

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.
18.04.2025 г



РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ
Б1.В.ДВ.02.01 Анализ больших данных

1. Код и наименование направления подготовки:

38.03.05 Бизнес-информатика

2. Профиль подготовки:

Бизнес-аналитика и системы автоматизации предприятий

3. Квалификация выпускника: бакалавр

4. Форма обучения: очная

5. Кафедра, отвечающая за реализацию дисциплины: ММИО

6. Составители программы: Ухлова В.В., к.ф.-м.н, доцент кафедры ММИО

7. Рекомендована: НМС факультета Прикладной математики, информатики и механики № 6 от 17.03.2025

8. Учебный год: 2028/2029

Семестр(ы): 7

9. Цели и задачи учебной дисциплины

Целями освоения учебной дисциплины являются: формирование целостного представления о современных технологиях работы с данными, методам и алгоритмам работы с большими массивами данных, которые позволяют обрабатывать, интерпретировать, оформлять и представлять профессиональному обществу результаты исследований.

Задачи учебной дисциплины:

- изучение процессов консолидации, анализа, обработки больших данных;
- получение знаний и умений, необходимых для проведения анализа предметной области, выявления информационных потребностей организации;
- получение знаний и умений, необходимых для проведения аналитического исследования в соответствии с согласованными требованиями;
- приобретение навыков формализации требований к данным, формализации и построения концептуальной модели решения профессиональных задач;
- приобретение навыков разработки требований и выбора инструментальных средств и технологий проектирования прикладных моделей и программ.

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, управлении и информационно-	ПК-6.1	Оценивает эффективность ИТ-инноваций	использовать инструментальные средства для работы с данными, в том числе, с большими данными; проводить исследования и испытания методов, моделей,

	коммуникативных технологиях			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
1. Лекции			
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 (ПММ, Ухлова В.В.) https://edu.vsu.ru/course/view.php?id=19266
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 (ПММ, Ухлова В.В.) https://edu.vsu.ru/course/view.php?id=19266

13.2. Темы (разделы) дисциплины и виды занятий

№ п/п	Наименование темы (раздела) дисциплины	Виды занятий (часов)				
		Лекции	Практические	Лабораторные	Самостоятельная работа	Всего
1	Понятие Data science и Big Data, область применения	4	-	-	6	10
2	Основные процессы в Data science	8	-	-	10	18
3	Технологии консолидации, обработки и управления большими данными	2	-	4	4	10
4	Методы работы с данными	2	-	12	20	34
	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 зачёта, 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: https://e.lanbook.com/book/198599 (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: https://e.lanbook.com/book/149343 (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: https://e.lanbook.com/book/176903 (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: https://e.lanbook.com/book/93571 (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: https://e.lanbook.com
7	Электронный каталог Научной библиотеки Воронежского государственного университета. – Режим доступа: http://www.lib.vsu.ru .
8	Анализ больших данных (ПММ, Ухлова В.В.)/ В.В. Ухлова. — Образовательный портал «Электронный университет ВГУ». — Режим доступа: https://edu.vsu.ru/course/view.php?id=19266

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:

Учебная аудитория для проведения лекций: специализированная мебель, компьютер (ноутбук), мультимедийное оборудование (проектор, экран, средства звуковоспроизведения), допускается использование переносного оборудования.

Программное обеспечение:

ОС Windows 8 (10), интернет-браузер (Google Chrome, Mozilla Firefox), с возможностью подключения к сети «Интернет» и платформе Электронного университета ВГУ (LMS moodle), ПО Adobe Reader, пакет стандартных офисных приложений для работы с документами (MS Office, Мой Офис, Libre Office).

Учебная аудитория для проведения практических занятий, лабораторных работ, организации самостоятельной работы, проведения текущих и промежуточных аттестаций: специализированная мебель, компьютер (ноутбук) для преподавателя, персональные компьютеры для возможности организации индивидуальной работы обучающихся, мультимедийное оборудование (проектор, экран, средства звуковоспроизведения), допускается использование переносного оборудования.

ОС Windows 8 (10), ОС Linux, интернет-браузер (Google Chrome, Mozilla Firefox), с возможностью подключения к сети «Интернет» и платформе Электронного университета ВГУ (LMS moodle), ПО Adobe Reader, пакет стандартных офисных приложений для работы с документами (MS Office, Мой Офис, Libre Office), специализированное ПО по тематике дисциплины (Power BI) (допускается демоверсия или виртуальный аналог ПО).

19. Оценочные средства для проведения текущей и промежуточной аттестаций

Контроль успеваемости по дисциплине осуществляется с помощью следующих оценочных средств:

- перечень вопросов (тест),
- лабораторные работы.

Порядок оценки освоения обучающимися учебного материала определяется содержанием следующих разделов дисциплины:

№ п/п	Наименования раздела дисциплины	Компетенция(и)	Индикатор(ы) достижения компетенции	Оценочные средства
1	Понятие Data science и Big Data, область применения.	ПК-4	ПК-4.1	Лабораторные работы 1-5
2	Технологии консолидации, обработки и управления большими данными.	ПК-6	ПК-6.1, ПК-6.2, ПК-6.3	Лабораторные работы 1-5
3	Основные процессы в Data science.	ПК-6	ПК-6.1, ПК-6.2, ПК-6.3	Лабораторные работы 1-5
4	Методы работы с данными	ПК-6	ПК-6.1, ПК-6.2, ПК-6.3	Лабораторные работы 1-5
Промежуточная аттестация, форма контроля – зачет				Перечень вопросов (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:

- перечень вопросов (тест),
- лабораторные работы.

Перечень заданий теста

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 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, 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 `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 `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.

Лабораторные работы 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:

- 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:

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 зачете используется «зачтено» и «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