

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 «17» мая 2022 г., протокол № 4/22

Председатель / М.А. Волков
«17» мая 2022 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"

код направления (специальности), полное наименование

Направленность (профиль/специализация) Интеллектуальные инфокоммуникационные технологии и сети

полное наименование

Форма обучения очная

очная, заочная, очно-заочная (указать только те, которые реализуются)

Дата введения в учебный процесс УлГУ:

«1» сентября 2022 г.

Программа актуализирована на заседании кафедры: протокол № 9 от 31.05.2023 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
Липатова Светлана Валерьевна	Телекоммуникационных технологий и сетей	доцент, к.т.н., доцент

СОГЛАСОВАНО

Заведующий кафедрой телекоммуникационных технологий и сетей, реализующей дисциплину и выпускающей

 / Смагин А.А. /

Подпись / ФИО

«17» мая 2022 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, 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

4.

Код и наименование реализуемой компетенции	Перечень планируемых результатов обучения по дисциплине (модулю), соотнесённых с индикаторами достижения компетенций
ПК-3 Способностью обеспечения информационной безопасности системного программного обеспечения инфокоммуникационной системы организации	Знать: причины возникновения тренда больших данных; процессы анализа больших данных; основные подходы к обработке больших массивов данных; Уметь: формулировать алгоритмы; выбирать подходящий инструмент анализа больших данных; выбирать подходящую технологию хранения больших данных. Владеть: Современными инструментами работы с большими данными.
ПК-4 Способностью организовывать и проводить экспериментальные испытания с целью оценки и улучшения качества предоставляемых услуг связи, соответствия требованиям технических регламентов, международных и национальных стандартов и иных нормативных документов	Знать: причины возникновения тренда больших данных; процессы анализа больших данных; основные подходы к обработке больших массивов данных; Уметь: формулировать алгоритмы; выбирать подходящий инструмент анализа больших данных; выбирать подходящую технологию хранения больших данных. Владеть: Современными инструментами работы с большими данными.
ПК-10 Способен к выполнению работ по обеспечению функционирования	Знать: причины возникновения тренда больших данных; процессы анализа больших данных; основные подходы к обработке больших массивов данных;

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

телекоммуникационного оборудования корпоративных сетей с учетом требований информационной безопасности	Уметь: формулировать алгоритмы; выбирать подходящий инструмент анализа больших данных; выбирать подходящую технологию хранения больших данных. Владеть: Современными инструментами работы с большими данными.
--	--

5. ОБЩАЯ ТРУДОЕМКОСТЬ ДИСЦИПЛИНЫ

4.1. Объем дисциплины в зачётных единицах (всего) 3

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

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	36/36*
Лекции	18	18/18*
Семинары и практические занятия	18	18/18*
Лабораторные работы, практикумы	-	-
Самостоятельная работа	72	72
Форма текущего контроля знаний и контроля самостоятельной работы	Опрос, реферат	Опрос, реферат
Курсовая работа	-	-
Виды промежуточной аттестации (экзамен, зачет)	зачёт	зачёт
Всего часов по дисциплине	108	108

*Количество часов работы ППС с обучающимися в дистанционном формате с применением электронного обучения

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	17	2	2	-	-	13	Опрос
Тема 2. Хранилища данных. Технологии OLTP, OLAP, ETL	19	3	3	-	2	13	Опрос
Тема 3.	19	3	3	-	2	13	Опрос,

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

Технологии NoSQL							реферат
Тема 4. Экосистема Hadoop.	17	2	2	-	2	13	Опрос
Тема 5. Распределённые файловые системы	18	2	2	-	2	14	Опрос
Тема 6. MapReduce: методология и технология распределённых вычислений.	18	2	2	-	2	14	Опрос
Тема 7. Обработка данных в реальном времени.	18	2	2	-	-	14	Опрос
Тема 8. Массово-параллельная структура - Massive Parallel Processing.	18	2	2	-	-	14	Опрос
Итого	144	18	18	-	-	72	-

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

6. 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 –

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

свертки результатов. Примеры функций. Различные решения: Apache Hadoop, Erlang, в MongoDB, в CouchDB.

Тема 7. Обработка данных в реальном времени.

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.

7. 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.

8. ЛАБОРАТОРНЫЕ РАБОТЫ, ПРАКТИКУМЫ

Данный вид работы не предусмотрен УП.

9. ТЕМАТИКА КУРСОВЫХ, КОНТРОЛЬНЫХ РАБОТ, РЕФЕРАТОВ

Данный вид работы не предусмотрен УП.

10. 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.

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

12 Основные отличия NoSQL от реляционных баз данных. Достоинства и недостатки этих технологий.
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.

11. САМОСТОЯТЕЛЬНАЯ РАБОТА ОБУЧАЮЩИХСЯ

Форма обучения очная

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;	13	Опрос
Тема 2. Хранилища данных. Технологии OLTP, OLAP, ETL	чтение основной и дополнительной литературы, самостоятельное изучение материала по литературным источникам;	13	Опрос
Тема 3. Технологии NoSQL	чтение основной и дополнительной литературы, самостоятельное изучение материала по литературным источникам;	13	Опрос, реферат
Тема 4. Экосистема Hadoop.	чтение основной и дополнительной литературы, самостоятельное изучение материала по литературным источникам;	13	Опрос
Тема 5. Распределённые файловые системы	самостоятельное выполнение практических заданий репродуктивного типа (ответы на вопросы, тренировочные упражнения, задачи, тесты);	14	Опрос, реферат
Тема 6. MapReduce: методология и технология распределенных вычислений.	самостоятельное выполнение практических заданий репродуктивного типа (ответы на вопросы, тренировочные упражнения, задачи, тесты);	14	Опрос
Тема 7. Обработка данных в реальном	самостоятельное выполнение практических заданий репродуктивного	14	Опрос, реферат

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

времени.	типа (ответы на вопросы, тренировочные упражнения, задачи, тесты);		
Тема 8. Массово-параллельная структура - Massive Parallel Processing.	самостоятельное выполнение практических заданий репродуктивного типа (ответы на вопросы, тренировочные упражнения, задачи, тесты);	14	Опрос

10. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

а) Основная литература

- 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>
- 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 литература:

- 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>
- 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>
- 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

- 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:

ДИРЕКТОР НБ / БУРХАНОВА М.М. / 2022
 Должность сотрудника научной библиотеки ФИО подпись 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, [2022]. – 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, [2022]. - 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, [2022]. – 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, [2022]. – URL: <https://www.rosmedlib.ru>. – Режим доступа: для зарегистрир. пользователей. – Текст : электронный.

1.5. Большая медицинская библиотека : электронно-библиотечная система : сайт / ООО Букап. – Томск, [2022]. – 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, [2022]. – URL: <https://e.lanbook.com>. – Режим доступа: для зарегистрир. пользователей. – Текст : электронный.

1.7. ЭБС Znanium.com : электронно-библиотечная система : сайт / ООО Знаниум. - Москва, [2022]. - URL: <http://znanium.com>. – Режим доступа : для зарегистрир. пользователей. - Текст : электронный.

1.8. Clinical Collection : научно-информационная база данных EBSCO // EBSCOhost : [портал]. – URL: <http://web.b.ebscohost.com/ehost/search/advanced?vid=1&sid=9f57a3e1-1191-414b-8763-e97828f9f7e1%40sessionmgr102>. – Режим доступа : для авториз. пользователей. – Текст : электронный.

1.9. База данных «Русский как иностранный» : электронно-образовательный ресурс для иностранных студентов : сайт / ООО Компания «Ай Пи Ар Медиа». – Саратов, [2022]. – URL: <https://ros-edu.ru>. – Режим доступа: для зарегистрир. пользователей. – Текст : электронный.

2. КонсультантПлюс [Электронный ресурс]: справочная правовая система. /ООО «Консультант Плюс» - Электрон. дан. - Москва : КонсультантПлюс, [2022].

3. Базы данных периодических изданий:

3.1. База данных периодических изданий EastView : электронные журналы / ООО ИВИС. - Москва, [2022]. – URL: <https://dlib.eastview.com/browse/udb/12>. – Режим доступа : для авториз. пользователей. – Текст : электронный.

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

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

должность

С.В. Липатова

ФИО