


Министерство образования и науки Российской Федерации Ульяновский государственный университет	Форма	
Ф - Рабочая программа по дисциплине		

## УТВЕРЖДЕНО

решением Ученого совета факультета математики,  
информационных и авиационных технологий  
от «16» мая 2023 г. протокол 4/23  
Председатель М.А. Волков



## РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Дисциплина	Технологии хранения и обработки больших объемов данных
Факультет	Математики, информационных и авиационных технологий
Кафедра	Информационных технологий
Курс	1

Направление: **01.04.02 «Прикладная математика и информатика»**  
*код направления (специальности), полное наименование*

Профиль: **«Имитационное моделирование и анализ данных»**  
*полное наименование*

Форма обучения: очная

Дата введения в учебный процесс УлГУ: «01» 09 2023 г.



Программа актуализирована на заседании кафедры: протокол № \_\_\_ от \_\_\_ 20\_\_\_ г.


Программа актуализирована на заседании кафедры: протокол № \_\_\_ от \_\_\_ 20\_\_\_ г.

Программа актуализирована на заседании кафедры: протокол № \_\_\_ от \_\_\_ 20\_\_\_ г.

Сведения о разработчиках:

ФИО	Кафедра	Должность, ученая степень, звание
Шабалин Александр Станиславович	Информационных технологий	доцент, к.ф.-м.н.

СОГЛАСОВАНО	СОГЛАСОВАНО
Заведующий кафедрой информационных технологий, реализующей дисциплину	Заведующий выпускающей кафедрой прикладной математики
 _____ Подпись / Волков М.А. ФИО «16» мая 2023 г.	 _____ Подпись / Бутов А.А. ФИО «16» мая 2023 г.

Министерство образования и науки Российской Федерации Ульяновский государственный университет	Форма	
Ф - Рабочая программа по дисциплине		

## 1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины «Технологии хранения и обработки больших объемов данных» является теоретическая и практическая подготовка студентов к работе с большими данными. Знания, полученные в результате освоения дисциплины, помогут при сборе и анализе огромных объемов структурированной или неструктурированной информации, при разработке моделей данных и получении новых знаний. Все это необходимо выпускнику, освоившему программу магистратуры, для решения различных задач практической и научно-исследовательской деятельности.

Задачи освоения дисциплины:

- приобретение студентами знаний о технологиях подготовки, хранения, обработки и анализа больших данных;
- применение статистических и математических методов для анализа больших объемов информации;
- приобретение практических навыков работы с методами Map Reduce.


## 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина «Технологии хранения и обработки больших объемов данных» относится к базовой части Блока Б1.О.04 «Дисциплины (модули)» Основной Образовательной Программы по направлению подготовки магистров 01.04.02 Прикладная математика информатика.

## 3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ), СООТНЕСЕННЫХ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Процесс изучения дисциплины направлен на формирование следующих компетенций:

Код и наименование реализуемой компетенции	Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с индикаторами достижения компетенций
способность комбинировать и адаптировать существующие информационно-коммуникационные технологии для решения задач в области профессиональной деятельности с учетом требований информационной безопасности (ОПК-4)	<b>Знать:</b> причины возникновения тренда больших данных; процессы анализа больших данных; основные подходы к обработке больших массивов данных; <b>Уметь:</b> формулировать алгоритмы; выбирать подходящий инструмент анализа больших данных; выбирать подходящую технологию хранения больших данных. <b>Владеть:</b> Современными инструментами работы с большими данными.
способность разрабатывать требования к программным	<b>Знать:</b> причины возникновения тренда больших данных; процессы анализа больших данных; основные

Министерство образования и науки Российской Федерации Ульяновский государственный университет	Форма	
Ф - Рабочая программа по дисциплине		

продуктам математическому обеспечению, приложениям, системам, информационной инфраструктуре ПК-7	и	подходы к обработке больших массивов данных; <b>Уметь:</b> формулировать алгоритмы; выбирать подходящий инструмент анализа больших данных; выбирать подходящую технологию хранения больших данных. <b>Владеть:</b> Современными инструментами работы с большими данными.
--	---	--


#### 4. ОБЪЕМ ДИСЦИПЛИНЫ

4.1. Объем дисциплины в зачетных единицах (всего) – 2 ЗЕ.

4.2. По видам учебной работы (в часах):

Вид учебной работы	Количество часов (форма обучения очная)			
	Всего по плану	В т.ч. по семестрам		
		1	4	5
1	2	3	4	5
Контактная работа обучающихся с преподавателем в соответствии с УП	18	18/18		
Аудиторные занятия:	18	18		
Лекции	18	18/18		
практические и семинарские занятия	Нет	Нет		
лабораторные работы (лабораторный практикум)	нет	нет		
Самостоятельная работа	54	54		
Текущий контроль (количество и вид: конт. работа, коллоквиум, реферат)	нет	Нет		
Курсовая работа	нет	Нет		
Виды промежуточной аттестации (экзамен, зачет)	Зачёт	Зачёт		
Всего часов по дисциплине	72	72		


*\*В случае необходимости использования в учебном процессе частично/исключительно дистанционных образовательных технологий в таблице через слеш указывается количество часов работы ППС с обучающимися для проведения занятий в дистанционном формате с применением электронного обучения*

Министерство образования и науки Российской Федерации Ульяновский государственный университет	Форма	
Ф - Рабочая программа по дисциплине		

### 4.3. Содержание дисциплины (модуля). Распределение часов по темам и видам учебной работы:

Форма обучения – очная

		Аудиторные занятия			Самостоятельная работа
		лекции	практические занятия, семинары	лабораторная работа	
1	2	3	4	5	7
Тема 1. Вводный обзор: что такое Big Data и для чего нужен.	8	2			6
Тема 2. Обзор реляционных баз данных.	8	2			6
Тема 3. Предметно-ориентированные информационные базы данных.	8	2			6
Тема 4. MapReduce: методология и технология распределенных вычислений.	8	2			6
Тема 5. Введение в Hadoop.	8	2			6
Тема 6. Обработка данных в реальном времени.	8	2			6
Тема 7. Массово-параллельная структура - Massive Parallel Processing.	8	2			6
Тема 8. Вычисление дескриптивных статистик для больших объемов данных.	8	2			6
Тема 9. Data Mining и Big Data.	8	2			6
Зачет					

Министерство образования и науки Российской Федерации Ульяновский государственный университет	Форма	
Ф - Рабочая программа по дисциплине		

Итого	72	18		54
-------	----	----	--	----

## 5. СОДЕРЖАНИЕ КУРСА

**Тема 1.** Вводный обзор: что такое Big Data и для чего нужен.

с каких объемов начинается Big Data

реляционные и нереляционные базы данных

потоки данных

**Тема 2.** Обзор реляционных баз данных.

SQL-сервер: основные принципы, примеры

NoSQL базы данных: обзор, примеры

**Тема 3.** Предметно-ориентированные информационные базы данных.

Data Warehausig

**Тема 4.** MapReduce: методология и технология распределенных вычислений.

Этап Map – предварительной обработки

Этап Reduce – свертки результатов

Примеры функций

**Тема 5.** Введение в Hadoop.

основные принципы Hadoop

компоненты Hadoop

работа с нереляционными данными

примеры использования

MapReduce в Hadoop

настройки Hive и Pig

**Тема 6.** Обработка данных в реальном времени.

Storm,

Spark,

Impal

**Тема 7.** Массово-параллельная структура - Massive Parallel Processing

масштабирование реляционных баз данных

параллельное выполнение запросов к БД

архитектура Hub and Spoke

**Тема 8.** Вычисление дескриптивных статистик для больших объемов данных.

частоты,

средние,

стандартные отклонения,

медианы,

квартили

**Тема 9.** Data Mining и Big Data

кластеризация, сегментация, алгоритмы к-средних, EM - Expectation-maximization

иерархическая кластеризация


классификация данных

предиктивный анализ

регрессионные деревья

правила ассоциаций

machine learning.

Министерство образования и науки Российской Федерации Ульяновский государственный университет	Форма	
Ф - Рабочая программа по дисциплине		

## 6. ТЕМЫ ПРАКТИЧЕСКИХ И СЕМИНАРСКИХ ЗАНЯТИЙ

Не предусмотрены учебным планом

## 7. ЛАБОРАТОРНЫЕ РАБОТЫ (ЛАБОРАТОРНЫЙ ПРАКТИКУМ)

Не предусмотрены учебным планом

## 8. ТЕМАТИКА КУРСОВЫХ, КОНТРОЛЬНЫХ РАБОТ, РЕФЕРАТОВ


Курсовые и контрольные работы, рефераты не предусмотрены учебным планом.

## 9. ПЕРЕЧЕНЬ ВОПРОСОВ К ЗАЧЕТУ

1.	Технологии BigData: дать определение для BigData, назначение BigData, история появления и основные принципы BigData. Достоинства и недостатки BigData.
2.	Технологии управления знаниями, визуализации знаний и интеллектуальные карты. Дать определение понятиям, назначение технологии, привести примеры программного обеспечения для визуализации знаний и построения интеллектуальных карт.
3.	Данные, информация, знания, модели. Наука о данных.
4.	Эволюционное развитие архитектур и данных.
5.	Критерии больших данных. Источники больших данных. Интернет вещей. Робототехника.
6.	Возможные этапы работы с большими данными.
7.	Примеры и истории успеха работы с большими данными: торговля, финансы, кадры.
8.	Обзор подходов к работе с данными: от языка простых запросов до методов анализа больших данных.
9.	Интеллектуальный анализ данных: краткий обзор подходов.
10.	Генетические алгоритмы.
11.	Деревья принятия решений.
12.	Визуализация больших данных.
13.	Специфика хранения и обработки больших данных.
14.	Парадигма MapReduce
15.	Файловая система HDFS.
16.	Особенности хранилищ данных NoSQL.
17.	Архитектура высоконагруженных систем.

## 10. САМОСТОЯТЕЛЬНАЯ РАБОТА СТУДЕНТОВ

Название разделов и тем	Вид самостоятельной работы	Объем в часах	Форма контроля
Тема 1.	Проработка учебного материала, лабораторные работы, подготовка к сдаче зачета	6	Зачет, проверка лабораторных работ
Тема 2.	Проработка учебного материала, лабораторные работы, подготовка к сдаче зачета	6	Зачет, проверка лабораторных работ
Тема 3.	Проработка учебного материала,	6	Зачет, проверка

Министерство образования и науки Российской Федерации Ульяновский государственный университет	Форма	
Ф - Рабочая программа по дисциплине		

	лабораторные работы, подготовка к сдаче зачета		лабораторных работ
Тема 4.	Проработка учебного материала, лабораторные работы, подготовка к сдаче зачета	6	Зачет, проверка лабораторных работ
Тема 5.	Проработка учебного материала, лабораторные работы, подготовка к сдаче зачета	6	Зачет, проверка лабораторных работ
Тема 6.	Проработка учебного материала, лабораторные работы, подготовка к сдаче зачета	6	Зачет, проверка лабораторных работ
Тема 7.	Проработка учебного материала, лабораторные работы, подготовка к сдаче зачета	6	Зачет, проверка лабораторных работ
Тема 8.	Проработка учебного материала, лабораторные работы, подготовка к сдаче экзамена	6	Зачет, проверка лабораторных работ
Тема 9.	Проработка учебного материала, лабораторные работы, подготовка к сдаче зачета	6	Зачет, проверка лабораторных работ

По данной дисциплине организуется и проводится внеаудиторная самостоятельная работа.

Самостоятельная работа студентов, предусмотренная учебным планом в объеме не менее 50-70% общего количества часов, должна соответствовать более глубокому усвоению изучаемого курса, формировать навыки исследовательской работы и ориентировать студентов на умение применять теоретические знания на практике.

Самостоятельная работа по данной дисциплине состоит из следующих модулей:

- подготовка к лабораторным занятиям;
- подготовка к экзамену.

При подготовке к лабораторным занятиям и контрольным мероприятиям рекомендуется руководствоваться учебниками и учебными пособиями, в том числе и информацией, полученной в INTERNET.


Задания для самостоятельной работы требует дополнительной проработки и анализа рассматриваемого преподавателем материала в объеме запланированных часов.

Задания по самостоятельной работе оформлены в виде таблицы с указанием конкретного вида самостоятельной работы:

- проработка учебного материала (по конспектам лекций учебной и научной литературе) и подготовка лабораторным занятиям;
- поиск и обзор научных публикаций и электронных источников информации;

Студентам рекомендуется следующий порядок организации самостоятельной работы над темами и подготовки к практическим занятиям:

- ознакомиться с содержанием темы;
- прочитать материал лекций, при этом нужно составить себе общее представление об излагаемых вопросах;
- прочитать параграфы учебника, относящиеся к данной теме;
- перейти к тщательному изучению материала, усвоить теоретические положения и выводы, при этом нужно записывать основные положения темы (формулировки, определения, термины, воспроизводить отдельные схемы и чертежи из учебника и конспекта лекций);

Министерство образования и науки Российской Федерации Ульяновский государственный университет	Форма	
Ф - Рабочая программа по дисциплине		

РЕЗУЛЬТАТЫ САМОСТОЯТЕЛЬНОЙ РАБОТЫ КОНТРОЛИРУЮТСЯ ПРЕПОДАВАТЕЛЕМ И УЧИТЫВАЮТСЯ ПРИ АТТЕСТАЦИИ СТУДЕНТА (ЭКЗАМЕН).

## 11. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

### Список рекомендуемой литературы

Список рекомендуемой литературы

#### а) Основная литература

1. Жуковский, О. И. Информационные технологии и анализ данных [Электронный ресурс] : учебное пособие / О. И. Жуковский. — Электрон. текстовые данные. — Томск : Томский государственный университет систем управления и радиоэлектроники, Эль Контент, 2014. — 130 с. — 978-5-4332-0158-3. — Режим доступа: <http://www.iprbookshop.ru/72106.html>
2. Воронова, Л. И. Machine Learning: регрессионные методы интеллектуального анализа данных [Электронный ресурс] : учебное пособие / Л. И. Воронова, В. И. Воронов. — Электрон. текстовые данные. — М. : Московский технический университет связи и информатики, 2018. — 82 с. — 2227-8397. — Режим доступа: <http://www.iprbookshop.ru/81325.html>


#### Дополнительная литература:

1. Федин, Ф. О. Анализ данных. Часть 1. Подготовка данных к анализу [Электронный ресурс] : учебное пособие / Ф. О. Федин, Ф. Ф. Федин. — Электрон. текстовые данные. — М. : Московский городской педагогический университет, 2012. — 204 с. — 2227-8397. — Режим доступа: <http://www.iprbookshop.ru/26444.html>
2. Федин, Ф. О. Анализ данных. Часть 2. Инструменты Data Mining [Электронный ресурс] : учебное пособие / Ф. О. Федин, Ф. Ф. Федин. — Электрон. текстовые данные. — М. : Московский городской педагогический университет, 2012. — 308 с. — 2227-8397. — Режим доступа: <http://www.iprbookshop.ru/26445.html>
3. Лесковец, Ю. Анализ больших наборов данных / Лесковец Ю. , Раджараман А. , Джеффри Д. Ульман - Москва : ДМК Пресс, 2016. - 498 с. - ISBN 978-5-97060-190-7. - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785970601907.html>
4. Вольфсон, М. Б. Анализ данных : учебное пособие / М. Б. Вольфсон. — Санкт-Петербург : СПбГУТ им. М.А. Бонч-Бруевича, 2015. — 81 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/180254>

#### Учебно-методическая

1. Головин В.А. Методические указания для самостоятельной работы студентов по дисциплине «Технологии хранения и обработки больших объемов данных» / В.А. Головин. Ульяновск: УлГУ, 2019/ В. А. Головин; УлГУ, Фак. математики, информ. и авиац. технологий. - Ульяновск : УлГУ, 2019 - Загл. с экрана; Неопубликованный ресурс. - Электрон. текстовые дан. (1 файл : 752 КБ). - Текст : электронный.— URL: <http://lib.ulsu.ru/MegaPro/Download/MObject/7120>.
2. Липатова С. В. Методические рекомендации для семинарских (практических) занятий, лабораторного практикума и самостоятельной работы по дисциплине «Методы разработки систем управления базами данных» для студентов направления 01.04.02 «Прикладная математика и информатика» / С. В. Липатова. - Ульяновск : УлГУ, 2022. - 36 с. - Неопубликованный ресурс. - URL: <http://lib.ulsu.ru/MegaPro/Download/MObject/13454>. - Режим доступа: ЭБС УлГУ. - Текст : электронный.



Министерство образования и науки Российской Федерации Ульяновский государственный университет	Форма	
Ф - Рабочая программа по дисциплине		

Согласовано:

Специалист ведущий НБ УлГУ  
Должность сотрудника научной библиотеки

Боброва Н.А.  
ФИО

  
подпись

/ \_\_\_\_\_ / 2023  
дата

## б) программное обеспечение:

Для образовательного процесса по данной дисциплине необходим стационарный класс ПК с установленным следующим программным обеспечением:

- операционная среда MS Windows;
- пакет приложений MS Office
- СУБД MS SQL и Eclipse;

### 1. Электронно-библиотечные системы:

*Профессиональные базы данных, информационно-справочные системы*

#### 1. Электронно-библиотечные системы:

1.1. Цифровой образовательный ресурс IPRsmart : электронно-библиотечная система : сайт / ООО Компания «Ай Пи Ар Медиа». - Саратов, [2023]. – URL: <http://www.iprbookshop.ru>. – Режим доступа: для зарегистрир. пользователей. - Текст : электронный.

1.2. Образовательная платформа ЮРАЙТ : образовательный ресурс, электронная библиотека : сайт / ООО Электронное издательство «ЮРАЙТ». – Москва, [2023]. - URL: <https://urait.ru>. – Режим доступа: для зарегистрир. пользователей. - Текст : электронный.

1.3. База данных «Электронная библиотека технического ВУЗа (ЭБС «Консультант студента») : электронно-библиотечная система : сайт / ООО «Политехресурс». – Москва, [2023]. – URL: <https://www.studentlibrary.ru/cgi-bin/mb4x>. – Режим доступа: для зарегистрир. пользователей. – Текст : электронный.

1.4. Консультант врача. Электронная медицинская библиотека : база данных : сайт / ООО «Высшая школа организации и управления здравоохранением-Комплексный медицинский консалтинг». – Москва, [2023]. – URL: <https://www.rosmedlib.ru>. – Режим доступа: для зарегистрир. пользователей. – Текст : электронный.

1.5. Большая медицинская библиотека : электронно-библиотечная система : сайт / ООО «Букап». – Томск, [2023]. – URL: <https://www.books-up.ru/ru/library/> . – Режим доступа: для зарегистрир. пользователей. – Текст : электронный.

1.6. ЭБС Лань : электронно-библиотечная система : сайт / ООО ЭБС «Лань». – Санкт-Петербург, [2023]. – URL: <https://e.lanbook.com>. – Режим доступа: для зарегистрир. пользователей. – Текст : электронный.


1.7. ЭБС **Znanium.com** : электронно-библиотечная система : сайт / ООО «Знаниум». - Москва, [2023]. - URL: <http://znanium.com> . – Режим доступа : для зарегистрир. пользователей. - Текст : электронный.

**2. КонсультантПлюс** [Электронный ресурс]: справочная правовая система. / ООО «Консультант Плюс» - Электрон. дан. - Москва : КонсультантПлюс, [2023].

#### 3. Базы данных периодических изданий:

3.1. eLIBRARY.RU: научная электронная библиотека : сайт / ООО «Научная Электронная Библиотека». – Москва, [2023]. – URL: <http://elibrary.ru>. – Режим доступа : для авториз. пользователей. – Текст : электронный

3.2. Электронная библиотека «Издательского дома «Гребенников» (Grebinnikon) : электронная библиотека / ООО ИД «Гребенников». – Москва, [2023]. – URL:

Министерство образования и науки Российской Федерации Ульяновский государственный университет	Форма	
Ф - Рабочая программа по дисциплине		

<https://id2.action-media.ru/Personal/Products>. – Режим доступа : для авториз. пользователей.  
– Текст : электронный.

**4. Федеральная государственная информационная система «Национальная электронная библиотека»** : электронная библиотека : сайт / ФГБУ РГБ. – Москва, [2023].  
– URL: <https://нэб.рф>. – Режим доступа : для пользователей научной библиотеки. – Текст : электронный.

**5. Российское образование** : федеральный портал / учредитель ФГАУ «ФИЦТО».  
– URL: <http://www.edu.ru>. – Текст : электронный.

**6. Электронная библиотечная система УлГУ** : модуль «Электронная библиотека» АБИС Мега-ПРО / ООО «Дата Экспресс». – URL: <http://lib.ulsu.ru/MegaPro/Web>. – Режим доступа : для пользователей научной библиотеки. – Текст : электронный.

Согласовано:  
 УИТТ |  Бурдаков А.П. |  | \_\_\_\_\_  
 Должность сотрудника УИТТ | ФИО | подпись | дата

## 12. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Аудитории для проведения лекций, семинарских занятий, для проведения лабораторных работ, для проведения текущего контроля и промежуточной аттестации.

Помещение 3/321. Аудитории укомплектованы специализированной мебелью, учебной доской. Аудитории для проведения лекций оборудованы мультимедийным оборудованием для представления информации большой аудитории. 432017, Ульяновская область, г. Ульяновск, ул. Набережная реки Свияги, д. 106 (3 корпус).

Помещения для самостоятельной работы оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде, электронно-библиотечной системе.

## 13. СПЕЦИАЛЬНЫЕ УСЛОВИЯ ДЛЯ ОБУЧАЮЩИХСЯ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ

В случае необходимости, обучающимся из числа лиц с ограниченными возможностями здоровья (по заявлению обучающимся) могут предлагаться одни из следующих вариантов восприятия информации с учетом их индивидуальных психофизических возможностей:

- для лиц с нарушением зрения: в форме электронного документа, индивидуальные консультации с привлечением тифлосурдопереводчика, индивидуальные задания и консультация;
- для лиц с нарушением слуха: в форме электронного документа, индивидуальные консультации с привлечением сурдопереводчика, индивидуальные задания и консультация;
- для лиц с нарушением опорно-двигательного аппарата: в форме электронного документа, индивидуальные задания и консультация.

Разработчик  | Шабалин А.С.  
 (Подпись) | ФИО