы для зачета)	47-50	Проверка преподавателем Отметка в системе «зачтено – не зачтено»
2	Машинное обучение	УК-6, ОПК-2	Тест	1-8, 20-40	Компьютерное тестирование Процентная шкала. 0-100 %; 0-59,99% - неудовлетворительно; 60-74,99% - удовлетворительно; 75- 84,99% -хорошо; 85-100% - отлично.
			Собеседование (вопросы для зачета)	53-58	Проверка преподавателем Отметка в системе «зачтено – не зачтено»
			Кейс-задание	41-43	Проверка преподавателем Отметка в системе «зачтено – не зачтено»
3	Работа с данными	УК-2	Тест	9-14	Компьютерное тестирование Процентная шкала. 0-100 %; 0-59,99% - неудовлетворительно; 60-74,99% - удовлетворительно; 75- 84,99% -хорошо; 85-100% - отлично.
			Собеседование (вопросы для зачета)	51-52	Проверка преподавателем Отметка в системе «зачтено – не зачтено»
			Домашнее задание	44-46	Проверка преподавателем Отметка в системе «зачтено – не зачтено»

3 Оценочные материалы для промежуточной аттестации

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

Для оценки знаний, умений, навыков студентов по дисциплине применяется бально-рейтинговая система оценки сформированности компетенций студента.

Бально-рейтинговая система оценки осуществляется в течение всего семестра при проведении аудиторных занятий и контроля самостоятельной работы.

Показателями ОМ являются: текущий опрос в виде собеседования на лабораторных работах, тестовые задания и самостоятельно (домашнее задание). Оценки выставляются в соответствии с графиком контроля текущей успеваемости студентов в автоматизированную систему баз данных (АСУБД) «Рейтинг студентов».

Обучающийся, набравший в семестре более 60 % от максимально возможной бально-рейтинговой оценки работы в семестре получает зачет автоматически.

Студент, набравший за текущую работу в семестре менее 60 %, т.к. не выполнил всю работу в семестре по объективным причинам (болезнь, официальное освобождение и т.п.) допускается до зачета, однако ему дополнительно задаются вопросы на собеседовании по разделам, выносимым на зачет.

Аттестация обучающегося по дисциплине проводится в форме тестирования и предусматривает возможность последующего собеседования (зачета). Зачет проводится в виде тестового задания.

Каждый вариант теста включает 30 контрольных заданий, из них:

- 10 контрольных заданий на проверку знаний;
- 10 контрольных заданий на проверку умений;
- 10 контрольных заданий на проверку навыков;

В случае неудовлетворительной сдачи зачета студенту предоставляется право повторной сдачи в срок, установленный для ликвидации академической задолженности по итогам соответствующей сессии. При повторной сдаче зачета количество набранных студентом баллов на предыдущем зачете не учитывается.

3.1 Тесты (тестовые задания и кейс-задания)

3.1.1 Шифр и наименование компетенции

ОПК-2 Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности

№ задания	Тестовое задание
1.	Какие оконные функции можно использовать для того, чтобы сравнивать текущее значение строки с предыдущим или следующим? Выберите один вариант: 1. LAG и LEAD 2. RANK и DENSE_RANK 3. NTAILE и NTILE
2.	Какой оператор используется для возврата в первом операторе SELECT всех строк, которые не возвращаются вторым оператором SELECT? Выберите один вариант: 1. UNION ALL 2. MINUS 3. INTERSECT
3.	При каком условии выполняется оператор соединения UNION? Выберите один вариант: 1. Запросы имеют одно общее поле 2. Равное количество столбцов в обоих запросах 3. Запросы имеют одинаковое количество строк
4.	Какая команда используется для создания виртуальной таблицы? Выберите один вариант: 1. CREATE VIEW 2. CREAT VIRTUAL TABLE 3. ALTER VIEW
5.	Какой оператор объединяет таблицы без удаления дубликатов? Выберите один вариант: 1. UNION ALL 2. UNION 3. FULL JOIN

6.	<p>Зачем нужен оператор INTERSECT? Выберите один вариант:</p> <ol style="list-style-type: none"> 1. Для удаления дублированных строк 2. Для нахождения пересечения двух множеств 3. Для полного соединения двух таблиц
7.	<p>В чем разница между операторами UNION и JOIN? Выберите один вариант:</p> <ol style="list-style-type: none"> 1. UNION можно использовать только для объединения подзапросов 2. UNION возвращает дубликаты, а JOIN - нет 3. JOIN объединяет столбцы, а UNION – строки таблиц
8.	<p>Структурированные данные - это</p> <p>Выберите один вариант:</p> <ol style="list-style-type: none"> 1. Данные, которые соответствуют общеизвестным форматам (doc, docx, pdf, jpeg и тд) 2. Данные с чётко заданным форматом и однозначным обращением к ним 3. Данные с гибким форматом, в котором информация не представлена в однозначном виде
УК-2 Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	
9.	<p>Неструктурированные данные - это</p> <ol style="list-style-type: none"> 1. Данные, которые соответствуют общеизвестным форматам (doc, docx, pdf, jpeg и тд) 2. Данные с чётко заданным форматом и однозначным обращением к ним 3. Данные с гибким форматом, в котором информация не представлена в однозначном виде
10.	<p>Какие наборы признаков заложены в 3V?</p> <ol style="list-style-type: none"> 1. vocal, velocity, variety 2. vacation, velocity, variety 3. volume, velocity, variety 4. violence, velocity, variety
11.	<p>В какой компании создали MapReduce?</p> <ol style="list-style-type: none"> 1. Google 2. Yahoo 3. Amazon 4. Facebook
12.	<p>Из чего состоит вычислительная модель MapReduce?</p> <ol style="list-style-type: none"> 1. Map, reduce 2. Map, resort, reduce 3. Map, shuffle, reduce 4. Map, calculation, reduce
13.	<p>Из каких узлов состоит MapReduce?</p> <ol style="list-style-type: none"> 1. Master, work 2. Master, slave 3. Master, pork
14.	<p>Где используется файловая система HDFS?</p> <ol style="list-style-type: none"> 1. Apache Hadoop 2. Oracle DB 3. Mongo DB
УК-6 Способен управлять своим временем, выстраивать и реализовывать траекторию саморазвития на основе принципов образования в течение всей жизни	
15.	<p>Из чего состоит архитектура master/slave HDFS?</p> <ol style="list-style-type: none"> 1. MainNode, StorageNodes, Client 2. MainNode, DataNodes, Client 3. NameNode, DataNodes, Client 4. NameNode, DataNodes, Patient
16.	<p>Power BI – это</p> <ol style="list-style-type: none"> 1. Семейство продуктов от Microsoft для бизнес-аналитики 2. Десктопное приложение от Microsoft для бизнес-аналитики 3. Надстройка для MS Excel 4. Веб-приложение от Microsoft для бизнес-аналитики
17.	<p>Power BI предназначен для:</p> <ol style="list-style-type: none"> 1. Создания приложений 2. Обработки данных 3. Визуализации 4. Извлечения данных из источников

18.	Верно ли следующее утверждение? Бизнес-аналитические задачи можно решать только при помощи Power BI. 1. Да 2. Нет
19.	Power Query предназначен для: 1. Создания модели данных и связей между ними 2. Извлечения, преобразования и очистки данных
ОПК-2 Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности	
20.	Чем отличаются меры от вычисляемых столбцов? 1. В случае мер хранится только формула 2. Меры хранятся со всей моделью 3. Меры рассчитывается для каждой строки таблицы 4. Меры вычисляются только при использовании 5. Меры вычисляются моментально после создания
21.	Дах – это 1. Язык запросов для Power Pivot 2. Язык запросов для Power Query
22.	M language – это 1. Язык запросов для Power Pivot 2. Язык запросов для Power Query
23.	Как выбирать визуализацию в Power BI? 1. Используя готовые шаблоны выбора визуализаций 2. По принципу "что больше нравится" 3. Поняв предварительно, какие данные визуализируем (сравнение, распределение, связь переменных, структуру и т.д.), и кто будет «потребителем» результатов анализа
24.	Почему Apache Spark называют "ленивым"? 1. Трансформации данных накапливаются, а исполняются лишь при вызове действия 2. Самые затратные трансформации не исполняются 3. Spark работает медленнее Pandas 4. Трансформацию данных всегда нужно ждать большое количество времени
25.	Что такое "драйвер" в Apache Spark? 1. Связующее звено между распределителем ресурсов и контроллером отказоустойчивости приложения 2. Процесс, выполняющий задачи и конфигурирующий работу приложения 3. Планировщик исполнения запросов к базам данных 4. Программа, которая создает SparkContext и распределяет задачи между исполнителями
26.	Что делают исполнители при отказе работы одного из них? 1. Перехватывают работу на рабочие узлы 2. Завершают работу основной программы 3. Останавливаются и ждут перезагрузки вышедшего из строя исполнителя 4. Перезагружают программу и начинают всю работу заново
27.	Формат хранения данных, разработанный для Apache Spark – это 1. Xlsx 2. Csv 3. Json 4. Parquet
28.	Что такое А/В тестирование? 1. Метод неформальной оценки 2. Метод проверки идей или гипотез 3. Алгоритмическое тестирование 4. Вербальное тестирование
29.	Какие критерии позволяют сказать, есть ли статистически значимые различия между двумя группами? 1. Критерий Стьюдента 2. Критерий Манна-Уитни 3. Критерий Менделеева-Клайперона 4. Критерий Шапиро-Уилка

30.	Какой критерий является легко интерпретируемым? 1. Критерий Уилкоксона 2. Критерий Коши 3. t- критерий Стьюдента 4. Критерий Манна-Уитни
31.	Какой метод лучше подходит для поиска различий между 10 группами? 1. Критерий Уилкоксона 2. Критерий Манна-Уитни 3. Дисперсионный анализ 4. t- критерий Стьюдента
32.	Какой из критериев чувствителен к распределению в данных? 1. t- критерий Стьюдента 2. Критерий Уилкоксона 3. Дисперсионный анализ 4. Критерий Манна-Уитни
33.	Какую архитектуру модели стоит использовать для определения вероятности оттока клиента? 1. Градиентный бустинг 2. Модель арифметической прогрессии 3. Модель геометрической прогрессии 4. Нейросеть
34.	Какую задачу должна решать ML-модель при определении точной стоимости объекта недвижимости? 1. Персонализация 2. Регрессия 3. Классификация 4. Кластеризация
35.	Как называется математический подход, при котором определяются "плотные острова" в неразмеченных данных? 1. Машина опорных векторов 2. Кластеризация 3. Логистическая регрессия 4. Персонализация
36.	Какой тип архитектуры используется для машинного перевода? 1. Трансформер 2. Байесовский классификатор 3. Метод ближайших соседей 4. Рекуррентная нейросеть 5. Логистическая регрессия
37.	Вы построили нейронную сеть, в которой к входным данным добавляется выход от предыдущего слоя. Как называется данная архитектура? 1. Сверточная нейронная сеть 2. Рекуррентная нейронная сеть
38.	При решении какой задачи обычно используют сверточные нейронные сети? 1. При решении задачи детекции мошеннических операций на данных по финансовым транзакциям 2. При решении задачи детекции автомобилей на изображениях
39.	Может ли помочь автоэнкодер служить инструментом для понижения размерности? 1. Может 2. Не может
40.	Какую проблему помогает решить padding в операции свертки? 1. Позволяет уменьшить входное изображение путем выделения максимального элемента из блока 2. Позволяет чаще задействовать данные на границах изображения в операциях свертки

3.2. Кейс- задания (примеры)

ОПК-2 Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности

№ задания	Формулировка задачи
41.	<p>Дана выгрузка по продажам компании-поставщика тяжелой техники в одном из регионов России. Требуется посчитать какой доход принесла каждая продаваемая модель техники.</p> <p>В среде MSExcel (версия не ниже 10) решение нужно выполнить следующие действия:</p> <ol style="list-style-type: none"> 1. Отформатировать выгрузку как таблицу. 2. Функцией ВПР «вытянуть» в выгрузку цены из листа с ценами моделей 3. Добавить в выгрузку столбец с суммой сделки 4. На листе с ценами моделей добавить столбец с суммарной выручкой по каждой модели (используя функцию СУММЕСЛИ). 5. Добавить строку Итог.
42.	<p>Дана выгрузка по продажам компании-поставщика тяжелой техники в одном из регионов России. Требуется узнать сколько денег заработала фирма в каждом городе на каждой модели в каждом месяце каждым продавцом.</p> <p>В среде MSExcel (версия не ниже 10) нужно выполнить следующие действия:</p> <ol style="list-style-type: none"> 1. Подключится к исходному диапазону с помощью Power Query 2. Подключится к диапазону с ценами с помощью Power Query 3. Объединить диапазоны из п.п. 1 и 2 4. Добавить столбец с произведением цены на количество 5. Построить сводную таблицу с отображением выручки в разрезе городов, месяцев, продавцов и товаров
43.	<p>Дана выгрузка по продажам компании-поставщика тяжелой техники в одном из регионов России. Требуется узнать сколько денег заработала фирма в каждом городе на каждой модели в каждом месяце каждым продавцом.</p> <p>В среде MSExcel (версия не ниже 10) нужно выполнить следующие действия:</p> <ol style="list-style-type: none"> 1. Загрузить исходные данные в Power Query 2. В редакторе Power Query заполнить пропущенные данные в столбце “Модель” (<i>Преобразование/Заполнить/Вниз</i>) 3. Коме первых двух столбцов (“Модель” и “Дата продажи”) к остальным применить <i>Отменить свертывание столбцов/ Отменить свертывание других</i> 4. Изменить наименование получившихся двух последних столбцов На “Город” и “Количество” соответственно. 5. Загрузить данные в Power Pivot (в усиленную сводную таблицу) 6. Связать таблицы <i>Выгрузка</i> и <i>Цены моделей</i> (<i>Представление диаграммы</i> в группе <i>Просмотр</i>). 7. Добавить и связать третью таблицу <i>Calendar</i> (<i>Конструктор/Таблица дат/Создать</i>) 8. На языке DAX ввести формулу для расчета выручки: 9. Создать сводную таблицу и сохранить. 10. Импортировать в Power BI сводную таблицу. 11. Создать интерактивный отчет (интерактивные элементы <i>Карта, Гистограмма, Срез</i>)

3.3 Домашнее задание (примеры)

УК-2 Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений

№ задания	Формулировка задачи
44.	<ol style="list-style-type: none"> 1. Найти набор данных в репозитории Kaggle [2] или [3] и скачать его на свой компьютер 2. Выполнить загрузку в разных вариантах: полной, отдельных столбцов, части строк 3. Выполнить обработку данных и сохранить результаты во внешнем файле

45.	Изучить основные функции библиотеки numpy и научиться их использовать при выполнении операций с многомерными массивами
46.	Изучить функции библиотеки numpy по агрегированию данных. Научиться выполнять агрегирование данных, содержащихся в многомерных массивах.

3.4 Вопросы к зачёту

УК-6 Способен управлять своим временем, выстраивать и реализовывать траекторию саморазвития на основе принципов образования в течение всей жизни

№	Текст вопроса
47.	Понятие Больших данных
48.	Особенности сбора, хранения, обработки и анализа Больших данных.
49.	Требования к распределенным информационным системам
50.	Средства построения распределенных информационных систем
УК-2 Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	
51.	Технология Map-Reduce
52.	Система Apache Hadoop
ОПК-2 Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности	
53.	Базы данных NoSQL. Особенности, классификация.
54.	Возможности NoSQL-баз данных по обеспечению целостности, доступности.
55.	Способы репликации и кластеризации баз данных.
56.	Документо-ориентированные базы данных
57.	Возможности СУБД MongoDB
58.	Работа с документо-ориентированными БД.

Критерии и шкалы оценки:

- **оценка «зачтено»** выставляется студенту, если домашнее задание является самостоятельным, оригинальным текстом, в котором прослеживается авторская позиция, продуманная система аргументов, а также наличествуют обоснованные выводы; используются термины, понятия по дисциплине, в рамках которой выполняется работа; полностью соответствует выбранной теме, цели и задачам; текст домашнего задания логически выстроен, имеет четкую структуру; работа соответствует всем техническим требованиям; домашнее задание выполнено в установленный срок.

- **оценка «не зачтено»**, выставляется студенту, если домашнее задание не является самостоятельным, оригинальным текстом, в котором не прослеживается авторская позиция, не продумана система аргументов, а также отсутствуют обоснованные выводы; не используются термины, понятия по дисциплине, в рамках которой выполняется работа; не соответствует выбранной теме, цели и задачам; текст домашнего задания композиционно не выстроен; работа не соответствует техническим требованиям; домашнее задание не выполнено в установленный срок.

4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Процедуры оценивания в ходе изучения дисциплины знаний, умений и навыков, характеризующих этапы формирования компетенций, регламентируются положениями:

- П ВГУИТ 2.4.03 Положение о курсовых экзаменах и зачетах;
- П ВГУИТ 4.1.02 Положение о рейтинговой оценке текущей успеваемости.

Для оценки знаний, умений, навыков обучающихся по дисциплине применяется рейтинговая система. Итоговая оценка по дисциплине определяется на основании определения среднеарифметического значения баллов по каждому заданию.

Зачет по дисциплине выставляется в зачетную ведомость по результатам работы в семестре после выполнения всех видов учебной работы, предусмотренных рабочей программой дисциплины (с

отметкой «зачтено») и получении по результатам тестирования по всем разделам дисциплины не менее 60 %.

5. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания для каждого результата обучения по дисциплине

Результаты обучения по этапам формирования компетенций	Предмет оценки (продукт или процесс)	Показатель оценивания	Критерии оценивания сформированности компетенций	Шкала оценивания	
				Академическая оценка или баллы	Уровень освоения компетенции
УК-2 Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений					
Знать технологии хранения и обработки больших данных	Собеседование (зачет)	Корректно работает с разнородными данными	Обучающийся твердо знает материал, грамотно и по существу излагает его, но допускает в ответе некоторые неточности	Зачтено/ 60-100	Освоена (базовый)
			Обучающийся не раскрыл содержание материала, допускает грубые ошибки в формулировках основных понятий дисциплины	Не зачтено/ 0-59,99	Не освоена (недостаточный)
Уметь формулировать задачи в сфере анализа больших данных	Выполнение заданий на практических занятиях	Применение знаний по обработке и хранению больших данных к решению заданий на практических занятиях	Обучающийся твердо знает материал, грамотно и по существу излагает его, но допускает в ответе некоторые неточности	Зачтено/ 60-100	Освоена (повышенный)
			Обучающийся не раскрыл содержание материала, допускает грубые ошибки в формулировках основных понятий дисциплины	Не зачтено/ 0-59,99	Не освоена (недостаточный)
Владеть навыками использования современных информационных технологий для решения прикладных задач	Домашнее задание	Демонстрация навыков владения технологиями хранения и обработки больших данных	Обучающийся твердо знает материал, грамотно и по существу излагает его, но допускает в ответе некоторые неточности	Зачтено/ 60-100	Освоена (повышенный)
			Обучающийся не раскрыл содержание материала, допускает грубые ошибки в формулировках основных понятий дисциплины	Не зачтено/ 0-59,99	Не освоена (недостаточный)
УК-6 Способен управлять своим временем, выстраивать и реализовывать траекторию саморазвития на основе принципов образования в течение всей жизни					
Знать основные тенденции развития	Собеседование (зачет)	Корректно формирует требования к информационным	Обучающийся твердо знает материал, грамотно и по существу излагает его, но допускает в ответе некоторые неточности	Зачтено/ 60-100	Освоена (базовый)

цифровых технологий в области анализа больших объемов данных		системам	Обучающийся не раскрыл содержание материала, допускает грубые ошибки в формулировках основных понятий дисциплины	Не зачтено/ 0-59,99	Не освоена (недостаточный)
Уметь использовать результаты анализа данных для принятия эффективных решений при согласовании требований к ИС с заказчиком	Выполнение заданий на практических занятиях	Применение знаний об основных тенденциях развития цифровых технологиях к решению заданий на практических занятиях	Обучающийся твердо знает материал, грамотно и по существу излагает его, но допускает в ответе некоторые неточности	Зачтено/ 60-100	Освоена (повышенный)
			Обучающийся не раскрыл содержание материала, допускает грубые ошибки в формулировках основных понятий дисциплины	Не зачтено/ 0-59,99	Не освоена (недостаточный)
Владеть навыками использования современных методов анализа больших данных	Домашнее задание	Демонстрация навыков использования современных методов анализа больших данных	Обучающийся твердо знает материал, грамотно и по существу излагает его, но допускает в ответе некоторые неточности	Зачтено/ 60-100	Освоена (повышенный)
			Обучающийся не раскрыл содержание материала, допускает грубые ошибки в формулировках основных понятий дисциплины	Не зачтено/ 0-59,99	Не освоена (недостаточный)
ОПК-2 Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности					
Знать специфику анализа больших данных	Собеседование (зачет)	Корректно применяет знания о специфике анализа больших данных	Обучающийся твердо знает материал, грамотно и по существу излагает его, но допускает в ответе некоторые неточности	Зачтено/ 60-100	Освоена (повышенный)
			Обучающийся не раскрыл содержание материала, допускает грубые ошибки в формулировках основных понятий дисциплины	Не зачтено/ 0-59,99	Не освоена (недостаточный)
Уметь формулировать задачи в сфере анализа больших данных	Выполнение заданий на практических занятиях	Корректно формулирует задачи анализа больших данных при выполнении заданий на практических занятиях	Обучающийся твердо знает материал, грамотно и по существу излагает его, но допускает в ответе некоторые неточности	Зачтено/ 60-100	Освоена (повышенный)
			Обучающийся не раскрыл содержание материала, допускает грубые ошибки в формулировках основных понятий дисциплины	Не зачтено/ 0-59,99	Не освоена (недостаточный)

<p>Владеть инструментальными средствами обработки статистической информации, а также навыками применения вероятностно - статистического подхода к постановке и решению прикладных задач</p>	<p>Домашнее задание</p>	<p>Демонстрация навыков владения инструментальными средствами обработки статистической информации при решении прикладных задач</p>	<p>Обучающийся твердо знает материал, грамотно и по существу излагает его, но допускает в ответе некоторые неточности</p>	<p>Зачтено/ 60-100</p>	<p>Освоена (повышенный)</p>
			<p>Обучающийся не раскрыл содержание материала, допускает грубые ошибки в формулировках основных понятий дисциплины</p>	<p>Не зачтено/ 0-59,99</p>	<p>Не освоена (недостаточный)</p>