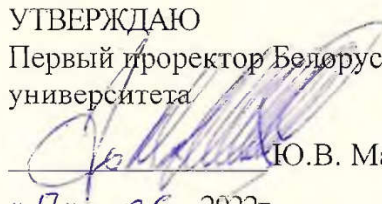


Межгосударственное образовательное учреждение высшего образования
«Белорусско-Российский университет»

УТВЕРЖДАЮ

Первый проректор Белорусско-Российского
университета

 Ю.В. Машин

« 17 » 06 2022г.

Регистрационный № УД-090304/Б.Р.В.16/р

МЕТОДЫ СБОРА, ОБРАБОТКИ И АНАЛИЗА ДАННЫХ

(наименование дисциплины)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Направление подготовки 09.03.04 Программная инженерия

Направленность (профиль) Разработка программно-информационных систем

Квалификация Бакалавр

	Форма обучения
	Очная
Курс	4
Семестр	7
Лекции, часы	14
Лабораторные занятия, часы	30
Зачет, семестр	7
Контактная работа по учебным занятиям, часы	44
Самостоятельная работа, часы	28
Всего часов / зачетных единиц	72/2

Кафедра-разработчик программы: Программное обеспечение информационных технологий

(название кафедры)

Составитель: К.В. Захарченков, к.т.н.; Ю.В. Вайнилович, к.т.н.

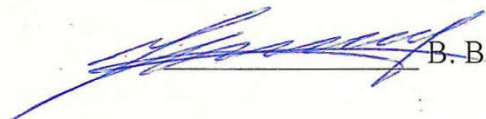
(И.О. Фамилия, ученая степень, ученое звание)

Могилев, 2022

Рабочая программа составлена в соответствии с федеральным государственным образовательным стандартом высшего профессионального образования по направлению подготовки 09.03.04 – «Программная инженерия» (уровень бакалавриата), утвержденным приказом № 920 от 19.09.2017г. и учебным планом Рег. № 090304-5 от 25.03.2022.

Рассмотрена и рекомендована к утверждению кафедрой «Программное обеспечение информационных технологий» «08» апреля 2022 г., протокол № 10.

Зав. кафедрой ПОИТ

 В. В. Кутузов

Одобрена и рекомендована к утверждению Научно-методическим советом Белорусско-Российского университета

«15» июня 2022 г., протокол № 7.

Зам. председателя
Научно-методического совета

 С.А. Сухоцкий

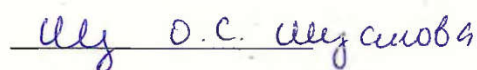
Рецензент:

С.В. Миренков, начальник управления информационных технологий ОАО «Лента»

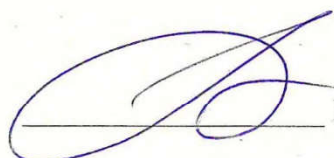
(И.О. Фамилия, должность, ученая степень, ученое звание рецензента)

Рабочая программа согласована:

Ведущий библиотекарь



Начальник учебно-методического
отдела

 В.А. Кемова

1 ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

1.1 Цель учебной дисциплины

Основной целью преподавания дисциплины является получение теоретических знаний о методах сбора, обработки и анализа данных в распределенных системах, а также приобретение практических навыков автоматизированной обработки больших объемов данных в сложных программных системах.

Дисциплина «Методы сбора, обработки и анализа данных» обеспечивает приобретение знаний и умений в соответствии с государственным образовательным стандартом, содействует формированию мировоззрения и системного мышления.

1.2 Планируемые результаты изучения дисциплины

В результате освоения учебной дисциплины студент должен

знать:

- процессы сбора, обработки и анализа больших объемов данных, основы машинного обучения, методы распределенного хранения и обработки больших объемов данных с использованием современных информационных технологий, методы глубокого анализа текста, методы визуализации данных.

уметь:

- решать задачи хранения, сбора, обработки и анализа больших объемов данных в распределенных системах.

владеть:

- навыками использования современных программных средств для хранения, сбора, обработки и анализа больших объемов данных в распределенных системах.

1.3 Место учебной дисциплины в системе подготовки студента

Дисциплина относится к блоку 1 «Часть Блока 1, формируемая участниками образовательных отношений»

Перечень учебных дисциплин, изучаемых ранее, усвоение которых необходимо для изучения данной дисциплины:

- программирование;
- базы данных;
- технологии командной разработки приложений.

Перечень учебных дисциплин (циклов дисциплин), которые будут опираться на данную дисциплину:

- современные системы программирования (8 семестр);
- архитектура программных систем.

Кроме того, знания, полученные при изучении дисциплины на лабораторных занятиях, будут использованы при прохождении преддипломной практики, а также при подготовке выпускной квалификационной работы.

1.4 Требования к освоению учебной дисциплины

Освоение данной учебной дисциплины должно обеспечивать формирование следующих компетенций:

Коды формируемых компетенций	Наименования формируемых компетенций
ПК-3	Способность оформления методических материалов и пособий по применению программных систем
ПК-6	Владение навыками моделирования, анализа и использования формальных методов конструирования программного обеспечения

2 СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Вклад дисциплины в формирование результатов обучения выпускника (компетенций) и достижение обобщенных результатов обучения происходит путём освоения содержания обучения и достижения частных результатов обучения, описанных в данном разделе.

2.1 Содержание учебной дисциплины

Номер тем	Наименование тем	Содержание	Коды формируемых компетенций
1	Введение.	Введение Область применения больших данных и их преимущества Структурированные и неструктурированные данные Данные на естественном языке Машинные данные Графовые и сетевые данные	ПК-3 ПК-6
2	Процесс анализа больших объемов данных	Сбор данных Подготовка данных Исследование данных Моделирование данных Визуальное представление результатов обработки данных	ПК-3 ПК-6
3	Программное обеспечение для автоматизации обработки больших объемов данных	Распределенные файловые системы Средства распределенного программирования приложений Программные средства интеграции данных Программные средства машинного обучения Программные средства сравнительного анализа данных Развертывание распределенной системы обработки данных	ПК-3 ПК-6
4	Машинное обучение	Понятие машинного обучения Типы машинного обучения Процесс моделирования при машинном обучении Обучение модели Проверка адекватности модели Прогнозирование новых наблюдений	ПК-3 ПК-6
5	Общие методы обработки больших объемов данных	Проектирование алгоритмов обработки больших объемов данных Проектирование архитектуры распределенной системы обработки больших объемов данных	ПК-3 ПК-6
6	Средства распределенного хранения и обработки больших данных	Проблемы создания распределенных баз данных Hadoop Spark	ПК-3 ПК-6
7	Глубокий анализ текста	Применение глубокого анализа текста Методы глубокого анализа текста Концепция набора слов Выделение основы и лемматизация Классификатор на базе дерева принятия решений	ПК-3 ПК-6

2.2 Учебно-методическая карта учебной дисциплины

№ недели	Лекции (наименование тем)	Часы	Практические (семинарские) занятия	Часы	Лабораторные занятия	Часы	Самостоятельная работа, часы	Форма контроля знаний	Баллы (max)
Модуль 1									
1	Тема 1. Введение	2			Лр №1 – Использование Hadoop для анализа больших объемов данных	2	1		
2					Лр №1 - Использование Hadoop для анализа больших объемов данных	2	1		
3	Тема 2. Процесс анализа больших объемов данных	2			Лр №1 - Использование Hadoop для анализа больших объемов данных	2	2	ЗЛР	15
4					Лр №2 – Сбор данных для анализа	2	2		
5	Тема 3. Программное обеспечение для автоматизации обработки больших объемов данных	2			Лр №2 - Сбор данных для анализа	2	2		
6					Лр №2 - Сбор данных для анализа	2	2	ЗЛР	15
7	Тема 4. Машинное обучение	2			Лр №3 – Построение модели	2	2		
8					Лр №3 - Построение модели	2	2	ПКУ	30
Модуль 2									
9	Тема 5. Общие методы обработки больших объемов данных	2			Лр №3 - Построение модели	2	2	ЗЛР	10
10					Лр №4 - Исследование модели	2	2		
11	Тема 6. Средства распределенного хранения и обработки больших данных	2			Лр №4 - Исследование модели	2	2		
12					Лр №4 - Исследование модели	2	2	ЗЛР	10
13	Тема 7. Глубокий анализ текста	2			Лр №5 - Анализ и визуальное представление результатов моделирования	2	2		
14					Лр №5 - Анализ и визуальное представление результатов моделирования	2	2		
15					Лр №5 - Анализ и визуальное представление результатов моделирования	2	2	ЗЛР ПКУ ПА (зачет)	10 30 40
Итого		14				30	28		100

Принятые обозначения:

Текущий контроль –

ЗЛР – защита лабораторной работы;

ПКУ – промежуточный контроль успеваемости.

ПА - Промежуточная аттестация.

Итоговая оценка определяется как сумма текущего контроля и промежуточной аттестации и соответствует баллам:

Зачет

Оценка	Зачтено	Незачтено
Баллы	51–100	0–50

3 ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

При изучении дисциплины используется модульно-рейтинговая система оценки знаний студентов. Применение форм и методов проведения занятий при изучении различных тем курса представлено в таблице.

№ п/п	Форма проведения занятия*	Вид аудиторных занятий**			Всего часов
		Лекции	Практические занятия	Лабораторные занятия	
1	Мультимедиа	Лекции №1-7			14
2	С использованием ЭВМ			Лр№1 -Лр№5	30
	ИТОГО				44

4 ОЦЕНОЧНЫЕ СРЕДСТВА

Используемые оценочные средства по учебной дисциплине представлены в таблице и хранятся на кафедре.

№ п/п	Вид оценочных средств	Количество комплектов
1	Вопросы к зачету	1
2	Задания для защиты лабораторных работ	5

5 МЕТОДИКА И КРИТЕРИИ ОЦЕНКИ КОМПЕТЕНЦИЙ СТУДЕНТОВ

5.1 Уровни сформированности компетенций

№ п/п	Уровни сформированности компетенции	Содержательное описание уровня*	Результаты обучения**
ПК-3: Способность оформления методических материалов и пособий по применению программных систем			
ИПК-3.2: Способен оформлять документацию по программным системам, использующимся в научных исследованиях			
1	Пороговый уровень	Знание стандартов оформления методических материалов	Знание и понимание процессов оформления

			методических материалов по применению программных систем в научных исследованиях
2	Продвинутый уровень	Применение стандартов оформления при подготовке документации программных систем	Анализ документации по применению программных систем для сбора, анализа и обработки данных
3	Высокий уровень	Синтез и оценка способов описания архитектуры программных систем	Синтез способов описания процессов сбора, анализа и обработки данных
ПК-6: Владение навыками моделирования, анализа и использования формальных методов конструирования программного обеспечения			
ИПК-6.1: Владеет навыками моделирования и анализа, используемыми при конструировании программного обеспечения			
1	Пороговый уровень	Знание методов сбора, анализа и обработки больших объемов данных	Знание и понимание процессов сбора, анализа и обработки больших объемов данных
2	Продвинутый уровень	Применение методов моделирования и анализа больших объемов данных в распределенных системах	Анализ структуры модели и результатов моделирования.
3	Высокий уровень	Синтез и оценка способов моделирования и анализа больших объемов данных в распределенных системах	Синтез способов построения моделей и анализа результатов моделирования

5.2 Методика оценки знаний, умений и навыков студентов

Результаты обучения	Оценочные средства*
ПК-3: Способность оформления методических материалов и пособий по применению программных систем	
Знание и понимание процессов оформления методических материалов по применению программных систем в научных исследованиях	Задание для защиты лабораторных работ №№ 1-5. Вопросы к зачету.
Анализ документации по применению программных систем для сбора, анализа и обработки данных	Задание для защиты лабораторных работ №№ 1-5. Вопросы к зачету.
Синтез способов описания процессов сбора, анализа и обработки данных	Задание для защиты лабораторных работ №№ 1-5. Вопросы к зачету.
ПК-6: Владение навыками моделирования, анализа и использования формальных методов конструирования программного обеспечения	
Знание и понимание процессов сбора, анализа и обработки больших объемов данных	Задание для защиты лабораторных работ №№ 1-5. Вопросы к зачету.
Анализ структуры модели и результатов моделирования.	Задание для защиты лабораторных работ №№ 1-5. Вопросы к зачету.
Синтез способов построения моделей и анализа результатов моделирования	Задание для защиты лабораторных работ №№ 1-5. Вопросы к зачету.

5.3 Критерии оценки лабораторных работ

Каждая выполненная и защищенная лабораторная работа оцениваются в диапазоне от 7 до 15 баллов. При этом 6 баллов начисляется за выполнение работы и от 1 до 9 баллов за оформление отчета и защиту работы в зависимости от качества оформления и уровня знаний студента по тематике работы. Если по окончании модуля лабораторная работа выполнена, но не защищена, то баллы по ней не начисляются и она попадает в разряд задолженности.

5.4 Критерии оценки зачета

Контрольное задание включает 2 теоретических вопроса. Теоретические вопросы выбираются из разных дидактических единиц. Каждый вопрос оценивается положительной оценкой в диапазоне от 10 до 20 баллов. Ответы на вопросы оцениваются по следующим критериям.

Теоретические вопросы:

- ♦ **19-20 баллов** – студент глубоко понимает пройденный материал, отвечает четко и всесторонне, умеет оценивать факты, использует научную терминологию, самостоятельно рассуждает, отличается способностью обосновать выводы и разъяснять их в логической последовательности, дает развернутый ответ на поставленный вопрос и четко отвечает на дополнительные вопросы.
- ♦ **17-18 баллов** – студент глубоко понимает пройденный материал, отвечает четко и всесторонне, умеет оценивать факты, самостоятельно рассуждает, отличается способностью обосновать выводы и разъяснять их в логической последовательности, но допускает отдельные неточности, в том числе и на дополнительные вопросы.
- ♦ **15-16 баллов** – студент хорошо понимает пройденный материал, отвечает правильно, умеет оценивать факты, самостоятельно рассуждает, обосновывает выводы и разъясняет их, но допускает ошибки общего характера.
- ♦ **13-14 баллов** – студент понимает пройденный материал, но не может теоретически обосновать некоторые выводы, допускает ошибки общего характера.
- ♦ **11-12 баллов** – студент отвечает в основном правильно на поставленный вопрос, но чувствуется механическое заучивание материала, отсутствует логическая последовательность при изложении ответа, не может ответить на дополнительные вопросы.
- ♦ **10 баллов** – в ответе студента имеются существенные недостатки, материал охвачен «половинчато», в рассуждениях допускаются ошибки.
- ♦ **Ниже 10 баллов** – студент имеет общее представление о вопросе, ответ студента правилен лишь частично, при разъяснении материала допускаются серьезные ошибки, отсутствует техническая терминология, не может исправить ошибки с помощью наводящих вопросов;

6 МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ОРГАНИЗАЦИИ И ВЫПОЛНЕНИЮ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ ПО УЧЕБНОЙ ДИСЦИПЛИНЕ

Самостоятельная работа студентов (СРС) направлена на закрепление и углубление освоения учебного материала, развитие практических умений. СРС включает следующие виды самостоятельной работы студентов:

Перечень контрольных вопросов и заданий для самостоятельной работы студентов приведен в приложении и хранится на кафедре.

Для СРС рекомендуется использовать источники, приведенные в п. 7.

7 УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

7.1 Основная литература

№ п/п	Библиографическое описание	Гриф	Количество экземпляров/URL
1	Мартишин, С. А. Базы данных: Работа с распределенными базами данных и файловыми системами на примере MongoDB и HDFS с использованием Node.js, Express.js, Apache Spark и Scala : учебное пособие / С.А. Мартишин, В.Л. Симонов, М.В. Храпченко. — М. : ИНФРА-М, 2021. — 235 с.	Рекомендовано в качестве учебного пособия для студентов, бакалавров и магистрантов направлений подготовки «Информатика и вычислительная техника», «Информационные системы и технологии», «Программная инженерия», «Прикладная математика и информатика», «Прикладная информатика», «Управление в технических системах», «Бизнес-информатика», «Педагогическое образование» (профиль «Информатика»)	https://znanium.com/catalog/product/1214862

7.2 Дополнительная литература

№ п/п	Библиографическое описание	Гриф	Количество экземпляров
1	Варфоломеева, А. О. Информационные системы предприятия : учебное пособие / А.О. Варфоломеева, А.В. Коряковский, В.П. Романов. — 2-е изд., перераб. и доп. — М. : ИНФРА-М, 2022. — 330 с. — (Высшее образование: Бакалавриат).	Рекомендовано Учебно-методическим объединением по образованию в области прикладной информатики в качестве учебного пособия для студентов высших учебных заведений, обучающихся по направлению 09.03.03 «Прикладная информатика» и экономическим специальностям	https://znanium.com/catalog/product/1844303

2	Григорьев, А. А. Передача, хранение и обработка больших объемов научных данных : учебное пособие / А.А. Григорьев, Е.А. Исаев, П.А. Тарасов. — М. : ИНФРА-М, 2021. — 207 с. — (Высшее образование: Бакалавриат).	Рекомендовано Межрегиональным учебно-методическим советом профессионального образования в качестве учебного пособия для студентов высших учебных заведений, обучающихся по направлениям подготовки 01.03.02 «Прикладная математика и информатика», 09.03.01 «Информатика и вычислительная техника», 38.03.05 Бизнес-информатика» квалификация (степень) «бакалавр»)	https://znanium.com/catalog/product/1073525
3	Келлехер, Д. Наука о данных: базовый курс / Джон Келлехер, Брендан Тирни ; пер. с англ.. - М. : Альпина Паблишер, 2020. - 222 с.	—	https://znanium.com/catalog/product/1221800

7.3 Перечень ресурсов сети Интернет по изучаемой дисциплине

1. *Intuit.ru*
2. *Edx.com*

7.4 Перечень наглядных и других пособий, методических рекомендаций по проведению учебных занятий, а также методических материалов к используемым в образовательном процессе техническим средствам

7.4.1 Методические рекомендации

1) Захарченков К. В., Вайнилович Ю.В. Методы сбора, обработки и анализа данных. Методические рекомендации к лабораторным работам для студентов специальности 09 03 04 «Программная инженерия» дневной формы обучения (электронный вариант).

7.4.3 Информационные технологии

- Тема 1. Введение.
- Тема 2. Процесс анализа больших объемов данных.
- Тема 3. Программное обеспечение для автоматизации обработки больших объемов данных.
- Тема 4. Машинное обучение.
- Тема 5. Общие методы обработки больших объемов данных.
- Тема 6. Средства распределенного хранения и обработки больших данных.
- Тема 7. Глубокий анализ текста.

7.4.3 Перечень программного обеспечения, используемого в образовательном процессе

- 1. MS Visual Studio 2019 Community Edition (свободно распространяемое)*
- 2. Eclipse (свободно распространяемое)*

8 МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Материально-техническое обеспечение дисциплины содержится в паспорте лаборатории а. 517/2, рег. № паспорта лаборатории № ПУЛ - 4 517/2-21; в паспорте лаборатории а. 518/2, рег. № паспорта лаборатории № ПУЛ - 4 518/2-21.