


Министерство науки и высшего образования РФ Ульяновский государственный университет	Форма	
Ф-Рабочая программа дисциплины		

УТВЕРЖДЕНО

решением Учёного совета факультета математики,
информационных и авиационных технологий

от «16» мая 2023 г., протокол № 4/23

Председатель _____ / М.А. Волков
«16» мая 2023 г.



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Дисциплина	Технологии обработки больших данных
Факультет	Факультет математики, информационных и авиационных технологий
Кафедра	Телекоммуникационные технологии и сети
Курс	1

Направление (специальность) 11.04.02 "Инфокоммуникационные технологии и системы связи "

Уровень магистратура

Профиль подготовки Интеллектуальные телекоммуникационные системы и сети

Форма обучения очная

Дата введения в учебный процесс УлГУ:

«1» сентября 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, учёная степень, звание
Курилова Оксана Леонидовна	Телекоммуникационных технологий и сетей	доцент, к.т.н.

СОГЛАСОВАНО
Заведующий кафедрой телекоммуникационных технологий и сетей, реализующей дисциплину и выпускающей
 / Смагин А.А. /
Подпись ФИО
«16» мая 2023 г.

Министерство науки и высшего образования РФ Ульяновский государственный университет	Форма	
Ф-Рабочая программа дисциплины		

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ:

Цели освоения дисциплины: формирование общекультурных и профессиональных компетенций, необходимых для реализации информационно-аналитической и научно-исследовательской деятельности.

Задачи освоения дисциплины: приобретение в рамках освоения, предусмотренного курсом занятий следующих знаний, умений и навыков, характеризующих определённый уровень сформированности целевых компетенций (см. подробнее п.3):

- приобретение студентами знаний о технологиях подготовки, хранения, обработки и анализа больших данных;
- применение статистических и математических методов для анализа больших объёмов информации;
- приобретение практических навыков работы с нереляционными базами данных;
- приобретение студентами знаний о экосистеме Hadoop.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП:

Дисциплина «Технологии обработки больших данных» относится к числу дисциплин блока Б1.В.07, формируемая участниками образовательных отношений, предназначенного для студентов, обучающихся по направлению: 11.04.02 "Инфокоммуникационные технологии и системы связи".

Основные положения дисциплины используются в дальнейшем при изучении таких дисциплин как: «Проектная деятельность»; «Научно-исследовательская работа».


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

№ п/п	Индекс компетенции	Содержание компетенции (или ее части)	Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с индикаторами достижения компетенций		
			знать	Уметь	Владеть
1	2	3	4	5	6
1	ПК-1 (ПК-1и)	Способен исследовать применение интеллектуальных систем для различных предметных областей	причины возникновения тренда больших данных; процессы анализа больших данных;	формулировать алгоритмы; выбирать подходящий инструмент анализа больших данных	Современными инструментами работы с большими данными.
2	ПК-2 (ПК-4и)	Способен адаптировать и применять методы и алгоритмы машинного обучения для решения прикладных задач в различных предметных областях	основные подходы к обработке больших массивов данных;	выбирать подходящую технологию хранения больших данных.	Современными инструментами работы с большими данными.

4. ОБЩАЯ ТРУДОЕМКОСТЬ ДИСЦИПЛИНЫ

4.1. Объем дисциплины в зачётных единицах (всего) 2


4.2. Объем дисциплины по видам учебной работы (в часах) 72 ч.

Министерство науки и высшего образования РФ Ульяновский государственный университет	Форма	
Ф-Рабочая программа дисциплины		

Вид учебной работы	Количество часов (форма обучения очная)	
	Всего по плану	В т.ч. по семестрам
		1
1	2	3
Контактная работа обучающихся с преподавателем в соответствии с УП	36	36
Аудиторные занятия:	36	36
Лекции	18	18
Семинары и практические занятия	18	18
Лабораторные работы, практикумы	-	-
Самостоятельная работа	36	36
Форма текущего контроля знаний и контроля самостоятельной работы: тестирование, контр. работа, коллоквиум, реферат и др. (не менее 2 видов)		
Курсовая работа	-	-
Виды промежуточной аттестации (экзамен, зачет)	зачёт	зачёт
Всего часов по дисциплине	72	72

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

Название разделов и тем	Всего	Виды учебных занятий					Форма текущего контроля знаний
		Аудиторные занятия			Занятия в интерактивной форме	Самостоятельная работа	
		Лекции	Практические занятия, семинары	Лабораторные работы, практикумы			
1	2	3	4	5	6	7	
Тема 1. Введение в BigData	9	2	2	-	-	5	Опрос
Тема 2. Хранилища данных. Технологии OLTP, OLAP, ETL	11	3	3	-	2	5	Опрос
Тема 3. Технологии NoSQL	9	3	3	-	2	3	Опрос, реферат

Министерство науки и высшего образования РФ Ульяновский государственный университет	Форма	
Ф-Рабочая программа дисциплины		

Тема 4. Экосистема Hadoop.	7	2	2	-	2	3	Опрос
Тема 5. Распределённые файловые системы	9	2	2	-	2	5	Опрос
Тема 6. MapReduce: методология и технология распределённых вычислений.	9	2	2	-	2	5	Опрос
Тема 7. Обработка данных в реальном времени.	9	2	2	-	-	5	Опрос
Тема 8. Массово-параллельная структура - Massive Parallel Processing.	9	2	2	-	-	5	Опрос
Итого	72	18	18	-	10*	36	-

* Столбец «Занятия в интерактивной форме» в подсчёте итогов не участвует.

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Тема 1. Введение в BigData.

Основные вызовы больших данных(V). Определение термина "большие данные". Классификация Big Data. Роли игроков на рынке Big Data. Профессии Big Data. Примеры успешных кейсов.

Тема 2. Хранилища данных. Технологии OLTP, OLAP, ETL.

Принципы технологии OLTP. Понятие транзакции. Способы организации транзакций и принципы блокировки доступа к данным. Назначение технологии.

Определение и свойства хранилищ данных, виды данных, хранящихся в хранилищах. Многомерная модель представления данных. Технологии BI и ETL, OLAP. Виды реализации многомерной модели данных. СУБД, обеспечивающие поддержку OLAP.

Тема 3. Технологии NoSQL.

Горизонтальное и вертикальное масштабирование. CAP-теорема. История термина NoSQL и его трактование. BASE-архитектура (Basically Available, Soft-state, Eventually consistent). Графовые, колончатые, документо-ориентированные модели модель и ключ-значение. Термин NewSQL. СУБД: HBase, Cassandra, Neo4j, MongoDB.

Тема 4. Экосистема Hadoop.

Основные принципы Hadoop, компоненты, примеры использования. Проблемы разработки для крупных параллельных систем


Тема 5. Распределённые файловые системы.

Структура РФС, требования, примеры: HDFS, Google, LustreFS. HDFS: архитектура, основные узлы, ограничения, основные команды.

Тема 6. MapReduce: методология и технология распределённых вычислений.

Основные идеи MapReduce, Этапы Map – предварительной обработки и Reduce – свертки результатов. Примеры функций. Различные решения: Apache Hadoop, Erlang, в MongoDB, в CouchDB.

Тема 7. Обработка данных в реальном времени.

Министерство науки и высшего образования РФ Ульяновский государственный университет	Форма	
Ф-Рабочая программа дисциплины		

Storm, Spark, Impal: основные идеи, сравнение достоинств и недостатков. Примеры использования.

Тема 8. Массово-параллельная структура - Massive Parallel Processing.

Масштабирование реляционных баз данных. Параллельное выполнение запросов к БД. Архитектура Hub and Spoke.

6. ТЕМЫ ПРАКТИЧЕСКИХ И СЕМИНАРСКИХ ЗАНЯТИЙ

Тема 2. Хранилища данных. Технологии OLTP, OLAP, ETL.

Рассмотрение особенностей процесса ETL на базе библиотеки PETA на языке Python.

Тема 3. Технологии NoSQL.

Рассмотрение особенностей использования документно-ориентированной базы данных на примере MongoDB.

Тема 4 - 5. Экосистема Hadoop.

Рассмотрение особенностей экосистемы Hadoop в виртуальной машине Cloudera QuickStart.

Тема 6. MapReduce: методология и технология распределенных вычислений.

Рассмотрение особенностей технологии MapReduce в базе данных MongoDB.

7. ЛАБОРАТОРНЫЕ РАБОТЫ, ПРАКТИКУМЫ


Данный вид работы не предусмотрен УП.

8. ТЕМАТИКА КУРСОВЫХ, КОНТРОЛЬНЫХ РАБОТ, РЕФЕРАТОВ

Данный вид работы не предусмотрен УП.

9. ПЕРЕЧЕНЬ ВОПРОСОВ К ЭКЗАМЕНУ (ЗАЧЕТУ)

1	Данные, информация, знания, модели. Наука о данных.
2	Эволюционное развитие архитектур и данных.
3	Технологии BigData: определение для BigData, назначение BigData, ключевые характеристики, история появления и основные принципы BigData. Основные виды данных.
4	Критерии больших данных. Источники больших данных. Интернет вещей.
5	Возможные этапы работы с большими данными.
6	Технология OLTP. Основные принципы, назначение, понятие транзакции, уровни блокировки, организация многопользовательского доступа.
7	Хранилища данных (определение и свойства хранилищ данных, виды данных, хранящихся в хранилищах, многомерная модель представления данных)
8	Технологии BI и ETL, OLAP.
9	Виды реализации многомерной модели данных. СУБД, обеспечивающие поддержку OLAP.
10	Специфика хранения и обработки больших данных.
11	Распределённые файловые системы. Файловая система HDFS.
12	Основные отличия NoSQL от реляционных баз данных. Достоинства и недостатки этих технологий.

Министерство науки и высшего образования РФ Ульяновский государственный университет	Форма	
Ф-Рабочая программа дисциплины		


13 Парадигма Map Reduce. Принцип работы. Слабые и сильные стороны, области применимости.
14 Модели представления данных NoSQL: графовая, колончатая, ключ-значение.
15 Масштабирование реляционных баз данных.
16 Параллельное выполнение запросов к БД.
17 Архитектура Hub and Spoke.
18 Storm, Spark, Impal: основные идеи, сравнение достоинств и недостатков.

10. САМОСТОЯТЕЛЬНАЯ РАБОТА ОБУЧАЮЩИХСЯ

Содержание, требования, условия и порядок организации самостоятельной работы обучающихся с учетом формы обучения определяются в соответствии с «Положением об организации самостоятельной работы обучающихся», утвержденным Ученым советом УлГУ (протокол №8/268 от 26.03.2019 г.).

Методические рекомендации для семинарских (практических) занятий и самостоятельной работы представлены в учебно-методической литературе [1].

Название разделов и тем	Вид самостоятельной работы	Объем в часах	Форма контроля
Тема 1. Введение в BigData	чтение основной и дополнительной литературы, самостоятельное изучение материала по литературным источникам;	5	Опрос
Тема 2. Хранилища данных. Технологии OLTP, OLAP, ETL	чтение основной и дополнительной литературы, самостоятельное изучение материала по литературным источникам;	5	Опрос
Тема 3. Технологии NoSQL	чтение основной и дополнительной литературы, самостоятельное изучение материала по литературным источникам;	3	Опрос, реферат
Тема 4. Экосистема Hadoop.	чтение основной и дополнительной литературы, самостоятельное изучение материала по литературным источникам;	3	Опрос
Тема 5. Распределённые файловые системы	самостоятельное выполнение практических заданий репродуктивного типа (ответы на вопросы, тренировочные упражнения, задачи, тесты);	5	Опрос, реферат
Тема 6. MapReduce: методология и технология распреде-	самостоятельное выполнение практических заданий репродуктивного типа (ответы на вопросы, тренировочные упражнения, задачи, тесты);	5	Опрос

Министерство науки и высшего образования РФ Ульяновский государственный университет	Форма	
Ф-Рабочая программа дисциплины		

ленных вычислений.			
Тема 7. Обработка данных в реальном времени.	самостоятельное выполнение практических заданий репродуктивного типа (ответы на вопросы, тренировочные упражнения, задачи, тесты);	5	Опрос, реферат
Тема 8. Массово-параллельная структура - Massive Parallel Processing.	самостоятельное выполнение практических заданий репродуктивного типа (ответы на вопросы, тренировочные упражнения, задачи, тесты);	5	Опрос
Всего		36	

11. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

а) Список рекомендуемой литературы

основная


1. Воронова, Л. И. Big Data. Методы и средства анализа : учебное пособие / Л. И. Воронова, В. И. Воронов. — Москва : Московский технический университет связи и информатики, 2016. — 33 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/61463.html>
2. Иванова, С. М. Теория информации. Хранение и передача данных : учебное пособие / С. М. Иванова, З. В. Ильиченкова. — Москва : РТУ МИРЭА, 2022. — 75 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/256583>

дополнительная

1. Анализ данных : Учебник для вузов / Мхитарян Владимир Сергеевич, Архипова Марина Юрьевна, Дуброва Татьяна Абрамовна [и др.]; под ред. Мхитаряна В.С. - Москва : Юрайт, 2022. - 490 с. - (Высшее образование). - URL: <https://urait.ru/bcode/489100>
- 2.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>
- 3.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>

учебно-методическая

1. Липатова С. В. Методические рекомендации для лабораторного практикума и самостоятельной работы по дисциплине «Технологии обработки больших данных» для студентов направления 11.04.02 "Инфокоммуникационные технологии и системы связи" / С. В. Липатова. - Ульяновск : УлГУ, 2022. - 29 с. - Неопубликованный ресурс. - URL: <http://lib.ulsu.ru/MegaPro/Download/MObject/13307>. - Режим доступа: ЭБС УлГУ. - Текст : электронный.

Министерство науки и высшего образования РФ Ульяновский государственный университет	Форма	
Ф-Рабочая программа дисциплины		

Согласовано:

Специалист ведущий НБ УлГУ
Должность сотрудника научной библиотеки

Боброва Н.А.
ФИО


подпись

/ _____ 2023
дата

б) Программное обеспечение

1. Anaconda (open source)

в) 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: <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.пф>. – Режим доступа : для пользователей научной библиотеки. – Текст : электронный.

5. Российское образование : федеральный портал / учредитель ФГАУ «ФИЦТО». –

Министерство науки и высшего образования РФ Ульяновский государственный университет	Форма	
Ф-Рабочая программа дисциплины		

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:

 |  |  | _____
 Должность сотрудника УИТиТ | ФИО | подпись | дата

12. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ:

Аудитории для проведения лекций (лекционные аудитории 3 корпуса УлГУ), семинарских занятий (лекционные аудитории 3 корпуса УлГУ), для выполнения лабораторных работ и практикумов (дисплейные классы 1 корпуса УлГУ), для проведения текущего контроля и промежуточной аттестации (лекционные аудитории 3 корпуса УлГУ).

Аудитории укомплектованы специализированной мебелью, учебной доской. Аудитории для проведения лекций оборудованы мультимедийным оборудованием для предоставления информации большой аудитории. Помещения для самостоятельной работы оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде, электронно-библиотечной системе. Перечень оборудования, используемого в учебном процессе, указывается в соответствии со сведениями о материально-техническом обеспечении и оснащённости образовательного процесса, размещёнными на официальном сайте УлГУ в разделе «Сведения об образовательной организации».

13. СПЕЦИАЛЬНЫЕ УСЛОВИЯ ДЛЯ ОБУЧАЮЩИХСЯ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ

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

– для лиц с нарушениями зрения: в форме электронного документа; индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания и консультации;

– для лиц с нарушениями слуха: в печатной форме; в форме электронного документа; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания и консультации;

– для лиц с нарушениями опорно-двигательного аппарата: в печатной форме; в форме электронного документа; индивидуальные задания и консультации.

В случае необходимости использования в учебном процессе частично/исключительно дистанционных образовательных технологий, организация работы ППС с обучающимися с ОВЗ и инвалидами предусматривается в электронной информационно-образовательной среде с учетом их индивидуальных психофизических особенностей.

Разработчик: _____



подпись

_____ доцент кафедры

должность

Курилова Оксана Леонидовна

ФИО