

МИНОБРНАУКИ РОССИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ  
ВЫСШЕГО ОБРАЗОВАНИЯ  
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
(ФГБОУ ВО «ВГУ»)

УТВЕРЖДАЮ  
Заведующий кафедрой  
Математических методов исследования операций  
Азарнова Т.В.  
22.03.2024 г



**РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ**  
**Б1.В.ДВ.02.02 Технологии Big Data в управлении бизнесом**

- 1. Код и наименование направления подготовки / специальности:**  
**38.03.05 Бизнес-информатика**
- 2. Профиль подготовки / специализация/магистерская программа:**  
Бизнес-аналитика и системы автоматизации предприятий
- 3. Квалификация (степень) выпускника:** бакалавр
- 4. Форма обучения:** очная
- 5. Кафедра, отвечающая за реализацию дисциплины:** математических методов исследования операций
- 6. Составители программы:** Ухлова В.В., к.ф.-м.н, доцент кафедры математических методов исследования операций
- 7. Рекомендована:** НМС факультета Прикладной математики, информатики и механики № 5 от 22.03.2024
- 8. Учебный год:** 2027/2028 **Семестр(ы):** 7

## 9. Цели и задачи учебной дисциплины

Цели дисциплины: освоение основных технологий, методов и алгоритмов работы с большими массивами данных, которые позволяют обрабатывать, анализировать, интерпретировать, оформлять и представлять профессиональному обществу результаты исследований, позволяют разрабатывать профессионально-ориентированные информационные системы с учетом возможностей современных интеллектуальных информационных технологий.

Задачами курса являются:

- знакомство с основными процессами консолидации, анализа, обработки и управления больших данных;
- изучение и совершенствование методов, алгоритмов и инструментальных средств работы с большими данными для возможности проведения аналитических исследований в рамках профессиональной деятельности;
- освоение основных навыков ведения проектов в области больших данных, в том числе, по созданию и внедрению профессионально-ориентированных информационных систем с учетом возможностей современных интеллектуальных информационных технологий.

## 10. Место учебной дисциплины в структуре ОПОП:

дисциплина относится к части, формируемой участниками образовательных отношений, дисциплинам по выбору блока Б1.В.ДВ.02 дисциплин учебного плана.

## 11. Планируемые результаты обучения по дисциплине/модулю (знания, умения, навыки), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями) и индикаторами их достижения

Код	Название компетенции	Код(ы)	Индикаторы(ы)	Планируемые результаты обучения
ПК-1	Способен использовать методы математического и статистического анализа, экономико-математические методы для решения задач в области бизнес-аналитики	ПК-1.3	Использует математический аппарат и инструментальные средства для обработки, анализа и систематизации данных для целей бизнес-анализа	Знать: основные технологии консолидации, обработки и управления большими данными, позволяющие осуществлять поиск, сбор и хранение информации из открытых источников и специализированных баз данных; основные методологии анализа данных; алгоритмы обработки данных.
ПК-3	Способен решать стандартные задачи профессиональной деятельности с учетом основных требований информационной безопасности	ПК-3.2	Анализирует возможность применения существующих ИС и ИКТ для управления предприятием	основные методики исследования и испытания разработанных методов, моделей, алгоритмов, технологий и инструментальных средств по работе с данными. Уметь: осуществлять информационный поиск с использованием открытых источников информации и специализированных баз данных;
ПК-6	Способен осуществлять стратегическое планирование с учетом анализа инноваций в экономике,	ПК-6.1	Оценивает эффективность ИТ-инноваций	использовать инструментальные средства для работы с данными, в том

	управлении и информационно-коммуникативных технологиях			числе, с большими данными; проводить исследования и испытания методов, моделей, алгоритмов и инструментальных средств работы с большими данными. Владеть навыками инсталляции и настройки ПО для работы с большими данными.
--	--	--	--	---

## 12. Объем дисциплины в зачетных единицах/часах в соответствии с учебным планом —2/72

Форма промежуточной аттестации - зачет.

## 13. Трудоемкость по видам учебной работы

Вид учебной работы	Трудоемкость (часы)				
	Всего	В том числе в интерактивной форме	По семестрам/ сессиям		
			Семестр № 7		.....
Аудиторные занятия					
в том числе: лекции	16		16		
практические					
лабораторные	16		16		
Самостоятельная работа	40		40		
Форма промежуточной аттестации	-		-		
Итого:	72		72		

### 13.1. Содержание дисциплины

№ п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК
<b>1. Лекции</b>			
1.1	Понятие Data science и Big Data, область применения, тенденции развития	1.1.1 Термины и определения. Особенности технологий. Сферы применения, состояние и тенденции развития. 1.1.2 Бизнес-кейсы Big Data. 1.1.3 Рынок Big Data в России и мире.	Анализ больших данных (38.03.05, Ухлова В.В.)
1.2	Технологии консолидации, обработки и управления большими данными	1.2.1 Платформа Hadoop: архитектура и принцип работы. Организация файловой системы HDFS. Концепция Map Reduce. Система YARN. Экосистема Hadoop.	
		1.2.2 Платформа Spark: архитектура и принцип работы. Файловые системы для работы в Spark.	
		1.2.3 Сравнение Hadoop и Spark: инфраструктура, работа ПО, задачи.	
		1.2.4 Базы данных NoSQL	
1.3	Основные процессы в Data science	1.3.1. Жизненный цикл аналитики больших данных: процессы сбора, подготовки, исследования и отображения данных. 1.3.2 Методы моделирования данных. 1.3.3 Визуализация данных.	

2. Лабораторные работы			
2.1	Методы работы с данными	2.1.1 Загрузка данных. Проверка качества данных. Очистка данных. Отображение данных. 2.1.2 Организация хранения данных. 2.1.3 Методы обработки и анализа данных. 2.1.4 Инструменты управления данными. 2.1.5 Выбор и установка ПО для работы с большими данными.	Анализ больших данных (38.03.05, Ухлова В.В.)

### 13.2. Темы (разделы) дисциплины и виды занятий

№ п/п	Наименование темы (раздела) дисциплины	Виды занятий (часов)				Всего
		Лекции	Практические	Лабораторные	Самостоятельная работа	
1	Понятие Data science и Big Data, область применения	4			4	8
2	Основные процессы в Data science	8			4	12
3	Технологии консолидации, обработки и управления большими данными	4		8	20	32
4	Методы работы с данными			8	12	20
Контроль						-
Итого:		16		16	40	72

### 14. Методические указания для обучающихся по освоению дисциплины

Дисциплина реализуется по тематическому принципу, каждая тема представляет собой завершённый раздел курса. На первом занятии студент получает информацию для доступа к комплексу учебно-методических материалов.

Лекционные занятия посвящены рассмотрению теоретических основ дисциплины: вводятся основные понятия, изучаются базовые технологии, разбираются основные процессы работы с большими данными. Лабораторные работы предназначены для формирования умений и навыков, закреплённых компетенций по ОПОП. Они организовываются в виде выполнения отдельных заданий. По окончании изучения дисциплины проводится тестирование.

Самостоятельная работа студентов включает в себя проработку учебного материала лекций, разбор заданий лабораторных работ, подготовку к экзамену. Для успешного освоения дисциплины рекомендуется подробно конспектировать лекционный материал, просматривать презентации по соответствующей теме, чтобы систематизировать изучаемый материал, выполнять задания лабораторных работ.

Промежуточная аттестация по результатам обучения проводится в форме экзамена, контролирующего освоение ключевых положений дисциплины, составляющих основу знаний по дисциплине.

При использовании дистанционных образовательных технологий и электронного обучения следует выполнять все указания преподавателя по работе на LMS-платформе, своевременно подключаться к online-занятиям, соблюдать рекомендации по организации самостоятельной работы.

## 15. Перечень основной и дополнительной литературы, ресурсов интернет, необходимых для освоения дисциплины

а) основная литература:

№ п/п	Источник
1	Основы технологий Big Data [Электронный ресурс] : учебное пособие / Воронеж. гос. ун-т / В.В. Ухлова .— Электрон. текстовые дан. — Воронеж : Издательский дом ВГУ, 2020 .— Загл. с титула экрана .— Свободный доступ из интрасети ВГУ .— Текстовый файл .— <URL:http://www.lib.vsu.ru/ >.
2	Макшанов, А. В. Большие данные. Big Data / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — 2-е изд., стер. — Санкт-Петербург : Лань, 2022. — 188 с. — ISBN 978-5-8114-9690-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/book/198599">https://e.lanbook.com/book/198599</a> (дата обращения: 25.02.2024). — Режим доступа: для авториз. пользователей.
3	Макшанов, А. В. Современные технологии интеллектуального анализа данных : учебное пособие для спо / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — Санкт-Петербург : Лань, 2020. — 228 с. — ISBN 978-5-8114-5451-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/book/149343">https://e.lanbook.com/book/149343</a> (дата обращения: 25.02.2024). — Режим доступа: для авториз. пользователей.

б) дополнительная литература:

№ п/п	Источник
4	Макшанов, А. В. Системы поддержки принятия решений : учебное пособие для вузов / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — 2-е изд., стер. — Санкт-Петербург : Лань, 2021. — 108 с. — ISBN 978-5-8114-8489-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/book/176903">https://e.lanbook.com/book/176903</a> (дата обращения: 25.02.2024). — Режим доступа: для авториз. пользователей.
5	Юре, Л. Анализ больших наборов данных / Л. Юре, Р. Ананд, Д. У. Джеффри ; перевод с английского А. А. Слинкин. — Москва : ДМК Пресс, 2016. — 498 с. — ISBN 978-5-97060-190-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/book/93571">https://e.lanbook.com/book/93571</a> (дата обращения: 25.02.2024). — Режим доступа: для авториз. пользователей.

в) информационные электронно-образовательные ресурсы (официальные ресурсы интернет)\*:

№ п/п	Ресурс
6	Электронно-библиотечная система «Лань» - Режим доступа: <a href="https://e.lanbook.com">https://e.lanbook.com</a>
7	Электронный каталог Научной библиотеки Воронежского государственного университета. – Режим доступа: <a href="http://www.lib.vsu.ru">http://www.lib.vsu.ru</a> .
8	Технологии Big Data в управлении бизнесом (38.03.05, Ухлова В.В.)/ В.В. Ухлова. — Образовательный портал «Электронный университет ВГУ». — Режим доступа: <a href="https://edu.vsu.ru">https://edu.vsu.ru</a>

## **16. Перечень учебно-методического обеспечения для самостоятельной работы**

Самостоятельная работа обучающегося должна включать подготовку к тестированию, лабораторным занятиям и подготовку к промежуточной аттестации. При самостоятельной подготовке обучающийся пользуется конспектами лекций и литературой по тематике лекционного материала, заданий контрольных и лабораторных работ. Для обеспечения самостоятельной работы студентов в электронном курсе дисциплины на образовательном портале «Электронный университет ВГУ» сформирован учебно-методический комплекс, который включает в себя: программу курса, учебные пособия и справочные материалы, методические указания по выполнению лабораторных работ. Студенты получают доступ к данным материалам на первом занятии по дисциплине.

## **17. Образовательные технологии, используемые при реализации учебной дисциплины, включая дистанционные образовательные технологии (ДОТ), электронное обучение (ЭО), смешанное обучение):**

При реализации дисциплины используются следующие образовательные технологии: логическое построение дисциплины, обозначение теоретического и практического компонентов в учебном материале. Применяются разные типы лекций (вводная, обзорная, информационная, проблемная). Дисциплина реализуется с применением информационно-коммуникационных технологий.

Информационно-коммуникативные технологии для реализации учебной дисциплины:

- технологии синхронного и асинхронного взаимодействия студентов и преподавателя посредством служб (сервисов) по пересылке и получению электронных сообщений, в том числе, по сети Интернет;
- сервис электронной почты для оперативной связи преподавателя и студентов.

Дисциплина реализуется с применением электронного обучения и дистанционных образовательных технологий, для организации самостоятельной работы обучающихся используется онлайн-курс, размещенный на платформе Электронного университета ВГУ (LMS moodle), а также другие Интернет-ресурсы, приведенные в п.15в.

## **18. Материально-техническое обеспечение дисциплины:**

Лекционная аудитория должна быть оборудована учебной мебелью, компьютером, мультимедийным оборудованием (проектор, экран, средства звуковоспроизведения), допускается переносное оборудование.

Лабораторные работы должны проводиться в специализированной аудитории, оснащенной учебной мебелью и персональными компьютерами с доступом в сеть Интернет (компьютерные классы, студии), мультимедийным оборудованием (проектор, экран, средства звуковоспроизведения), Число рабочих мест в аудитории должно быть таким, чтобы обеспечивалась индивидуальная работа студента персональном компьютере.

Для самостоятельной работы необходимы компьютерные классы, помещения, оснащенные компьютерами с доступом к сети Интернет и платформе Электронного университета ВГУ (LMS moodle).

Программное обеспечение:

- ОС Windows 10, ОС Linux;

- пакет стандартных офисных приложений для работы с документами, таблицами и т.п. (МойОфис, LibreOffice);
- ПО Adobe Reader;
- специализированное ПО (ПО MatLab, Power BI);
- интернет-браузер (Mozilla Firefox, Яндекс).

## 19. Оценочные средства для проведения текущей и промежуточной аттестаций

№ п/п	Наименования раздела дисциплины	Компетенция(и)	Индикатор(ы) достижения компетенции	Оценочные средства
1	Понятие Data science и Big Data, область применения.	ПК-3 ПК-6	ПК-3.2 ПК-6.1	Контрольная работа
2	Технологии консолидации, обработки и управления большими данными.	ПК-1	ПК-1.3	Лабораторная работа 1-5
3	Основные процессы в Data science.	ПК-1 ПК-6	ПК-1.3 ПК-6.1	Лабораторная работа 1-5
4	Методы работы с данными	ПК-1	ПК-1.3	Лабораторная работа 1-5
Промежуточная аттестация, форма контроля - зачет				Тест

## 20 Типовые оценочные средства и методические материалы, определяющие процедуры оценивания

### 20.1 Текущий контроль успеваемости

Контроль успеваемости по дисциплине осуществляется с помощью следующих оценочных средств:

- контрольные работы,
- тест,
- лабораторные работы.

### Перечень заданий контрольной работы

Для исходного набора данных:

- 1) выполнить описание «идеальных» данных (тип данных, ограничения, шаблон и т.п.);
- 2) привести варианты возможных ошибок в данных;
- 3) составить алгоритм повышения качества данных;
- 4) продемонстрировать траекторию изменения данных при использовании разработанного алгоритма;
- 5) составить рекомендации, позволяющие получать исходный набор данных с более высоким качеством.

### Технология проведения

В качестве исходных данных студент берет любой набор из открытых источников (в формате xls/xlsx (количество записей должно быть более 50, атрибутов более 10). Если качество данных набора очень высокое, то искусственно «ухудшает» его.

Выполнение задания предусматривает использование информации из учебной и справочной литературы, а также ресурсов сети Интернет.

Критерии оценивания:

- оценка «зачтено» выставляется студенту, если студент дал правильные ответы на 90 и более процентов заданий;
- оценка «не зачтено» - даны правильные ответы на менее чем 90 процентов заданий.

### Перечень заданий теста

Пример компоновки вопросов теста (вопросы с вариантами ответов).

Вариант 1.

1. Приведите основные характеристики больших данных:

- a) Virtualization, Volume, Variability, Vehicle;
- б) Variety, Velocity, Volume, Value;
- в) Verification, Volume, Velocity, Visualization;
- г) Video, Value, Variety, Volume.

2. Расставьте в правильном порядке основные этапы процесса Data Science:

- a) назначение цели исследования, сбор данных, подготовка данных, исследование данных, моделирование данных, отображение данных;
- б) назначение цели исследования, сбор данных, подготовка данных, моделирование данных, исследование данных, отображение данных;
- в) назначение цели исследования, подготовка данных, сбор данных, моделирование данных, исследование данных, отображение данных;
- г) назначение цели исследования, сбор данных, подготовка данных, отображение данных, исследование данных, моделирование данных.

3. Поясните понятие:

Nadoop представляет собой...

- a) набор утилит, и программный каркас для выполнения распределённых программ, работающих на кластерах;
- б) распределённую СУБД, позволяющую обрабатывать большие данные;
- в) язык выполнения заданий в парадигме MapReduce;
- г) распределённую файловую систему для организации хранения файлов большого объёма.

4. Принцип MapReduce состоит в том, чтобы

- a) производить вычисления на узлах, где информация изначально была сохранена;
- б) использовать вычислительные мощности систем хранения;
- в) использовать функциональное программирование для решения задач массивно-параллельной обработки.

Технология проведения

Тест включает в себя 30 вопросов, вариант теста выбирается исходя из номера зачетки (последней цифры). Время на тестирование рассчитывается из соотношения 10 вопросов – 15 минут. Результаты тесты проверяются по ключу правильных ответов.



**Критерии оценивания:**

- оценка «зачтено» выставляется студенту, если студент дал правильные ответы на 90 и более процентов заданий (тест пройден);
- оценка «не зачтено» - даны правильные ответы на менее чем 90 процентов заданий (тест не пройден).

**Перечень заданий для лабораторных работ.****Лабораторная работа №1****Пример задания.**

Выполнить загрузку данных в аналитический контур. В качестве исходных данных использовать форматы xls, xlsx, txt, pdf. Проверить факт загрузки с использованием инструментов отображения данных.

**Технология проведения**

Студент выбирает вариант задания, ориентируясь на номер зачетки (последняя цифра). Файлы исходных данных заранее должны быть размещены на сервере (компьютере студента). Студенту разрешается пользоваться информацией из открытых источников.

**Критерии оценивания:**

- оценка «зачтено» выставляется студенту, если задания выполнены в полном объеме;
- оценка «не зачтено» - работа не выполнена или выполнена не в полном объеме.

**Лабораторная работа №2****Пример задания.**

Вычислить основные статистики данных, загруженных в аналитический контур. В качестве исходных данных использовать форматы xls, xlsx, txt, pdf. Отобразить полученные статистики с использованием соответствующих инструментов.

**Технология проведения**

Студент выбирает вариант задания, ориентируясь на номер зачетки (последняя цифра). Файлы исходных данных заранее должны быть размещены на сервере (компьютере студента). Студенту разрешается пользоваться информацией из открытых источников.

**Критерии оценивания:**

- оценка «зачтено» выставляется студенту, если задания выполнены в полном объеме;
- оценка «не зачтено» - работа не выполнена или выполнена не в полном объеме.

**Лабораторная работа №3****Пример задания.**

Выполнить проверку данных, загруженных в аналитический контур. В качестве исходных данных использовать форматы xls, xlsx, txt, pdf. Повысить качество данных с использованием соответствующих инструментов.

**Технология проведения**

Студент выбирает вариант задания, ориентируясь на номер зачетки (последняя цифра). Файлы исходных данных заранее должны быть размещены на сервере (компьютере студента). Студенту разрешается пользоваться информацией из открытых источников.

**Критерии оценивания:**

- оценка «зачтено» выставляется студенту, если задания выполнены в полном объеме;
- оценка «не зачтено» - работа не выполнена или выполнена не в полном объеме.

## Лабораторная работа №4

## Пример задания.

Рассчитать метрики для данных, загруженных в аналитический контур. В качестве исходных данных использовать форматы xlsx, txt, pdf. Отобразить полученные метрики с использованием соответствующих инструментов.

## Технология проведения

Студент выбирает вариант задания, ориентируясь на номер зачетки (последняя цифра). Файлы исходных данных заранее должны быть размещены на сервере (компьютере студента). Студенту разрешается пользоваться информацией из открытых источников.

## Критерии оценивания:

- оценка «зачтено» выставляется студенту, если задания выполнены в полном объеме;
- оценка «не зачтено» - работа не выполнена или выполнена не в полном объеме.

## Лабораторная работа №5

## Пример задания.

Выполнить визуализацию обработанных данных, отобразить рассчитанные в предыдущих работах статистики и метрики. В качестве исходных данных использовать форматы xlsx, txt, pdf.

## Технология проведения

Студент выбирает вариант задания, ориентируясь на номер зачетки (последняя цифра). Файлы исходных данных заранее должны быть размещены на сервере (компьютере студента). Студенту разрешается пользоваться информацией из открытых источников.

## Критерии оценивания:

- оценка «зачтено» выставляется студенту, если задания выполнены в полном объеме;
- оценка «не зачтено» - работа не выполнена или выполнена не в полном объеме.

## Лабораторная работа №6

## Пример задания.

Выполнить расчет хранилища данных для системы офисной системы видеонаблюдения. Параметры системы видеонаблюдения: 5 камер, разрешение 2.1, 1920x1080, частота 12к/с, кодек H.264. Период хранения данных составляет 3 месяца,

## Технология проведения

Студент выбирает вариант задания, ориентируясь на номер зачетки (последняя цифра). Время выполнения задания составляет 3 часа. Студенту разрешается пользоваться информацией из открытых источников.

## Критерии оценивания:

- оценка «отлично» выставляется студенту, если работа выполнена в полном объеме (приведены все расчеты и они правильные, даны пояснения);
- оценка «хорошо» - работа выполнена полностью, но имеются незначительные ошибки;
- оценка «удовлетворительно» - работа выполнена полностью, но в представленной части много ошибок или представлена часть работы и она без ошибок;
- оценка «неудовлетворительно» - работа не выполнена.

## Лабораторная работа №7

## Пример задания.

1. Обозначить бизнес-проблему.
2. Сформулировать бизнес-цели.
3. Обозначить бизнес-задачи.

4. Свести бизнес-задачу к аналитической задаче.
5. Определить потребности в ресурсах (указать источники, виды ресурсов, виды и содержание информации, которую можно получить).
6. Подобрать технологии (методы, модели, алгоритмы, инструментальные средства), позволяющие работать с определенными в п.6 ресурсами.
7. При необходимости дать рекомендации по доработке технологии (методы, модели, алгоритмы, инструментальные средства) из п.6.

#### Технология проведения

Предметную область студент выбирает самостоятельно, базируясь на информации из открытых источников. Время выполнения задания составляет 3 часа. Студенту разрешается пользоваться информацией из открытых источников.

#### Критерии оценивания:

- оценка «отлично» выставляется студенту, если работа выполнена в полном объеме, полученные результаты аргументированы;
- оценка «хорошо» - работа выполнена полностью, но полученные результаты не логичны или требуют уточнения;
- оценка «удовлетворительно» - работа выполнена полностью, но имеет место большое количество ошибок или представлена часть работы и она без ошибок;
- оценка «неудовлетворительно» - работа не выполнена.

Лабораторные работы №6-7 могут заменяться на подготовку реферата (презентацию) по одному из разделов дисциплины (одна работа равна одному реферату).

#### Примерные темы рефератов (презентаций)

Обзор инструментов работы с данными  
 Пример подбора инструментов обработки данных  
 Пример подбора инструментов очистки данных  
 Пример подбора инструментов отображения данных  
 Пример расчета статистик загруженных данных в ПО  
 Пример расчета метрик загруженных данных в ПО  
 Разработка алгоритма анализа больших данных в рамках поставленной задачи

#### Технология проведения

Тема выбирается обучающимся самостоятельно. При этом рекомендуется выбор тем в группе таким образом, чтобы они не повторялись.

#### Критерии оценки:

- оценка «зачтено» выставляется студенту, если:
  - изложенная информация является актуальной на момент представления реферата;
  - по содержанию реферат отражает все основные аспекты выбранной темы;
  - реферат оформлен в соответствии с рекомендациями по оформлению;
- оценка «не зачтено», если:
  - изложенная информация не является актуальной на момент представления реферата;
  - по содержанию реферат не отражает все основные аспекты выбранной темы;
  - реферат не оформлен в соответствии с рекомендациями по оформлению.

## 20.2 Промежуточная аттестация

Промежуточная аттестация по дисциплине осуществляется с помощью следующих оценочных средств: тест.

#### Тестовые задания.

Пример компоновки вопросов теста (вопросы с вариантами ответов).

## Вариант 1.

1. Приведите основные характеристики больших данных:

- а) Virtualization, Volume, Variability, Vehicle;
- б) Variety, Velocity, Volume, Value;
- в) Verification, Volume, Velocity, Visualization;
- г) Video, Value, Variety, Volume.

2. Расставьте в правильном порядке основные этапы процесса Data Science:

- а) назначение цели исследования, сбор данных, подготовка данных, исследование данных, моделирование данных, отображение данных;
- б) назначение цели исследования, сбор данных, подготовка данных, моделирование данных, исследование данных, отображение данных;
- в) назначение цели исследования, подготовка данных, сбор данных, моделирование данных, исследование данных, отображение данных;
- г) назначение цели исследования, сбор данных, подготовка данных, отображение данных, исследование данных, моделирование данных.

3. Поясните понятие: Hadoop представляет собой...

- а) набор утилит, и программный каркас для выполнения распределённых программ, работающих на кластерах;
- б) распределённую СУБД, позволяющую обрабатывать большие данные;
- в) язык выполнения заданий в парадигме MapReduce;
- г) распределённую файловую систему для организации хранения файлов большого объёма.

4. Принцип MapReduce состоит в том, чтобы

- а) производить вычисления на узлах, где информация изначально была сохранена;
- б) использовать вычислительные мощности систем хранения;
- в) использовать функциональное программирование для решения задач массивно-параллельной обработки.

Технология проведения: тест состоит из 50 вопросов. Вариант теста выбирается, исходя из номера зачетки (последней цифры). Время тестирования составляет 45 минут. Результаты теста проверяются по ключу правильных ответов.

Критерии оценивания:

- оценка «зачтено» выставляется студенту, если студент дал правильные ответы на 75 и более процентов заданий (тест пройден);
- оценка «не зачтено» - даны правильные ответы на менее чем 75 процентов заданий (тест не пройден).

Для оценивания результатов обучения на зачет используется шкала: «зачтено», «не зачтено».

Соотношение показателей, критериев и шкалы оценивания результатов обучения:

Критерии оценивания компетенций	Уровень сформированности компетенций	Шкала оценок
Итоговый тест зачтен. Посещение лекций базового блока составляет 90%. Выполнены лабораторные работы №1-5 или выполнены 3 из 5 лабораторных работ и контрольная работа зачтена.	<i>Базовый уровень</i>	Зачтено
Итоговый тест не зачтен и/или посещение лекций базового блока составляет менее 90% и/или не выполнены лабораторные работы №1-5.	<i>Ниже базового уровня</i>	Не зачтено