

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ

"БРАТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ"

УТВЕРЖДАЮ

Проректор по учебной работе

_____ Е.И.Луковникова

_____ 16 июня _____ 20 23 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.В.04 Введение в анализ больших данных

Закреплена за кафедрой **Информатики, математики и физики**

Учебный план bz090302_23_ИСиТ.plx

Направление: 09.03.02 Информационные системы и
технологии

Квалификация **Бакалавр**

Форма обучения **заочная**

Общая трудоемкость **4 ЗЕТ**

Виды контроля на курсах:

Контрольная работа 5, Экзамен 5

Распределение часов дисциплины по курсам

Курс	5		Итого	
	уп	рп		
Лекции	4	4	4	4
Лабораторные	10	10	10	10
В том числе инт.	2	2	2	2
В том числе в форме практ.подготовки	10	10	10	10
Итого ауд.	14	14	14	14
Контактная работа	14	14	14	14
Сам. работа	121	121	121	121
Часы на контроль	9	9	9	9
Итого	144	144	144	144

Программу составил(и):

б.с., ст.пр., Васильева Л.В. _____

Рабочая программа дисциплины

Введение в анализ больших данных

разработана в соответствии с ФГОС:

Федеральный государственный образовательный стандарт высшего образования - бакалавриат по направлению подготовки 09.03.02 Информационные системы и технологии (приказ Минобрнауки России от 19.09.2017 г. № 926)

составлена на основании учебного плана:

Направление: 09.03.02 Информационные системы и технологии
утвержденного приказом ректора от 17.02.2023 № 72.

Рабочая программа одобрена на заседании кафедры

Информатики, математики и физики

Протокол от 21 апреля 2023 г. № 9

Срок действия программы: 2023-2028 уч.г.

Зав. кафедрой Горохов Д.Б.

Председатель МКФ

старший преподаватель Латушкина С.В. 24 апреля 2023 г. № 9

Ответственный за реализацию ОПОП _____ Д.Б. Горохов

Директор библиотеки _____ Сотник Т.Ф.

№ регистрации _____ 40 _____
(методический отдел)

Визирование РПД для исполнения в очередном учебном году

Председатель МКФ

_____ 2024 г.

Рабочая программа пересмотрена, обсуждена и одобрена для
исполнения в 2024-2025 учебном году на заседании кафедры
Информатики, математики и физики

Внесены изменения/дополнения (Приложение _____)

Протокол от _____ 2024 г. № ____

Зав. кафедрой _____

Визирование РПД для исполнения в очередном учебном году

Председатель МКФ

_____ 2025 г.

Рабочая программа пересмотрена, обсуждена и одобрена для
исполнения в 2025-2026 учебном году на заседании кафедры
Информатики, математики и физики

Внесены изменения/дополнения (Приложение _____)

Протокол от _____ 2025 г. № ____

Зав. кафедрой _____

Визирование РПД для исполнения в очередном учебном году

Председатель МКФ

_____ 2026 г.

Рабочая программа пересмотрена, обсуждена и одобрена для
исполнения в 2026-2027 учебном году на заседании кафедры
Информатики, математики и физики

Внесены изменения/дополнения (Приложение _____)

Протокол от _____ 2026 г. № ____

Зав. кафедрой _____

Визирование РПД для исполнения в очередном учебном году

Председатель МКФ

_____ 2027 г.

Рабочая программа пересмотрена, обсуждена и одобрена для
исполнения в 2027-2028 учебном году на заседании кафедры
Информатики, математики и физики

Внесены изменения/дополнения (Приложение _____)

Протокол от _____ 2027 г. № ____

Зав. кафедрой _____

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1	Формирование профессиональной компетенции в области разработки и использования систем обработки и анализа больших массивов данных.
-----	--

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП

Цикл (раздел) ООП:	Б1.В.04
2.1	Требования к предварительной подготовке обучающегося:
2.1.1	Базы данных
2.1.2	Математическое моделирование
2.1.3	Программирование
2.1.4	Алгоритмы и структуры данных
2.1.5	Теория вероятностей и математическая статистика
2.1.6	Корпоративные информационные системы
2.2	Дисциплины и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:
2.2.1	Производственная (преддипломная) практика
2.2.2	Выполнение и защита выпускной квалификационной работы

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

ПК-2: Способность разрабатывать прототип информационной системы на базе типового решения и кодировать на языках программирования

Индикатор 1 | ПК-2.2. Разрабатывает код информационной системы (базы данных информационной системы), используя современные языки и технологии программирования

В результате освоения дисциплины обучающийся должен

3.1	Знать:
3.1.1	особенности анализа больших данных.
3.2	Уметь:
3.2.1	применять технологии анализа больших данных.
3.3	Владеть:
3.3.1	навыками разработки кода информационной системы с помощью технологий анализа больших данных.

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Код занятия	Вид занятия	Наименование разделов и тем	Семестр / Курс	Часов	Компетенции	Литература	Инте ракт.	Примечание
	Раздел	Раздел 1. Введение в большие данные						
1.1	Лек	Современные подходы к хранению, обработке и использованию данных. Предпосылки формирования тренда больших данных. Тенденции и перспективы рынка больших данных и их инструментов. Возможности для бизнеса. Большие данные (Big Data): понятие, сущность, ключевые признаки. Жизненный цикл больших данных. Экосистема больших данных.	5	1	ПК-2	Л1.1 Л1.2Л2.1 Л2.2 Л2.4	0,5	ПК-2.2 Лекция-визуализация
1.2	Ср	Подготовка к лекциям (установочная сессия), подготовка к экзамену (межсессионный период)	5	20	ПК-2	Л1.1 Л1.2Л2.1 Л2.2 Л2.4	0	ПК-2.2
1.3	Экзамен	Подготовка и сдача экзамена	5	1	ПК-2	Л1.1 Л1.2Л2.1 Л2.2 Л2.4	0	ПК-2.2

	Раздел	Раздел 2. Технологии хранения и анализа больших данных						
2.1	Лек	Архитектура системы обработки больших данных. Масштабирование и многоуровневое хранение данных. Парадигма NoSQL. Классификация NoSQL баз данных. Базы данных NewSQL и полихранилища. Высокопроизводительные вычисления. Параллельные алгоритмы для работы с данными. Вычислительная парадигма MapReduce. Визуализация данных и результатов анализа. Сложные методы аналитики. Программные платформы и системы для больших данных. Проект Apache Hadoop и его экосистема. Инструментарий Apache Hadoop.	5	3	ПК-2	Л1.1 Л1.2Л2.1 Л2.2 Л2.4	0,5	ПК-2.2 Лекция-визуализация
2.2	Лаб	MySQL для больших данных	5	5	ПК-2	Л1.1 Л1.2Л2.1 Л2.3 Л2.4 Э1	0	ПК-2.2
2.3	Лаб	Основы Hadoop	5	5	ПК-2	Л1.1 Л1.2Л2.1 Л2.2 Л2.4 Э2	1	ПК-2.2 Работа в малых группах
2.4	Ср	Подготовка к лекциям и лабораторным работам (установочная сессия); выполнение лабораторных работ, выполнение контрольной работы, подготовка к экзамену (межсессионный период)	5	101	ПК-2	Л1.1 Л1.2Л2.1 Л2.2 Л2.4 Э2	0	ПК-2.2
2.5	Контр.раб.	Подготовка к защите и защита контрольной работы	5	4	ПК-2	Л1.1 Л1.2Л2.1 Л2.2 Л2.4 Э2	0	ПК-2.2
2.6	Экзамен	Подготовка и сдача экзамена	5	4	ПК-2	Л1.1 Л1.2Л2.1 Л2.2 Л2.4	0	ПК-2.2

5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Традиционная (репродуктивная) технология (преподаватель знакомит обучающихся с порядком выполнения задания, наблюдает за выполнением и при необходимости корректирует работу обучающихся)

Технология компьютерного обучения(использование в учебном процессе компьютерных технологий и предоставляемых ими возможностей (электронные библиотеки))

Технология коллективного взаимодействия (работа в малых группах) (самостоятельное изучение обучающимися нового материала посредством сотрудничества в малых группах, дает возможность всем участникам участвовать в работе, практиковать навыки сотрудничества, межличностного общения)

Технология дистанционного обучения (получение образовательных услуг без посещения университета, с помощью современных систем телекоммуникации (электронная почта, Интернет и др.))

Образовательные технологии с использованием активных методов обучения (лекция-визуализация)

6. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

6.1. Контрольные вопросы и задания

ЛР 1. MySQL для больших данных.

Задание:

Изучить возможности MySQL для обработки больших данных.

Вопросы к защите:

1. Перечислите особенности обработки больших данных.
2. Какие инструменты MySQL могут использоваться для обработки больших данных?
3. Опишите механизмы выбора данных.

ЛР 2. Основы Hadoop (работа в малых группах - 1 час).

Задание:

1. Выполнить установку Hadoop кластера.
2. Освоить базовые операции обслуживания кластера и файловые операции HDFS.
3. Приобрести навыки управления ресурсами и запуском задач, использования основных инструментов и компонент Apache Hadoop экосистемы.

Вопросы к защите:

1. Опишите порядок установки Hadoop кластера.
2. Перечислите базовые операции обслуживания кластера.
3. Особенности файловых операций HDFS.
4. Перечислите основные компоненты Apache Hadoop экосистемы.
5. Назначение основных инструментов Apache Hadoop экосистемы.

Лекция-визуализация (0,5 часа), тема: Жизненный цикл больших данных. Экосистема больших данных.

Лекция-визуализация (0,5 часа), тема: Программные платформы и системы для больших данных. Проект Apache Hadoop и его экосистема.

6.2. Темы письменных работ

Целью контрольной работы является закрепление знаний об организации больших данных и практических навыков применения современных технологий их обработки и анализа.

Тема контрольной работы определяется вариантом индивидуального задания, который устанавливает предметную область получения данных, задачу обработки данных:

1. Социальные медиа.
2. Политика.
3. Наука и исследование.
4. Энергетика.
5. Здравоохранение.
6. Бизнес.
7. Транспорт.

6.3. Фонд оценочных средств

Экзаменационные вопросы

Раздел 1. Введение в большие данные.

1. Роль цифровой информации в 21 веке. Предпосылки формирования тренда больших данных.
2. Тенденции и перспективы рынка больших данных и их инструментов. Возможности для бизнеса.
3. Большие данные (Big Data): понятие, сущность, ключевые признаки.
4. Жизненный цикл больших данных.
5. Экосистема больших данных.

Раздел 2. Технологии хранения и анализа больших данных.

1. Архитектура системы обработки больших данных.
2. Масштабирование и многоуровневое хранение данных.
3. Парадигма NoSQL. Классификация NoSQL баз данных.
4. Базы данных NewSQL и полихранилища.
5. Высокопроизводительные вычисления. Вычислительная парадигма MapReduce.
6. Визуализация данных и результатов анализа.
7. Сложные методы аналитики.
8. Программные платформы и системы для больших данных.
9. Проект Apache Hadoop и его экосистема. Инструментарий Apache Hadoop.

6.4. Перечень видов оценочных средств

Лабораторные работы, контрольная работа, экзаменационные вопросы.

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

7.1. Рекомендуемая литература

7.1.1. Основная литература

Авторы,	Заглавие	Издательство, год	Кол-во	Эл. адрес
---------	----------	-------------------	--------	-----------

	Авторы,	Заглавие	Издательство, год	Кол-во	Эл. адрес
Л1. 1		Базы данных в высокопроизводительных информационных системах: учебное пособие	Ставрополь: Северо-Кавказский Федеральный университет (СКФУ), 2016	1	http://biblioclub.ru/index.php?page=book&id=466799
Л1. 2	Лесковец Ю., Раджараман А., Ульман Дж.	Анализ больших наборов данных: Учебное пособие	Москва: ДМК Пресс, 2016	1	https://ecat.brstu.ru/catalog/Ресурсы%20свободного%20доступа/Лесковец%20Ю.Анализ%20больших%20наборов%20данных.2016.pdf

7.1.2. Дополнительная литература

	Авторы,	Заглавие	Издательство, год	Кол-во	Эл. адрес
Л2. 1	Д. Келлехер, Б. Тирни	Наука о данных: базовый курс: учебное пособие	Москва : Альпина Паблишер, 2020	1	http://biblioclub.ru/index.php?page=book&id=598235
Л2. 2	Силен Д., Мейсман А., Али М.	Основы Data Science и Big Data. Python и наука о данных: учебное пособие	Санкт-Петербург: Питер, 2017	1	https://ecat.brstu.ru/catalog/Ресурсы%20свободного%20доступа/Силен%20Д.Основы%20Data%20Science%20и%20Big%20Data.%20Python%20и%20наука%20о%20данных.2017.pdf
Л2. 3	Чаллавала Ш. и др.	MySQL 8 для больших данных: учебное издание	Москва: ДМК Пресс, 2018	1	https://ecat.brstu.ru/catalog/Ресурсы%20свободного%20доступа/Чаллавала%20Ш. MySQL%208 для%20больших%20данных.2018.pdf
Л2. 4	Радченко И.А., Николаев И.Н.	Технологии и инфраструктура Big Data: учебное пособие	Санкт-Петербург: Университет ИТМО, 2018	1	https://ecat.brstu.ru/catalog/Ресурсы%20свободного%20доступа/Радченко%20И.А.%20Технологии%20и%20инфраструктура%20Big%20Data.2018.pdf

7.2. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"

Э1	Официальный сайт MySQL	https://www.mysql.com/
Э2	Официальный сайт проекта Apache Hadoop	https://hadoop.apache.org/

7.3.1 Перечень программного обеспечения

7.3.1.1	Adobe Acrobat Reader DC
7.3.1.2	LibreOffice
7.3.1.3	Chrome
7.3.1.4	MySQL Community Edition
7.3.1.5	Python IDLE
7.3.1.6	MySQL Server
7.3.1.7	Ай-Логос
7.3.1.8	Django
7.3.1.9	Office Professional Plus 2010
7.3.1.10	Jupyterlab

7.3.2 Перечень информационных справочных систем

7.3.2.1	Университетская информационная система РОССИЯ (УИС РОССИЯ)
7.3.2.2	Национальная электронная библиотека НЭБ
7.3.2.3	Научная электронная библиотека eLIBRARY.RU
7.3.2.4	Электронная библиотека БрГУ
7.3.2.5	Электронный каталог библиотеки БрГУ
7.3.2.6	«Университетская библиотека online»

8. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

2201	читальный зал №1	Комплект мебели (посадочных мест) Стеллажи Комплект мебели (посадочных мест) для библиотекаря Выставочные шкафы ПК i5-2500/H67/4Gb (монитор TFT19 Samsung) (10шт.); принтер HP Laser Jet P2055D (1шт.)
1348	Учебная аудитория (дисплейный класс)	Основное оборудование: -персональный компьютер i5-2500/H67/4Gb/500Gb – 1 шт; -системный блок AMD 690G/FA– 12 шт; -монитор TFT19 Samsung E1920NR – 13 шт; Дополнительно: Маркерная доска – 1 шт. Учебная мебель: Комплект мебели (посадочных мест /АРМ) - 24/12 шт. Комплект мебели (посадочных мест) для преподавателя – 1 шт.
1346	Учебная аудитория (дисплейный класс)	Основное оборудование: Системный блок CPU 5000/RAM 2Gb/HDD250Gb/2Gb- 16 шт. Монитор TFT 19" LG L1953S-SF- 16 шт. Интерактивная доска SMARTBoard 680I (77"/195,6 см) - 1 шт. Проектор мультимедийный торговой марки "CASIO" модель XJ-UT310WN с настенным креплением CASIO YM-80 - 1 шт. Принтер HP LaserJet P3005 - 1 шт. Коммутатор D-link DES1026G - 1 шт. Учебная мебель: Комплект мебели (посадочных мест/АРМ) – 32/16 шт. Комплект мебели (посадочных мест) для преподавателя – 1 шт.
1346	Учебная аудитория (дисплейный класс)	Основное оборудование: Системный блок CPU 5000/RAM 2Gb/HDD250Gb/2Gb- 16 шт. Монитор TFT 19" LG L1953S-SF- 16 шт. Интерактивная доска SMARTBoard 680I (77"/195,6 см) - 1 шт. Проектор мультимедийный торговой марки "CASIO" модель XJ-UT310WN с настенным креплением CASIO YM-80 - 1 шт. Принтер HP LaserJet P3005 - 1 шт. Коммутатор D-link DES1026G - 1 шт. Учебная мебель: Комплект мебели (посадочных мест/АРМ) – 32/16 шт. Комплект мебели (посадочных мест) для преподавателя – 1 шт.

9. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

Учебным планом предусмотрены следующие виды учебных занятий: лекции, лабораторные работы, контрольная работа, самостоятельная работа обучающегося, экзамен.

Лекции.
Написание конспекта лекций: краткое, последовательное изложение основных положений, формулировок, выводов, обобщений; техническое оформление записей (подчеркивание, выделение ключевых слов и терминов). Активная работа на лекции.

Лабораторные работы.
Выполнение заданий с использованием методических указаний по выполнению лабораторных работ, оформление отчетов, защита лабораторных работ.
Самостоятельная работа обучающихся.

- Подготовка к лекциям: систематическая работа с конспектом лекций (чтение записей; проверка терминов с помощью энциклопедий, словарей и справочников; обозначение вопросов, материал, которых вызывает трудности; попытка найти ответ в рекомендуемых источниках; подготовка вопросов преподавателю, если не удастся самостоятельно разобраться в материале).
- Подготовка к лабораторным работам: проработка материалов по теме лабораторной работы с использованием рекомендуемой литературы, конспекта лекций, ресурсов информационно-телекоммуникационной сети Интернет; оформление отчетов по лабораторным работам; подготовка к защите лабораторных работ.
- Выполнение контрольной работы: выполнение задания с использованием рекомендуемых источников; оформление отчета; подготовка к защите контрольной работы.
- Подготовка к экзамену: проработка вопросов с использованием конспекта лекций, рекомендуемых источников; подготовка вопросов преподавателю, если не удастся самостоятельно разобраться в материале.